Building institutional indicators:

Some theoretical perspectives and methodological propositions

by

ERIC BROUSSEAU

SOPHIE HARNAY

JEAN-FRANÇOIS SATTIN

This version: 9/29/2009

Very First Draft – Please do not quote

Abstract: As empirical assesments of institutions through composite indexes have becomed one of the most dynamic trend of research in institutional economics, some contributors still worry about their lack of theoretical leakage and about the robustness of the computational technics mobilized. This papers first define the gap between the institutional theory and the ones mobilized in institutional indicators. Precising what should be assessed by institutional measures, it derives then from the aggregation literature some propositions in order to check the validity of institutional composites. (JEL: D0, O43)

"Capital" is not what capital is called, it is what its name is called.

Joan Robinson,

The Production Function and the Theory of Capital,

The Review of Economic Studies, Vol. 21, N° 2, (1953 - 1954), p. 83

Social sciences have often suffered from a tension between conceptualization and measurement. The need for theorization leads to the emergence of concepts that may be abstract and difficult to operationalize when researchers when researchers seek to test predictions. Such dilemmas are common in sociology and economics where macro phenomena are often linked to barely observable concepts (BLALOCK [1971]). For instance debate amongst economic scholars about capital measurement should look familiar to people involved in institutional assessment and the foreword of Joan Robinson could easily be applied to "institutions" instead of "capital" (FELIPE AND FISHER, 2003; COHEN AND HARCOURT, 2005). Then as attempts to measures institutional characteristics have became more common in economics, one can ask if our understanding of institutional phenomena and current methodological toolbox enable us to do so.[1]

Governance indicators have emerged from an initiative to test and develop policy-relevant, politically acceptable, quantitative indicators of governance (KNACK AND AL., 2003). Recent years have seen a remarkable growth in such indicators, as well as a refinement of institutional measurement techniques. Even if estimates still differ, there should be more or less 175 such indicators developed by international organizations, research institutes, and private organizations (UNITED NATIONS, 2007). This phenomenal growth in governance indicators has partly been fuelled by a rising demand for such indicators by policymakers, researchers, and all kind of organizations — in particular those involved in trade, foreign direct investments and socio-political reforms — in order to assess institutional outcomes and to design institutional policies. Indeed, institutional indicators, that were initially devoted to private users (e.g. especially to assess hazards linked to trade and FDI), have also been recently widely used as policy tools in analysis and intensively used by national and international organization (such as the World Bank) in order to devise their funding policy. This reliance on institutional index is grounded on their capacity to provide simple

[1] Governance is a very broad concept that operates at every level of a society (NZONGOLA-NTALAJA, 2002). Many types of governance were defined in the literature depending on of the authority we focus on (MANNING ET KRANN, 2006). The fact that governance is just an expression of institutional arrangements present in a country lead us to consider governance and institutions as the two faces of a same concept. Thus we will use the two terms equally well throughout this paper.

comparisons across countries that can be used to analyze complex and sometimes elusive legal and political-economic issues. They seem to be easier to interpret than qualitative analysis or 'research' relying on wide set of statistical data. In the same time, the fact that institutional indicators could end-up sending misleading policy messages if inappropriately constructed or badly interpreted initiated a strong international controversy. In particular several papers question the reliability of most of institutional measures (e.g. ARRUNADA, 2007; VRIES, 2007). One should neither built nor trust institutional composites without being familiarized with the pitfalls associated to institutional measurements and aggregation techniques. Therefore the purpose of this survey is twofold. First it aims to summarize the existing literature on institutional measurement by focusing on recent advances in institutional theories and in assessments techniques. It aims also to define the conditions under which robust composites can actually exist and to provide some hints in order to check the consistency of institutional composites.

The paper is organized as follow. Our first section details theoretical flaws of the current institutional measures and presents the theoretical challenges associated to institutional measurement. In section 2, we discuss the technical conditions that should be taken into consideration to build more satisfactory institutional measures, suggesting some path of development for research in this domain. Section 3 briefly concludes the paper. In appendix an important note points out the parallel between the issues raised today by the computation of institutional aggregates and the debates that occurred in the past to build relevant macroeconomic aggregates.

1. Current challenges in conceptualizing institutional designs

Measuring institutions requires first to make clearer the aspect of institutions that is going to be targeted by the effort of quantification. Indeed, the concept of institution is very broad and encompass a number a quite heterogeneous set of collective constructs and social phenomena. This leads to highlight the fact that institutions are hardly measured independently of the purpose of the analysis that underlies the effort to measure it. Once the relevant aspect(s) of institutions referred to is(are) clarified, the concept has to be operationalized by indentifying the attributes that could be observed and compared from one institutional setting to another. Lastly, methods of measures have to be designed. The sequence definition => attributes =>

measure revels essential issues at each stage. Moreover, once measures have been realized, the ability to aggregate them so as to generate synthetic indexes that well capture the characteristics of attributes and the concept that is attempted to be measured (and therefore the inverse sequence) is also an essential issue. In the first part of this paper we will successively discuss the conceptual problems lying being the attempt to measure of institutions, to identify the relevant attributes, and to measure them; the second part of the paper focusing on the reverse sequence.

1.1. Definitional issues : what is an institution ?

1.1.1. The inherent biases in defining institutions

An important issue associated with the building of institutional indicators is directly related to the specification of the meaning of the concept and to the identification of its attributes (MUNCK AND VERKUILEN [2002]). Yet, as no unanimous definition is available, "measuring" institutions entails a risk of selecting a "wrong" or unsuitable definition of the concept. Therefore, any attempt to measure institutions may come up against two risks. On the one hand, absent a unified definition of institutions, it may adopt a maximalist definition that may turn unworkable because it includes attributes that are more suitable for the specification of other concepts. On the other hand, it runs the risk to confine itself within a minimalist definition that may lead to the omission of relevant attributes for the concept, that is, fundamental dimensions or features of the institutions at stake; a straightforward reason for such omissions being the very difficulty to measure the dimension or feature and to find a relevant proxy. Furthermore, the definition of institutional concepts may not always be imposed on by a theoretical background but rather driven by data availability, which is hardly satisficing from a scientific standpoint.

According to NORTH [1991, P. 97], institutions can be defined broadly as "the humanly devised constraints that structure political, economic and social interactions" and "consist of both informal constraints (sanctions, taboos, customs, traditions, and codes of conduct) and formal rules (constitutions, laws, property rights)". Such an extended definition is of little help to build indicators at the stage of their conceptualization. Williamson's partition of institutions among four social levels (WILLIAMSON [2000, P. 597]) may be more helpful. Williamson distinguishes a highest level, that he qualify as "the social-embededness level", which is composed of "norms, customs, moral and traditions", that he consider to be most of

the time taken as given by agents. They change only very slowly and are difficult to "monitor". The second level of analysis is qualified as "the institutional environment", understood as "the formal rules of the game" and includes "the executive, legislative, judicial, and bureaucratic functions of government as well as the distribution of powers across different levels of government". The "definition and enforcement of property rights and of contract laws [are] important features" (*IBID*., p. 598). The third level is "where the institutions of governance are located", with a focus on the governance of contracts. The fourth and last level deals with the governance of resource allocation. According to Williamson (*IBID*., p. 600), it is "the level at which neoclassical analysis works". While different levels are indentified, enabling to characterize different analytical purposes, hence relevant set of governance characteristic to focus on, the conceptualization of institutional indicators remains tricky because the various "levels" of analysis encompass quite wide set of issues and collective constructions.

In other words, absent a comprehensive theory of institutions or a consensus on those institutions that matter, most attempt to measure institutions have have been being based on partial definition of institutions, according to their own purposes, while leaving behind those institutions that are most uneasy to define and measure. The issue here is that institutions are clearly multi-dimensional and multi-purpose, and that any to focus on specific charactertics/dimensions leaving aside other may well lead to measure that will bias the vision. The dilemma is the following. Since it is impossible to take into considerations all aspects of the social constructs that contribute to establish "the rules of the game" — i.e. the basis on which human action is built and capability of decision are framed —, the definition of the relevant institutions to be measured depend upon the purpose of the analysis/policy to be undertaken. Institutional indicators will inherently be biased by their purpose. In addition, since considering the relevant aspect of institutions in function of an (analytical of political) purpose request to identify the causal relationship between institutional charactertics and the phenomena at stake, any measure might be flawed because of an imperfect theorization of the transmission mechanism through which institutions results in individual and collective effects. In the current state of the development of the economics of institutions, but this covers many other social sciences ast well, these "transmission" mechanisms are far from beeing well known. For instance, the mechanisms through which "democracy" generates its effect — e.g. checks and balances, accountability, challengeability of positions, public debate, etc. — are far from beeing well established.

Measures of institutions are thus inherently biased by the focus of the analyis in the frame of which they are developed. Moreover, they reflect the knowledge and beliefs about the transmission factors related to the stake. Both reasons explain the strong biases characterizing the existing availlable institutional "measures". For instance, since most of them have been developed in view of political reforms, almost no measure has been developed to grasp Williamson's first level of institutions, i.e. informal instituions. At best, characterization borrowed to other sciences, of proxied (like language), are relied upon as control variables in research work. This clearly reflects the fact that "the concept of embeddedness remains in need of greater theoretical specification" (SMELSER AND SWEDBERG [1994]). In the same manner, very few works deal with the relationships among levels of governance in the characterization of institutional regimes. For instance, the characteristics of property rights depends upon the combination of mechanisms operating at the four levels of Williamson's hierarchy — as analyzed and documented in details by Barzel (1989) —, which should lead any "indicator" on the matter not to focus only on the formal institutions. In the same line, too few works are able to hiearchize various institutional components according to the way they influence each other, which should lead to pay a greater attention to some "mother" characteristics; as for instance democratic tools that might be considered as useful "for building good institutions" (RODRIK [2000, P. 3]).

The absence of a well-established theory of institutions has other consequences. In particular, several pitfalls should be considered when building/using institutional indicators. First, the insufficient definition of institutional concepts — or the insufficient reliance on existing theories — is conducive to misusing concepts and contributes to semantic "dilution". Institutional indicators are sometimes built upon theoretical concepts that are defined in a restrictive or incorrect sense. At least what is measured differs strongly from the usual acception by the literature. Thomas [2007], for instance, remarks that whereas the concept of "voice" is well defined in the political science literature, it is "given a rough ride" in the Worldwide Governance Indicators (WGI). No reference is made to Hirschman explicitly, and voice and accountability are brought together within the same indicator – i. e. "voice and accountability".[2] In the same manner, also due to such a conceptual uncertainty, indicators

[2] According to THOMAS (2007, P. 18), "terms such as "voice," "accountability," and "rule of law" are concepts featured in existing theoretical literatures, but the definitions of the WGI neither refer to nor correspond with these literatures. For example, the indicator "Voice and Accountability" is defined as "the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association and free media."

that are expected to measure a same concept may actually rely on different concepts. For instance, "[T]here are numerous definitions of corruption in the academic literature and among donor agencies. Most of these definitions are quite broad, and often somewhat vague. Transparency International's definition, "the misuse of entrusted power for private gain," is representative" (KNACK [2006, P. 5]. Hence, "[R]egardless of one's preferred conceptual definition, the choice of measurement techniques from a limited set of feasible alternatives inevitably produces an implicit definition that may differ substantially from one's ideal. Any pair of assessment methodologies will measure a different (if unknown) mix of these various dimensions of corruption. For example, what weight should be given to central, state and local governments when assessing "corruption" for federal countries such as the United States or India? These sorts of questions typically are not explicitly answered in the methodology of existing country-level indicators". (*IBID.*, PP. 5-6). [3]

1.1.2. Causality and Focus

A challenge associated with definitional issues deals with the identification of those institutions that really matter or matter more than others. In absence of a comprehensive theory of institutions, no consensus really prevails on those institutions that should be taken into account when building institutional measures and indicators. Consistently with the dominant thinking in economics and following up the neo-institutional thought as well, the economic literature usually emphasizes the role of market-supporting institutions (RODRIK [2000]). Among those, institutions securing property rights are usually considered to be of paramount importance and an object of a wide array of measures. However, albeit important, they may not be the only relevant institution. Referring to BARDHAN [2004], PRZEWORSKI [2004, P. 8] underscores that other institutions have been brought to the front of the stage before, among others « those that coordinate investment » and those « that force savings ». In his list of institutions that matter, RODRIK [2000] not only includes property rights but also

---

(KAUFMANN AND AL. [2005]). In political science, the term "voice" derives from Hirschman's classic work, "Exit, Voice and Loyalty." Hirschman gives the following definition of "voice": "any attempt at all to change, rather than escape from, an objectionable state of affairs, whether through individual or collective petition to the management directly in charge, through appeal to a higher authority with the intention of forcing a change in management, or through various types of actions and protests including those that are meant to mobilize public opinion" (HIRSCHMAN [1970]). Accordingly, "voice" is not synonymous with accountability, freedom to select government, or other political freedoms. Nor are there well-known or well-documented relationships among them". She also remarks that "Rule of law" as it is developed by the WGI has no content validity since "it does not map to a definition of the Rule of Law".

[3] We discuss aggregation problems later in the paper.

regulatory institutions, institutions for macroeconomic stabilization, institutions for social insurance, and institutions of conflict management – all institutions likely to take the form of highly various institutional arrangements. More recently, the supporters of "Law and Finance" have put forward new contenders for the role of relevant variable. Among other important variables, they put strong emphasis on the role of financial institutions (See for instance LA PORTA AND AL. [1997, 1998, 1999, 2002]). However, in absence of theories that would really explain how specific institutions should lead to macroeconomic outcome, taking into consideration other institutional components, there is no certaintay that the most focussed, quoted, and used institutional indicators are the most relevant ones. In a provocative way, one can even claim that the emphasis put on a type of institution preferently may (also) be a fad.

It's worht to note in particular that the absence of a relevant theory plays an incrasing role as one consider more agregated phenomena. Indeed, micro-analysis (and microeconometrics) or institutions can lead to identify and even specify causal relationships between a "rule" and a set of "human behaviors", as the deterence effect of emprisonment (e.g. Drago, Galbiati Vertova, 2009). It is more complex however to assess the impact of an institutional setting on a complex social phenonmena, as the relationship between the judicial system (in all its dimensions: prevention, investigation, repression) on criminality (which is itself a multidimensionnal phenomena) given the impact of education, social cohesion, labor market, etc. Obviously, the relationship between an institutional characteristic (among many) and a macro-economic concept as growth is even less well established. This is one of the issue LLSV's analysis of legal origin. These authors focus on some specific aspect of the law, essentially the protection of shareholders against managerial and governmental hazards, and the development of financial market — which is only one of the several way to convey savings to productive investments. It is however heroic to make any grounded prediction on the impact of these micro-instituions on growth due to the many variables that influence the formation of GDP and the many possible arrangements to organize funding and choice in matter of investment.

Thus, while certainly some institutions may be more important than some others, nothing ensures that the indicators focus mainly on the former. True, empirical analyses allow for identifying institutions that do have an impact on a dependent variable. But so far, provided the current state of the art, one cannot claim that *those* institutions that are being tested are the most relevant ones or the only ones. Many researchers are aware of this limitation and recognize that the measure they rely on may fail to fully grasp institutional influence.

However, absent a first-best solution, they concentrate on those institutions that are most often taken into consideration and for which institutional measure exist. This might induce strong biases. First, nothing guarantee that the relevant institutional dimensions are taken into account. Second, as it will be develop later, this might also lead to focus on institutions that that are easier to "measure".

The first issue is well illustrated by the reliance on apparent "neutral" descriptors as of legal families, which universality and power of identification may be questionned. Indeed, they suppose that all legal dimensions vary the same way and when one switch from any country belonging to a given family to another one belonging to another. Since it is widely recognized that most institutional compact are highly specific due to history, collective preferences, political choices, specific equilibria reinforced by path-dependency effects, etc., it is doubtful that is is the case and two countries that can be classified in different legal family regarding familiy law may well belong to the same "family" in matter of intellectual property. In this case, the use "universal" descriptors to build institutional indicators may fail to grasp actual characteristic of legal systems and to relevantly assess the distance between two national legal regimes.

The opposite approach, consisting in focussing on "service" supplied by an institution may also fail to efficiently identify relevant institutional characteristics. From that perspective, the "one-size fits all" approach, consisting in comparing institutions from a single perspective, may come up against institutional understanding. For instance, the *Doing Business* reports released by the World Bank have been criticized strongly along this line. Reducing the quality of institutions to the (supposed) easyness in launching a new business, and ignoring other dimensions such as the security of property rights, or the protection of consumers, of business partners or of minority shareholders may fail to take into consideration some essential dimensions explaining behaviors and performances.

Focussing on specific institutional features, may also well bias analyses of institutional evolutions. Too selective approach may fail to characterize the initial situation of a country, and then to actually measure institutional change and improvement. Adapted approach should possibly be used in function to initial situations. For instance, "it is important to recognize that regulatory institutions may need to extend beyond the standard list covering anti-trust, financial supervision, securities regulation, and a few others. This is true especially in

developing countries where market failures may be more pervasive and the requisite market regulations more extensive" (RODRIK [2000, P. 7].

This set of last points raises the issue of the efficiency of (supposely) "universal descriptor" when institutional complementarities could exist. In the last section of this paper we get back to this. Since what has to be taken into accoutn is the way institutions impacts on bevaviors, it may well be more relevant to consider the (micro) "effects" of institutions rather than their "form", when tryinng to describe their characteristic. Risks are however obvious. In particular, it may reduce the vision of institution to a functionalist one (see below). Moreover, since one of the goal of institutional "measures" is to assess performances, it could lead to tautologies.

Another important issue concerns the purposes of institutions and the criterion against which one wants to assess them. From that standpoint, the debate remains very poor, to say the least. Different institutions may matter for different purposes, while only one dependent variable is generally considered in the literature: growth or alternative proxies for economic performance. This overlooks the ongoing discussion about the relevance of growth rates as a single indicator to measure development. Whereas progress towards a better understanding of the causality between institutions and economic growth has been made undoubtedly, no mention is ever done of other – possibly complementary – purposes of institutions, such as social cohesion, happiness... Certainly, "markets need to be supported by non-market institutions in order to perform well" (RODRIK [2000, P. 2])[4] and market-oriented reforms backed up by other institutions. Notwithstanding, "not all the institutions are there to serve the needs of the market economy first and foremost, even if their presence is required by the internal logic of private property and contract enforcement. The fact that a governance structure is needed to ensure that markets can do their work does not imply that the governance structure serves only that end. Non-market institutions will sometimes produce outcomes that are socially undesirable, such as the use of public office for private gains. They may also produce outcomes that restrict the free play of market forces in pursuit of a larger goal, such as social stability and cohesiveness" (*IBID.*, P. 5).

[4] As already pointed out, causal relationship are implicitly assumed, when the relevant institutions to focus on are decided. In particular, there is a strong focus on the performance of markets that is supposed to be directly linked to growth, while this certainly need to be clarified. Indeed, most developed economies rely on markets, but on markets that are framed by powerful institutions and in which the governments play a strong role.é

Along this line, it is worth to note that most institutional indicators are strongly biased by assumption on causal relationship that are simply implicitly assumed. In particular, there is a strong focus on the performance of markets that is supposed to be directly linked to growth, while this certainly need to be clarified. Indeed, most developed economies rely on markets, but on markets that are framed by powerful institutions and in which the governments play a strong role. Moreover, most indicators rely on an implicit conception of market that considers any regulation as burdensome and costly and more or less equates "good government" with "less government". From that perspective, most indicators tend to occult the whole array of socially and individually useful benefits provided by institutions. Thus, measures seldom include or valuate social services such as legal certainty and security or positive externalities that may derive from the acceptance of individual costs. For instance, measures like those developed in the *Doing Business* reports focus on the costs paid by user while neglecting "the fixed costs of creating and starting up formalization systems" (ARRUNADA [2007, P. 741]). In the same vein, the emphasis is put on the "quantity" of regulations at the expense of their quality, whereas opposite interpretations can be given to a same rule or institution. For instance, a higher length of procedures, red tape or compulsory requirements tend to be systematically construed as a cost rather than as a token of regulatory quality (KURTZ AND SCHRANK [2007, P. 543]). The point is that these assumptions not only bias any analysis, they might have disastrous political consequences. As pointed out by ARRUNADA [2007] "the recipe for simplification ends up doing little more than cutting back formalities while disregarding the value of the services offered" (P. 730]).

As signaled above, the second essential drawback of existing indicators and analysis, especially when they are applied to macro-analysis, is to focus on factors that are the easier to measure, with a strong risk of underestimation of the impact of less "observable" variables. For instance, most indicators concentrate on law *de jure* rather than *de facto*, due to the higher facility to measure formal law and "law in the books" in comparison with some other sources of law. In the same manner, they tend to favour mandatory rules against default rules, since information on the former is more easily obtained. More generally, informality is poorly taken into account due to straightforward statistical reasons. The measure of labour law by the OECD through employment protection legislation (EPL) indicators provides an illustration of such limitations. Not only do EPL indicators fail to take the size and impact of black labor markets into account, but they also consider mainly the wording of the law concerning the full-time employed jobs; thereby excluding many major legal exceptions that characterize

temporary, part-time, and special status workers. These specific arrangements concern a significant part of the labor force in most countries (OECD [2004]). The impact of formal law and of every formal rule is at best overestimated. At a broader level, this emphasis put on formality by institutional measures can lead to sever misinterpretation; as shown clearly by the idea that civil law countries may be more largely associated with burdensome regulation than common law countries, may well rely on the fact that case law is not taken into account in some indicators as those of the Doing Business initiative. Generally speaking, the focus of most indicators on the lower levels of Williamson's typology, is largelly explained by the greater availability of data, and by more precise and easier definition as well. This does not guarantee, however, that the scientific community and economists in particular, are focussing on the most relevant explaining factors.

1.2. Identifying Attributes and Aggregation Issues

Another consequence of insufficient conceptualization at a former stage of the measurement procedure is that some studies may be flawed by the discrepancy between the theoretically relevant and the operational measures of institutions. Namely, as noted by PRZEWORSKI [2004, P. 7], quoting ACEMOGLU AND AL. [2002, P. 1270], the measure of institutions used by most studies may sometimes « correspond poorly to the real concept that is relevant to development (which is likely to be a broad range of institutions, whereas we only have an index for a particular type of institutions) ». For instance, most studies emphasizing the key function of property rights use the risk of expropriation as an index of secure property rights (PRZEWORSKI [2004, P. 7]), although it may not comprehend all the complexity and dimensions of the concept. In some cases and for some institutions, thus, imperfect conceptualization of the indicator may be related to disputable identification of the attributes to be taken into consideration, and of the relevan proxies to assess.

Also linked with the issue of conceptualizing institutional indicators, the identification of attributes for an institutional concept raises the additional question of the relation of the various attributes of a concept to each other. Attention is needed in relation to this issue in order to avoid both redundancy and conflation problems, that arise frequently as no unanimous and clear definition exists for a concept (MUNCK AND VERKUILEN [2002, P. 12-14]). A careful organization of attributes into components and subcomponents of attributes within a conceptual tree according to their level of abstraction is helpful both to avoid those problems and at the further stage of aggregation. In this respect, institutional indicators, being

very young, may not have completely succeeded the initial stage of identification of their attributes. Due to definitional uncertainty on institutional concepts, further research is still needed at the theoretical level, as it has been carried out in other fields for initially as broad and encompassing concepts like democracy.

In this respect, a major concern derives from the difficulty for indicators to grasp institutional complementarities and substitutions as well. On the one hand, indicators implicitly assume that institutions are separable, which clearly entails the risk of measuring institutions in a piecemeal fashion and, therefore, deprives the investigator from any possibility to measure interconnections and feedbacks in the system[5]. This omission is particularly annoying whenever cross-system comparisons are at stake, since in that case the mere observation of various institutional components and features separately does no longer allow investigators to draw relevant comparisons. On the contrary, if various institutional items interact with each other within specific institutional frameworks, then the impact and performance of one component will become strongly dependent upon the others. For instance, the actual impact of a rule will be contingent on several other institutions as diverse as the enforcement capacities of the system, the rule of law, or the informal institutions such as social norms prevailing within the community. As a straightforward consequence of this institutional entanglement, identical institutions may thus be identified in various countries or legal systems formally but nevertheless perform in a dramatically different way in relation to the existence and performance of complementary institutions. In other words, the mere observation of a specific institution in a given context does not ensure that it will produce the same outcome as in an alternative institutional framework. Political reforms failing to take that point into account may thus turn inefficient as long as they mostly intend to change local systems in a piecemeal fashion. On the other hand, this systematic disregard for institutional complementarities goes with omitting institutional substitutions symmetrically. In that respect, indicators fail to recognize that the absence of an institution in a given institutional setting is not synonymous

[5] Indeed, most of the time measured proxies are added to each other (linear computation) or multiplied together (geometric computation)[5]. As noted by NARDO ET AL. [2004] compensability logics are quite different in both cases : linear aggregation implies full compensability, e.g. poor performance in some indicators can be compensated by sufficiently high values of other indicators, when geometric aggregation entails partial compensability, e.g. compensability is lower when the composite indicator contains indicators with low values. But one can ask why institutional interactions should follow one of these two schemes. On the contrary, one can think that most of the time institutions interact in a non constant way, with the true functional form of the institution being non-linear[5]. If the computed measures face some important variations across the different countries, it is then problematic to assess this last one in a linear way by doing some limited development to a neighborhood.

for the function performed by this specific institution not being fulfilled. As a matter of fact, one or several substitute institutions may actually perform an identical function (ARRUNADA, 2007, PP. 736 *SQ*). For instance, facultative formalities may substitute for compulsory procedures to reach the same institutional outcome. Thus, a country characterized by both a small number of compulsory formalities and a high number of facultative provisions may not be systematically associated with a lesser regulatory burden, but may simply shift compulsory requirements towards facultative additional services that agents may want to purchase to grant themselves a given level of legal certainty. A form of temporal substitution may also take place between *ex ante* and *ex post* procedures, as the outcome of a tradeoff between initial protection and prevention and future conflict resolution, as some systems may favor *ex ante* procedures such as registration procedures, thereby increasing the so-called regulatory burden, whereas some other systems may shift these costs *ex post* and resort to the judiciary whenever a conflict arises. Hence, legal security, conflict resolution and prevention, and any final service provided by institutions in that matter may be obtained through various institutional combinations but will be conducive to very different assessments through indicators, since the latter clearly fail to take institutional substitutions into account. The little attention devoted to institutional substitutions and to their effects therefore limits the relevance and scope of the scenario method, to that extent that it relies on an implicit assumption of common or "universal" institutions available and equally efficient for all countries and their alleged comparability rather than builds on a real functionalist view of institutions.

Thus, by breaking down complex institutions into a set of different variables that are analyzed independently of each other, one occults the relations and interactions between these different variables and, thus, one foregoes any systemic understanding of the institution as a whole. As highlighted by SIEMS [2005, P. 529], such an approach fails to recognize that "complicated systems cannot be understood by breaking them into simpler parts. Most things do not follow deterministic rules and a lot of truths are not the logical outcome of one set of rules because systems are often dynamic and chaotic". This systemic approach to institutions is the reason why some scholars refuse the idea of universal descriptors able to grasp institutional characteristics for the purpose of comparative analysis.

1.3. Building Indicators and Measuring

Once an institutional concept has been conceptualized properly, it is still necessary to design those indicators that will sum-up measures of their components (BOLLEN [2001, P. 7286]). Many institutional indicators rely on the idea that institutions can be measured by breaking them down into a plurality of variables. For instance, KNACK [2006, P. 5]) disaggregates corruption into many dimensions (by level of political system, by purpose of the improper action, by the actors involved in the corrupt transaction…). Another telling example is provided by the WGI describes the rule of law along several dimensions, aggregating several individual sources from 21 organizations. The problem is however to check the validity of the indicators, that is to ascertain that they actually measure what they purport to. Three types of validity are usually considered. Construct validity examines whether the indicator is associated with other constructs that it is supposed to relate to and not associated with those that it should not (BOLLEN [2001, P. 7285]). In other words, it is concerned with "whether the proposed operationalization captures the entire domain of a construct and includes nothing extraneous" (THOMAS [2007, P. 10]). Criterion validity compares the indicator to some standard variable that it should be correlated to if it is valid. Content validity assesses whether the indicators are capturing the concept for which the latent variable stands (BOLLEN [2001, P. 7285]).

From this standpoint, the construction of institutional indicators has been criticized along several lines. First, and in relation to the theoretical problems pointed out above, the construct and criterion validity are difficult to test. Without explicit references to a well-established theoretical background, the choice of specific indicators to back up a concept is easily accused of reflecting mostly the ideology, preferences, and beliefs of the investigator. As his or her private choices often remain implicit and unjustified, indicators can be easily called arbitrary and biased, especially when they serve as a basis for policy prescriptions. A second line of criticism emphasizes that the institutional dimensions on which measurement focuses in practice are frequently the easiest to measure. In other words, the selection of key variables is not driven by their relevance but rather influenced by the availability and cost of information. The lack of current validity in the building of indicators simply relates to the fact that institutional theory is developing and that efforts of conceptualization are needed to overcome the present limits of institutional measures. There is however obviously a dog and tail issue since part of the theoretical developments — especially on relevant causal relationships —

request advances in measurement techniques to test and qualibrate theoretical propositions. This requests, however, at least content validity of the indicators relied upon.

In that perspective it is worth to note that "[C]ontent validity does not require a definitional consensus, but it does require that researchers rigorously define what it is they wish to measure before they set out to measure it, and that the definition have as much in common as possible with the way the construct is typically defined and used" (THOMAS [2007, P. 19]). Further, "[T]hat the names of the governance indicators are divorced from the literatures dealing with constructs of the same names does not mean that the governance indicators are wrong or useless. They may simply be new constructs that are misleadingly named. To introduce the new construct, the authors would have to explain the nature and importance of the new construct, the distinctions between the new construct and existing constructs, and the predicted relationship of the new construct to other variables of interest before attempting an operationalization" (THOMAS [2007, P. 52]). However, absent an additional justification, the risk is that indicators lose content validity.

Indeed as pointed out by Thomas (2007), reliance on a vague definition of institutions may be conducive to some *ad hoc* definition of the institutions that one tries to measure. About the WGI, THOMAS [2007, P. 19], again, notes that "[A]n examination of the underlying indicators suggests that the WGI construct definitions are merely summary descriptions of the indicators in the cluster". Based on cluster analysis the indicators are far from enabling a precise reconciliation with the theory, resulting in poor tools to test it. The same remark is easily extended to other indicators. In practice, the definition of many institutional dimensions is based on *ex post* interpretation of what is measured, the later being driven by the availlability of data. This collides with the basic requirement of a rigorous conceptualization process. This is clearly unconsistent not only with best practices recommendations in statistical methodology but also with actual practices carried out by academic fields that are more in advance on that topic (see comparative political science for instance).

Beyond the issue consisting in establishing rigorous indicators, lies the measurement constraint. Many institutional measures rely on attempt to objectively assess institutional characteristic, essentially by focussing on formal aspects; informality being considered as subject to subjectivity. More generally, attempts to establish measures that are comparable across countries and across time, lead to focus on formality and simplifying assessment. This

raises the issue of biased measures. Biases car really matter because they may change across countries and accross time.[6]

For instance, indicators deal with legal compliance, legal enforcement and effectiveness only with difficulty. Several noticeable attempts to measure those fuzzy concepts try to go beyond the absence of a clear theoretical background. Thus, a first line draws on the activity and performance of courts and the judiciary system. For instance, the number of judicial decisions that are actually enforced provides a simple indicator of legal enforcement (See for instance Justice statistics at the domestic level; see also the comprehensive analysis of the European national systems carried out by the CEPEJ). Statistical information on amnesty legislation can also be useful. A second aspect of legal compliance is less documented. Indeed, legal compliance is concerned with other channels of enforcement that are less easily grasped. Namely, compliance is not only dependent upon the enforcement system, but also depends on market pressures, the self-enforcing nature of substantive law, social norms and culture (SIEMS [2005]). Therefore, some discrepancy may occur between the wording of the law and the actual practice. This point has been recognized widely by the economics analysis of law and several attempts to take "social compliance" into account have developed (Posner [2000]). However, we are far from benefitting of a satisfactory assessment of actual legal norms.

A closely related limitation is that most measures are flawed by the will to perform cross countries comparisons and therefore to measure domestic specificities. This raises two issues. First, universal descriptors may not be legitimate for all institutional settings and different institutions could not be compared on the basis of a limited set of criteria relevantly (GLENN [2001]; LEGRAND [1997]; PRZEWORSKI [2007]). Terminological ambiguities, cultural differences, the fact that rules and components of the judiciary do not perform identical functions in different institutional systems may not only affect the meaning of measures but

[6] At the empirical level, several studies provide us with telling examples for the complementary nature of formal and informal institutions. For instance, the mere existence of legal transplants and legal reforms has been shown insufficient to shape institutions. On the contrary, what really appears to matter is whether a legal order is already familiar with the basic principles of the transplanted law or not. Absent this "familiarity" with the transplanted institutions, a "transplant effect" may occur, resulting in a poor effectiveness of the imported legal order (Berkowitz & al. [2003]). In other words, the consequence of an initial mismatch between the pre-existing conditions and the transplanted law will be that nothing will ensure that the wording of the law – the way law is enforced within the society – will actually correspond to what is experimented by agents in their daily life. This parallels the requirement of "local knowledge" and "local experiment" preferably to "best-practice 'blueprints'" (Rodrik [2000, p. 3]).

also reduce the scope of comparisons (SIEMS [2005, P. 531]).[7] Second it can lead to an attempt to magnify contrasts among national characteristics, to the detriment of other, most relevant ones. For instance, in the area of law, most measures confine themselves to national law and, therefore, are in a way doomed to partial measurement, ignoring the strong movement of homogeneization of legal norms in many domains. Deriving from legal globalization and transnationalism and the development of legal networks, best practices, international codes and standards of public and private origin, legal arrangements pertaining to "soft law" have bloomed. Yet, « this complexity cannot be recognized by statistics, which simply focus on national rules and disregard public and private transnationalism » (SIEMS [2005, P. 530]). Several areas of law epitomize this remark, ranging from financial law — typically an area of law under global influence — to more strictly domestic law such as labour regulation.[8]

At the empirical level, several studies provide us with telling examples for the complementary nature of formal and informal institutions. For instance, the mere existence of legal transplants and legal reforms has been shown insufficient to shape institutions. On the contrary, what really appears to matter is whether a legal order is already familiar with the basic principles of the transplanted law or not. Absent this "familiarity" with the transplanted institutions, a "transplant effect" may occur, resulting in a poor effectiveness of the imported legal order (BERKOWITZ & AL. [2003]). In other words, the consequence of an initial mismatch between the pre-existing conditions and the transplanted law will be that nothing will ensure that the wording of the law – the way law is enforced within the society – will actually correspond to what is experimented by agents in their daily life. This parallels the requirement of "local knowledge" and "local experiment" preferably to "best-practice 'blueprints'" (RODRIK [2000, P. 3]).

2. The Severe Technical Requirements in matter of Measure and Indicators Building

Recognizing the similarity between the problems rose by the measure of "institutional capital" and the measure of "capital" (see the extended discussion in appendix) this section attempt to

---

[7]. This point provides a justification for developing measures based on a functionalist approach of institutions.

[8] However, one may not share the pessimistic diagnosis on statistics ability to include international and suprational arrangements as well, once their importance recognized. This more optimistic view relies on some ongoing attempts, such as those developed by the ILO recently

build upon the research developed to measure the stock of capital at the macroeconomic level to highlight the necessary conditions for building rigorous institutional indicators. Measuring institutions being a question, first, of aggregation of heterogeneous components, we discuss first aggregation techniques. This is followed by considerations on weighing methods. Lastly we get back on the measure of the aggregated variables.

The essential lesson to be drawn from establishing a parallel with the macroeconomic literature on the building of macroeconomic aggregates is that social scientists are facing a circular issue. Indeed, the building of satisfactory aggregates request well established theoretical relationship between the variables that are measured and the concept that is attempted to be measured (to test the quality of indicators). In the same time, measures are needed to test theories that are still developing, in particular on the phenomena of institutional complementarities, or on the relationship between institutional characteristics — design of rules, enforcement mechanisms, etc. — and 'intermediary' properties such as security of transaction, accountability of public decision makers, etc. The "circular" nature of the issue leads to highlight to path of development. First, micro and meso institutional analysis should be developed before any attempt to elaborate macro-institutional quantitative analysis. Indeed, aggregation problems are less tricky at micro-levels and the establishment of causal relationship is easier to control, in particular thanks to microeconometrics techniques (in the line of D. Levitt). The development of such analysis should allow to progressively develop the needed knowledge to build more satisfactory macro-institutional indicators. Second, techniques to "measure" institutional variables should be refined. We highlight in particular that agent's revealed preferences should be more systematically tracked to benefit from better measures.

2.1. Testing separability to check theoretical validity

Aggregated indicators may be informative if they assess the "good functional form" of the institutional bundle with relevant weights and a good dataset. As pointed out by KAUFMANN AND KRAAY (2007, PP. 3-4), they are usually justified on the following basis. First, they provide a useful summary of complex set or measures. This advantage can be relative, and should be assessed trough a tradeoff between synthesis of information and loss of conceptual precision. Second, they allow for a wider country coverage that may not be possible with individual indicators. Third, they contribute to the reduction of measurement errors and anomalies observed in individual indicators. The premise here is that each of the individual

data sources that are being used in the aggregation process provides an imperfect signal of the concept measured. Aggregating different sources thus allows for compensations among biases. It thus potentially allows to take information out of a vast array of governance data and indicators and to overcome some important limitations associated with the use of subjective indicators. The resulting composite indicators summarize the available evidence much better than individual indicators. Therefore, the former are frequently viewed as being more accurate, reliable and more useful for cross-country research than the latter. Last but not least, they should allow for sensitivity analysis and calculation of explicit margins of errors[9]. As pointed out by KNACK [2006, P. 4], however, "gains in statistical precision from aggregating data sources are far more modest than often claimed, because the assumption of independent error in measurement among data sources is violated".

Letting for the moment the quality of data aside, the question is how we can build, then test, the theoretical consistency of institutional composites? An essential result in matter of aggregation was provided by SONO [1945] and LEONTIEF [1947a; 1947b] more that sixty years ago. The Sono-Leontief's theorem provides the necessary and sufficient conditions for a twice differentiable function whose arguments are all non-negative, to be expressible as an aggregate. For a three order function, it states for instance that if $F(x_1, x_2, x_3)$ is continuously twice differentiable, then there exists two functions $h(x_1, x_2)$ and $G(h(x_1,x_2), x_3)$ such that $F(x_1, x_2, x_3)=G(h(x_1, x_2), x_3)$ if and only if F has the property $\frac{\partial(F_1/F_2)}{\partial x_3}=0$ where the subscripts denote differentiation**.**

To put it differently, this theorem states that an aggregation procedure is correct, and an overall construct valid, if and only if the marginal rates of substitution among variables in the aggregate are independent of the variables left out of it. It also means that the overall index should not change if one varies the respective contributions of two (institutional) dimensions in a same sub indicator, when leaving the value of the last one unchanged (AIZCORBE [1990]).

In the light of this theorem and due to institutional complexity, one can question the reliability of a large amount of institutional aggregates. Indeed, most institutional dimensions are interconnected trough various channels that are difficult to grasp as pointed out earlier in this

[9] This is the case at least for the WGI

paper. Most indicators, and in particular those of the World Bank on Governance, do not check at all for independence among variables aggregated in the same construct.

The Sono-Leontief's theorem provides the basis on which empirical tests of the separability of the functional form that can be grounded. A large number of empirical studies have thus checked the validity of diversified aggregates (amongst others HALAM AND POPE [1988]; CHAMBERS AND POPE [1991]; NAYGA AND CAPPS [1994], MAGNUS AND WOODLAND [1990]; BERNDT AND CHRISTENSEN [1974]; CHANG AND FRIEDLAENDER [1985]; BERNDT AND WOOD [1975]; CAPALBO AND DENNY [1986]; NORSWORTHY AND MALMQUIST [1983])[10].

Testing for strict separability is generally problematic because one need to implicitly place restrictions on the functional form of the aggregate when one chooses a functional form for the function (BLACKORBY AND AL. [1978]; DENNY AND FUSS [1977]). One implication is that, although the theoretical conditions represent a test of existence, their empirical counterparts actually test for the validity of whatever aggregate allowed by the underlying function of the functional form (AIZCORBE [1990])[11]. In order to avoid such stringency DENIS AND FUSS [1977] have introduced the notion of approximate separability. They relied on Taylor-series expansion to argue that if the functional form for the production function could be viewed as a second-order to an arbitrary function at a point. Then the choice of functional form has no implications for the possibly valid aggregate. According to this, separability tests would truly provide a test for existence, since no restrictions are placed on the form of the aggregate[12].

Each form of separability has its own testing design. Whatever the kind of test, a key point is that data on the overall performance or quality of the targeted institutional dimensions should be available, in addition to the measurement of the sub-indicators components. Indeed, the properties of the functional form can only be assessed if one has data on inputs and outputs of the functional form. This is one of the reason subjective measurement may be relied upon to perform such test in the context of institutions, provided that the ability to assess the overall performance of an institutional framework on a relevant basis (which may raise issues when

[10] Arthur Lewbel has recently proposed a method in order to check the validity of aggregates in demand analysis (LEWBEL [1996]). One can nevertheless wonder if his method could be implemented on institutional aggregates.

[11] For example, DENNY AND FUSS [1977] demonstrated that the translog only allows two types of aggregates, one log-linear in its arguments and the other quadratic in the logs.

[12] This point has been contested by WHITE [1980]. From his view, the functional form, once estimated, does not display properties normally exhibited by a Taylor-series expansion.

reasoning at an aggregate level due to the multi-purpose nature of institutions)[13]. This point will be discussed later. The central lesson at this point is that it is impossible to build an aggregate indicator without checking the marginal effects of each measured institutional dimension in the institutional aggregate. This request a good conceptualization of the way institutional characteristics impact on the institutional dimensions one attempt to measure. Moreover, one need to identify an independent variable that should be correlated with the construct, which request also a solid conceptualization, and therefore a well established theory. To illustrate, if one tries to build an indicator of the level of democratic governance in a country, one need to establish for instance a relationship between democracy and individual freedom to check whether the aggregate "leved of democracy" — built by aggregation of various measures of turnover in the various level of government, transparency of elections, degree of accountability of decision makers, etc. — is accurately computed. This supposes a significant level of maturity of the theory.

2.2. Weighing Methods

Testing for separability is a way to assess the validity of a functional form. The reliability of composites depends also upon the quantitative assesments of the contribution of each proxies. Carefull attention is thus needed to establish the proper weights for each dimensions introduced in the construct. The Sono-Leontief condition is concerned by the way institutional dimensions interact, but does not hold for the relative importance of two institutional effects. Computing the "good" weights for each components of the composite is then a second step.

Scholars have been paying a lot of attention to weighting strategies, and various techniques have been developed to establish the proper weights to build indexes. The currents methodologies do not have, however, the same reliability. Their capacity to provide good estimators is directly related to their ability to cope with a double tradeoff. First, as for measurement strategy, a choice has to be made between subjective and objective methods. Subjective methods are applied when the relative importance of the measures contributing to the composite is decided on the basis of personal opinion (which encompass the case where it is decided that all variables have the same weigh). Subjectivity can be due either to the unilateral choice of the author of the indicator, or to the one of a population of individuals that

[13] Such as those provided by LEE AND MANSFIELD [1996] for intellectual property protection for instance.

is surveyed to assess the relative importance of various institutional components. In any case, there are strong risks of bias due to insufficient theoretical grounding of the relative influence of the various institutional dimensions on the analytical construct. Strong criticisms in that sense have been addressed to most indicators: the *Doing Business* reports, the World Economic Forum competitiveness indicators, the World Bank's CPIA are criticized on the motive that their aggregation methods are based on choices hardly justifiable and barely justified[14]. The alternative is to compute weights by 'letting the data speak' (i.e. data analysis). The main advantage of this option is to produce weights that fit to the statistical properties of the data. Lack of theoretical background can however be problematic, because it is then difficult to identify precisely what is measured, leaving room for misinterpretation of the institutional dimensions assessed. That said, theoretical formalism should be preferred as it allows replicability of the index, and as it enable the community to see the theoretical limitation of the computed weights.

The second tradeoff is between empirical and theoretical grounding of the weighing. This is largely a 'dog and tail' issue since, by the end, weighs should be empirically estimated, while estimations suppose a well established theory.

Table 1 sum up this discussion. While the ideal would be to rely on methods in the south-east box of table 1, in practice one rely often either on the south-west methodologies that strongly depends upon the quality of data and raise issue of (theoretical) interpretation, or on the north-west type of methodology that question the relevancy of relying on individual opinion to decide an analytical issue (see table 1)[15].

---

[14] For instance, the CPIA, which consists of a set of criteria representing the different policy and institutional dimensions of an effective poverty reduction and growth strategy, utilizes 20 criteria grouped in 4 clusters, respectively i) economic management ii) structural policies iii) policies for social inclusion and equity iv) public sector management and institutions. Each of these four clusters is affected a 25% weight in the overall rating and within each cluster, all criteria receive equal weight, although components within a criterion may be weighted differently. The overall score is obtained by calculating the average score for each cluster and by averaging the scores of the four clusters. Yet, these weights are not given any serious justification in the general presentation of the CPIA rating. Worse, the weights used to build the aggregate indicators may even be unknown or impossible to find out for some indicators. Then, "[W]ith broader, multi-dimensional indicators such as ICRG, data users have no way of knowing exactly what the indicators are even attempting to measure.[…] This uncertainty problem is exacerbated for other corruption indicators for which no such criteria are made public at all, as is the case for corruption measures produced by two competitors of the ICRG: the Economics Intelligence Unit (EIU) and World Markets Research Centre (WMRC)." (KNACK [2006, PP. 13-14]).

[15] Several points should be noticed here. First the fact that non statistical weights are badly justified does not mean that there is no theoretical background. Even equal weighting such as in the CPIA indexes reveal some beliefs about the relative importance of the various sub-indicators. Moreover the theoretical and objective assessments rely on two distinct types of models. Regressions models and discriminatory measures quantify the relative effect of each policy action on the output, i.e. on the single indicator). The need of a "dependent variable" (here not in the form of a composite indicator) that accurately

Table 1. Tradeoffs in weightings methodologies[16]

	Empirical basis	Theoretical basis
Subjective Aggregation logic	○ Conjoint analysis ○ Public opinion ○ Analytic hierarchy process ○ Budget allocation	○ Non statistical assessment by the author
Objective Aggregation logic	○ Principal components and Factor analysis ○ Canonical correlations ○ Benefit of doubt Approach	○ Classic and multiple regression models ○ Unobserved component models ○ Discriminatory analysis

2.3. Measuring strategy: Avoiding biases in data collection

Up to the present, the (lack of) accuracy and credibility of primary source material has been a main subject of discussion around institutional indicators and measurement. In particular, the tendency to reuse data produced for other purposes has to over-reliance on a few source materials (BROUSSEAU AND AL. [2007]; VRIES [2007])[17]. This has been exacerbated by a tendency of IGOs to collate material from a number of sources and reproduce it uncritically within their own analysis, effectively giving it another layer of credence[18]. The recent passion for institutional assessment has been leading the research community to develop a large number of dataset (MALIK [2002]; UNITED NATIONS [2007]) and to put on the scholars' agenda the necessity to develop efforts to produce more "reliable" data.

The quality of a dataset should not be reduced to its technical property-such as its accuracy or its scope and country coverage-but should rather be assessed in terms of the issue that is analyzed. In this view, the quality of a dataset therefore refers mostly to its "fitness for use"

---

and satisfactorily measures the target in question is removed when one use unobserved components as it manage it like another unknown variable to estimate (see KAUFMANN AND AL. [1999]). Last but not least, discriminatory measures are the only ones to cope with weighting and aggregation in a simultaneous way (SATTIN AND BROUSSEAU [2007]).

[16] The choice of the aggregation rule should be in line with the chosen weighting methodology. For instance, benefice of doubt approaches require linear aggregation.

[17] This has been particularly true in some areas like the measurement of human rights and good governance. See the place for instance of by Amnesty International, Freedom House and the US Department of State as datasources.

[18] For instance, such reciprocal borrowings between sources can be expected from the use of World Bank data (Doing Business data) by other (most often policy-oriented) institutions to ground their decision and the parallel use of these other institutions by the World Bank to triangulate its own results. The scarcity of reliable data encourages such circular borrowings within the community but, thereby, may introduce biases into the analysis.

(UNITED NATIONS [2007]; OECD [2005])[19], such a concept being operationalized by a set of meta-concepts such as relevancy, credibility, coherence, etc. (OECD [2005]; IMF [2002]). If these guidelines can be helpful in order to assess the ex-post quality of the collected data, they give no information about the comparative advantages and the specific pitfalls of the various methodologies available in order to measure institutional phenomena.

The choice of a measurement strategy is mainly about the degree of subjectivity of the collected data. In regards of the future uses of the data, one should thus assess if the biases associated with each methodology could affect the reliability of its results[20]. Before discussing in details the issues rose by subjective measures, and the way to address them, it is worth to note that measurement standards tend to plead for subjective methods of assessment of institutional characteristics and outcomes vis-à-vis objective description or characteristics. This is indeed a consequence of the complexity of institutional systems, of the complementarities among institutional mechanisms, and of the interplay between formal and informal aspects of institutions. As pointed out in the first part of this paper, 'universal descriptors' tend to fail because various combinations of rules and enforcement mechanisms of different kind can result in similar frameworks for behaviors. Subjective assessment may therefore enables better than 'objective descriptions' to describe and measure how institutions actually perform and frame behaviors. It is however necessary to distinguish the subjectivity of the 'measurer' from the subjectivity of the surveyed individuals, whose rationality is relied upon to turn complex qualitative phenomena in reliable measures. As we will discuss in the following pages, the important issue is to be able to produce measures that are not biased by the default of rationality and the interest of the "experts" which are in charge of summing up their understanding of institutions in numbers or characteristics. Hence the usefulness of approach based on revealed preferences and practices.

One should notice that the subjectivity is not a problem on its own, but rather that it opens the door to biases and measurement errors that can drive to some serious concerns[21]. This is

---

[19] Some authors have attempted to define the attributes of a good dataset by explicating what should be its intrinsic properties, such as relevance, accuracy, credibility, etc. On this point see OECD [2005], NARDO AND AL. [2004] and MUNCK AND VERKUILEN [2002].

[20] Cost efficiency is also an important aspect of data collection. OECD [2005] considers that it as a complementary aspect of "good" quality that can have an impact on the other characteristics defined above.

[21] The wide array of possible biases associated with subjective indicators could deter to use them, and one can ask if the objective route is easier. Measurement is objective when different people do interpret the information in the same way. Objective indexes generally rely on various sources of information to study the provision of a given institutional good

especially the case when one considers the most relied upon method to compute institutional assessment: panels of experts.

First of all, the information on which experts base their assessment may be poorly reliable. For instance, experts may have access only to inaccurate or flawed sources, to their only partial experience, or to official statistics and documents that are poorly reliable or incomplete. In addition, experts do not systematically access first-hand knowledge about the aspect of the system they are called to assess. For example, an expert may be asked to assess the judicial system of a country where he never operated. His or her opinion reflects then the view of an external observer[22]. Critics insist then on the idea that institutional specificity and internal logic of a system are often neglected with such a methodology.

Second, experts may reflect particular interests and biases. This covers window dressing practices — for instance when governmental experts are requested to assess the quality of their national institutions —, and rent-seeking behaviors, for instance when entrepreneurs and firm managers, or other interest-groups, answer a surveys to favor implementation of specific policies (e.g. calling for a reduction of the costs of procedures in the spirit of the *Doing Business* reports). Many indicators, as those available in the Global Competitiveness Report by the World Economic Forum or the Executive Opinion Survey are based on samples of executives — preferably with international experience — from large exporting firms, who can have a very specific or partial vision, to say the least. An additional problem lies in that evaluation by highly skilled experts may create an illusion of objectivity and precision, while in fact their replies are opinions influenced by specific interests.

Lack of information and conflict of interest do not only induce measurement errors. It can lead to strong, self-reinforcing effects in the biases. High cost of information and assessment

---

(BROUSSEAU AND SATTIN [2007]). Measuring law, for instance, can rely on the analysis of legal documents, acts, or reports. The main advantage of objective indicators is that what is actually measured is clearer and enhances the relevancy of inter-countries comparisons. But they also suffer from some important drawbacks. First a skilled staff is generally needed to compute these indicators and it leads to high production costs. Moreover, for laws and governance programs, the construction and exploitation of objective indicators may not always be fully feasible. Finally for some institutional dimensions such as corruption, it may be uneasy to find alternatives to perceptions data. It is doubtful that they enable to avoid measuring superficial (possibly formal) institutional characteristics while not capturing the underlying processes at work. One should remind that institutions are operationalized through the beliefs of economic agents and thus as noted above perception matter. But absent a well defined theoretical framework that encompasses the institutional specificities of each country, there an increased risk to miss some important characteristics of specific institutional environment under analysis. Last but not least and as quoted above, testing separability with objective assessments turns to be problematic.

[22] For instance, in the case of the World Bank's CPIA (Country Policy and Institutional Assessment) or in the case of the WGI, the gathering of data is perception-based and there is no direct observation by actors involved in the system.

car lead expert to refer to the assessments of other experts to ground their own opinion, resulting in Asch effects (ASCH [1951]). This might also lead to "halo effect" (THORDIKE [1920]): expert's perceptions being wrongly affected by economic conditions, excessive optimism or pessimism over the economic, political, and social conditions. The corollary of this is that changes in expert assessments may reflect corrections in their judgment rather than effective evolutions. Also, time may be necessary to lead experts to perceive or recognize evolutions. Thus, improvements or deteriorations of a situation may actually occur and not be traceable through subjective perceptions. The stability of an indicator might thus correspond either to a stable situation or to an inability to recognize changes. Also, in some situations the perceptions of change can be hindered by the evolution of preferences. If the respondents' expectations grow at the same rate as a system's capacity, change will not be identified.

Users and public opinion surveys are another way of subjective assessment but they have also important drawbacks. In particular, laymen opinions are really problematic since expressed satisfaction might be strongly impacted by some cultural bias. Moreover, a strong discrepancy is often observed between opinion and actual performances, since many factors are likely to affect perception. Furthermore, the rent-seeking bias already observed in experts' assessments might also affect users' assessments as respondents might have an incentive to bias their responses when they expect this might lead to favorable policies[23] [24].

An interesting way to cope with the limitations of the traditional methodologies may be to rely on the actual choices of the agents in order to build the institutional assessments. Actually, as this method relies on an ex-post analysis, it does not suffer from the weaknesses of the traditional measurements. Indeed, it can be viewed as a direct application of the revealed preference theory in order to build institutional indicators, and as such, it should enable the practitioner to reach the best estimation of the agent's actual preferences.

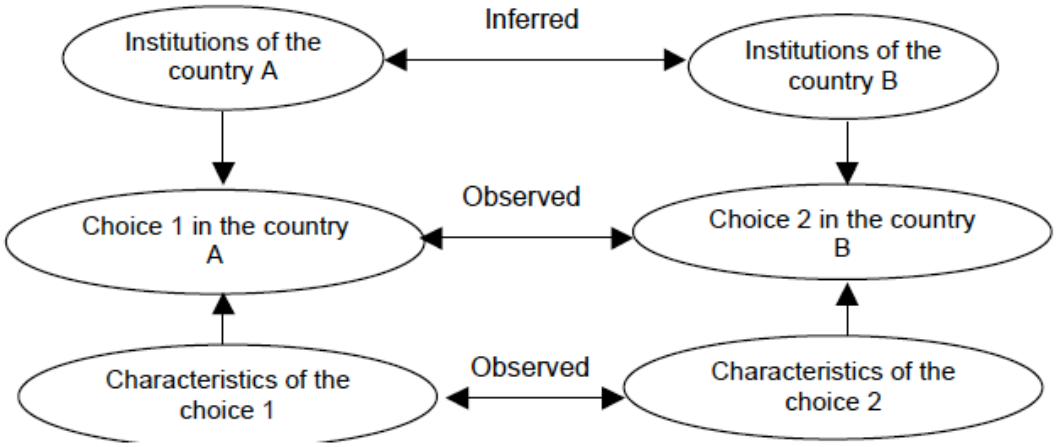
The basic principle of the approach is to assess the institutional differences among several countries (or any other variable identifying institutional boundaries) as residuals. This

[23] BEEPS (Business Environment and Enterprise Performance Survey), conceived as a survey of service users, may face such limitations.

[24] Lastly, a common bias to both experts and users' surveys is the confirmation bias that may affect databases that are being gradually completed. In that case, there is a risk that the authors favor new information confirming their former analyses and, therefore, produce data consistent with their previous results.

information is inferred from the observation of the agents' actual choices in the various environments and from the characteristics of these choices (see Figure 1).

Figure 1: The Comparison of Institutional Frameworks



In order to grasp the institutional effect, this methodology requires only a collection of data on some individual choices that react to the institutional background, at least for two countries. The data may be homogeneous or heterogeneous, in the sense that each observed action may also be conditioned by some other non-institutional factors. The first case is the trivial one, where the institutional impact can be grasped directly from the difference of the contractual choices' frequencies between the two countries. Unfortunately, this kind of "pure" situation is very unlikely to happen, because, in the real world, the final choice of the agents depends also of some individual or contextual characteristics besides the institutional effect. In these cases, a statistical correction is needed in order to take into account the characteristics of the transaction in the estimations, and thus to infer the institutional impact from the observed choices.

Thus, to assess the institutional effect, we have to cope with the differences in the characteristics of the choices between the various data sets. This problem has been solved in labor econometrics since OAXACA's pioneer work (OAXACA [1973]), in order to assess the racial or sexual discrimination. The first step is to estimate on each sample a statistical model that links the expressed choices to the individual or contextual characteristics. As we do not incorporate any variable that grasps the institutional factors, the models are biased and the institutional effects are reflected by changes in the coefficients among the estimations. These regressions enable us to know how the choices that are expressed in a country would have been affected if they had arisen in another institutional environment. Finally, with this

information, we are able to replicate the choices actually expressed in a county in different institutional frameworks. Then, as we compare two countries, it is possible to distinguish the part of the variation between the average expressed choices that comes from institutional differences, from the part that is due to changes in the characteristics of the choices.

A central feature of this approach is that the results change with the country of reference. Indeed, the different standardizations reflect the familiar index-number problem encountered whenever heterogeneous collections of goods are summed with two sets of prices (CAIN [1986]). Then, the final result is more an interval of variation that an indicator. Nevertheless, this decomposition method may return very precise information, and thus is still widely used in many contexts. Indeed, beside discrimination analysis, this methodology is the normal way to compute price indexes (Lapeyre and Paashe) or cost gap in accounting.

The possible applications of this approach in institutional analysis are numerous. For instance, it may be used to quantify the impact of the law on the criminal behavior by means of inter-countries comparisons. It may also be implemented onto governance structures to quantify the impact of the institutions that surround contracts. Brousseau and Sattin [2007] constitutes an attempt to apply this methodology on a sample of technology licensing agreements to compare the institutions that secure technology transfer in France, Germany and the US. The drawback of this methodology is of course that it is (micro-econometric) data demanding.

3. Conclusion

Despite a growing effort oriented towards increased reliability, institutional measures still suffers from a lack of theoretical conceptualization. As such, they fail to take into account some fundamental features and dimensions of institutions and to model properly the institutional interractions. This paper aims to assess of the current practice, and to make some proposition onn what should be done to increase the validity of institutional composites. Tests of separability are one way to discriminate between alternative indexes that can be proposed. But in order to be implemented properly , they need to rely both on reliable data as well as on well-designeds weighting scheme.

Perspectives for future researches are twofold. This work is obviously a call to generate new datasets and to use existing one more than carefully both in scientific research and in policy making. Methodological progress and refinined testing procedures are before all needed in

order to contribute to a better understanding of the causal relationship to be tested, especially to identify the institutional featrues to be measured. This is for instance the case when dealing with institutional innovations. Indeed, so far, most indicators implicitly rely on the disputable assumption that there exists a given and finite list of institutional arrangements among which to choose. This amounts not only to assume the intertemporal comparability of institutions but this is also to disregard the dynamics of institutions and dismiss any attempt to measure the capacity of adaptability and innovation of a legal system. From that standpoint, beyond some summary but insightful measures aimed to count legal and institutional reforms, new rules, and case reversals, few attempts have been carried out so far (PISTOR AND AL. [2003]).

Moreover the purpose here was not to plead for one silver bullet indicator that would be able to encompass the institutional complexity of a legal system and to disentangle institutional interractions and subtleties. In view of the state of the art, such a claim sounds clearly unrealistic. One should rather accept that only some dimensions and components of institutions could be measured. A corollary is that available indicators should be published and used with explicit and systematic acknowledgement of their limits in their attempt to measure. This would forestall most of critics against indicators and facilitate the interpretation of the measure.

Appendix: Building institutional indicators: what can be learn from the history of macro-production functions?

Despite a wide array of theoretical evidence showing that the conditions required to compute macro aggregates (such as consumption, capital and so forth…) are seldom encountered, large parts of growth theory still rely on aggregate production functions that are not theoretically relevant. Why?

As we will show, the answer to this paradox lies mainly in the specific dynamic that drove academic research in this field. As such, the analysis of the debates on the internal consistency of aggregates included in macro-production functions can be useful to understand the challenges that surround today the building of institutional indicators.

Amongst aggregation problems, those related to capital have deserved a special attention. While debates about the very nature of capital go back to the 19th century (with the works of K. Marx, T. Veblen, JB. Clark, and E. Fisher), a revival arose with the complaint of Joan Robinson in the 1950th. The question that triggered the debate was the unit in which the capital should be measured. Joan Robinson and her contenders, following Wicksell, argued that there was no 'natural unit' to assess the volume of accumulated capital in an economy. While labour and land could be compiled respectively in man-hours and acres, it was difficult to assess a stock of heterogeneous capital goods. The only way to escape the problem was to rely on the prices of capital in the macro-production function. However, that was exactly what the model was trying to derive from the marginal productivity of the aggregate. As such it was impossible to get any evaluation of capital as a measurable quantity independent of prices and therefore of distribution. To put it differently, aggregate capital, aggregate production function and the marginal productivity of production factors can only be defined when the rate of profits is given, and this implies that they cannot be relied upon to build a theory of the rate of profit or distribution (Felipe and Fisher, 2001). Moreover, instead of being well shaped, the production function with heterogeneous capital goods could display complex shapes with reswitching and capital reversing phenomena (Cohen and Harcourt, 2003).

Joan Robinson's complaint led to what is known now as the "Cambridge vs. Cambridge controversy". Twenty years of debates on these topics generated at least three kinds of solutions to aggregate capital and generate aggregate production functions (Felipe and Fisher, 2001).

A first solution is to define the conditions under which aggregation is possible. This results in a a set of theorems that deal with aggregation of goods (the Sono-Leontief theorem) or of firms (the Nataf, Gorman, Fisher, and Sato theorems). The point here is that the conditions they define are so stringent that it is difficult to imagine they could be matched in the real world. As a consequence, they cast doubts over the validity of most of the results in macroeconomics (think about the assumptions of endogenous growth models).

A second stream of researches has looked for theoretical refinements of the production function and was no more successful than the former. First of all, Swan (1956) introduced the concept of putty capital in order to merge heterogeneous goods into a simple malleable good. Solow (1963) attempted to bypass the problem of the valuation of capital by focusing on the rate of return on investment, understood as increases of capital stock. Finally, Samuelson (1962) presented the one commodity model as a parable and tried to integrate different kinds of commodities in the macro-production function (the surrogate production function); but with a set of such stringent conditions that he finally had to set back to the one commodity function. These various theoretical attempts only hid the problem without solving it. In 1966 Samuelson had to acknowledge that the conclusions of the neoclassical models couldn't be universally valid, and that reswitching and capital reversing should be common in the real world. Subsequent attempts from neoclassic economics included recourse to general equilibrium with Bliss and Hahn, but with no more success (Cohen et Harcourt, 2003).

The results of the "Cambridge vs. Cambridge" controversy and of the issues on aggregation have thus highlighted the lack of theoretical foundation of macro aggregate and of aggregate production functions. These issues seem however largely neglected today and, when one consider the current state of the art in macroeconomics it is difficult to believe that these debates have ever happened! Why? As Cohen and Harcourt point out, if the parties agreed on the nature of the problems, they differed on its empirical importance. This led the debate to become "ideological", partly explaining the de facto victory of the stronger academic influence of the American neoclassic economists. Moreover, Joan Robinson's contenders failed to produce both an alternative framework and reliable empirical works to ground their position.

On it side, even if it was poorly theoretically grounded, the aggregate production function, fitted quite well the macroeconomic data on growth. Solow (1956) was the first to justify the macro-production function on the empirical ground thanks to its capacity to capture the

essential features of the growth process as they draw from national accounting efforts. Then, why care should be taken of aggregation concerns if the empirical fits are good? As pointed out by Felipe and Fisher (2001), such a position is pure instrumentalism that cannot be defended from a methodological point of view. Moreover, such approach does not recognize the very nature of the technology and as such could be misleading. For instance, Fisher (1971) note that the shares of production factors are constant in developed economies not because technology could be represented by a Cobb Douglas, but rather because this functional form works empirically since, for unexplained reasons, the share of production factors remains fixed in most of the firms. Another argument in favor of reliance on aggregate production functions was to pretend that there was no other option. This is a variant of the instrumentalist position and, as such, it is not acceptable especially when one remembers that the whole meaning of macro aggregate is questionable.

Nevertheless, for reasons of pure convenience, the literature on aggregation ended by being ignored, and the neoclassical model became prevalent. Endogenous forces in academic research, such as social isomorphism and institutional pressures seem to have put an end to these controversies. A lot of young scholars are not even aware of this of literature, and it is now difficult to publish in this field without relying on the standard assumptions of the neoclassical model (Lavoie, 1992).

What lessons can be drawn from these developments when building institutional indicators? First, it seems interesting to develop the analogy between institutional and capital aggregates to define the conditions of validity of institutional indexes. Aggregation and measurement problems arise in both cases. Usher (1981), for instance, gives four contingent definitions of capital, and the 'natural unit' of institutional aggregates is not so obvious (if we except Djankov's Doing Business). But on the technical ground one should also take care of the limits of the analogy because there are some noticeable differences between capital and institutions. The aggregation literature was concerned by aggregations of goods and of firms' production functions. Institutions are collective phenomena and, as such, one does not need to sum them at the firm level to assess the institutional stock in a country. However institutional components have cumulative and combined and, as we have noticed in the paper, the way this should be computed has rarely been checked when building the existing institutional indexes. Moreover, as noticed above, a large part of the debate was related to repartition problems and derived from the economy seen as an economic circuit. So macroeconomic indicators were built and discuss following a logic that is far from the one characterizing institutional issues.

Nevertheless, it is possible that main message sent by this part of economic history relies on another ground. Indeed the case show us how academic isomorphism can sometimes validate procedures that are theoretically wrong when they are convenient to use and when their promoters have a large academic audience. As such it prompts us to define quickly the conditions under which institutional aggregate could exist and warns us against “easy institutional macroeconomics” contenders.

References

ACEMOGLU, D. AND S. JOHNSON [2005], "Unbundling Institutions," *Journal of Political Institutions*, 113, 5, 949-995.

ACEMOGLU, D., JOHNSON, S. AND J.A. ROBINSON [2002], Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution, "*Quarterly Journal of Economics,* 117, 1231-1294.

AIZCORBE, A. [1990], "Testing the Validity of Aggregates, " *Journal of Business & Economic Statistics*, 8, 4, 373-383

AOKI, M. [2001], *Toward a Comparative Institutional Analysis*, MIT Press.

ARRUÑADA, B. [2007], "Pitfalls to Avoid when Measuring the Institutional Environment: Is 'Doing Business' Damaging Business?," *Journal of Comparative Economics*, 35, 4, 729-47.

ASCH, S.E. [1951], "Effects of group pressure upon the modification and distortion of judgement, " In H. Guetzkow (ed.) *Groups, leadership and men*, Pittsburgh, PA: Carnegie Press

BARDHAN, P. [2004], "History, Institutions and Underdevelopment," Department of Economics, University of California, Berkeley.; Available at: http://emlab.berkeley.edu/users/webfac/bardhan/papers/Bardhan_Scarcity_Ch1.pdf

BERKOWITZ D., PISTOR, K. AND J-F. RICHARD, [2003], "The Transplant Effect," *American Journal of Comparative Law*, 51, 2, 163-203.

BERNDT, E. R. AND L. CHRISTENSEN [1974], "Testing for the Existence of a Consistent Aggregate Index of Labor Inputs," *American Economic Review*, 64, 391-403.

BERNDT, E.R. AND D. WOOD [1975], "Technology, Prices and the Derived Demand for Energy," *Review of Economics and Statistics*, 57, 259-268

BLALOCK, H.M., [1971], "Aggregation and Measurement Error, " *Social Forces*, 50, 2, 151-165

BOLLEN, K.A., [2001], "Indicator: Methodology," in Smeler N. and P. Balfes (eds), *International Encyclopedia of the Social and Behavioural Sciences*, vol. 11, Amsterdam, Elsevier, pp. 7282-7287.

BROUSSEAU, E AND J-F SATTIN [2007] "Les Contrats pour Evaluer le Droit: Quelques Propositions Méthodologiques Fondées sur l'Analyse du Droit des Brevets et des Contrats de Licence de Technologie, " *Revue Economique*, 58, 6, 1363-1386.

BROUSSEAU, E, S. HARNAY AND J-F SATTIN, [2007], *Measuring Law and Institutions: A Critical Survey of Current Practices and Some Recommendations to Build Indicators*, Rapport pour le groupement Attractivité Economique du Droit.

BUSCAGLIA E.AND P. DOMINGO, [1996], "Political and Economic Impediments to Judicial Reform in Latin America," in *Law and Economics of Development*, E. Buscaglia, W. Ratliff, R. Cooter (eds), JAI Press.

CAIN, G.G. [1986], "The Economic Analysis of Labor Market Discrimination: A Survey," pp. 693-785 in: O. Ashenfelter and R. Layard (eds), *Handbook of Labor Economics*, Vol. 1, Elsevier Science Publishers: North Holland.

CAPALBO, S.M., AND M.G. DENNY [1986], "Testing Long-Run Productivity Models for the Canadian and U.S. Agricultural Sectors," *American Journal of Agricultural Economics*, 68, 615-625.

CHAMBERS, R.G. AND R.D. POPE [1991], "Testing for Consistent Aggregation, M.G. *American Journal of Agricultural Economics*, 73, 3, 808-818

COHEN, A. AND G.C. HARCOURT [2003], "Whatever Happened to the Cambridge Capital Theory Controversies?,"*Journal of Economic Perspectives*, 17, 1, 199–214.

DE LONG, J. B. AND A. SHLEIFER [1993], "Princes and Merchants : European City Growth Before the Industrial Revolution , " *Journal of Law and Economics*, 36, 671-702.
DE SOTO, H. [1989] *The Other Path*, New York, Harper and Row.
DE SOTO, H. [1996], "Property Rights and economic Progress in Developing Countries: The Peruvian Experience, " in *Law and Economics of Development*, E. Buscaglia, W. Ratliff, r; Cooter (eds), JAI Press.
DENNY, M. AND M. FUSS [1977] "The Use of Approximation Analysis to Test for Separability and the Existence of Consistent Aggregates," American Economic Review, 67, 404-418.
FELIPE, J. AND FISHER, F.M. [2003], "Aggregation in Production Functions: What Applied Economists Should Know," *Metroeconomica*, 54, 208-262.
Fisher, F., 1971, Aggregate production functions and the explanation of wages: A simulation experiment, The Review of Economics and Statistics, 3, 4, 305-325
FISHER, F.M. [1969], "The Existence of Aggregate Production Functions," Econometrica, 37, 4, 553-577
GLENN, H.P. [2001], "Are Legal Traditions Incommensurable?," *American Journal of Comparative Law*, 49, 1., 133-145.
GREIF, A. [2000], "The Fundamental Problem of exchange : A Research Agenda in Historical Institutional Analysis, " *European Review of Economic History*, 4, 3, 251-284.
HALL, R.E. AND C.I. JONES [1999] "Why Do Some Countries Produce So Much More Output Per Worker Than Others?," *The Quarterly Journal of Economics*, 114, 183-116
IJIRI, Y. [1971], "Fundamental Queries in Aggregation," *Journal of the American Statistical Association*, 66, 336, 766- 782
IMF [2002], *Globalization: Threat or Opportunity;* Available at : http://www.imf.org/external/np/exr/ib/2000/041200.html
KAHAN , D.M. [2000], "Gentle Nudges vs. Hard Shoves: Solving the Sticky Norms Problem, " *University of Chicago Law Review*, 67, 3, 607-645.
KAUFMANN, D. AND A. KRAAY [2007], "On Measuring Governance: Framing Issues for Debate," MPRA Paper 8187, University Library of Munich, Germany
KNACK, S. [2006], "Measuring Corruption in Eastern Europe and Central Asia: A Critique of the Cross-Country Indicators," World Bank Policy Research Working Paper 3968.
KNACK, S., KUGLER., M AND N. MANNING [2003], "Second-generation Governance Indicators," *International Review of Administrative Sciences*, 69, 03, 345-364.
KURTZ, M.J. AND A. SCHRANK, [2007], "Growth and Governance: Models, Measures, and Mechanisms, " *The Journal of Politics*, 69, 2, 538-554.
LA PORTA R., LOPEZ-DE-SILANES, F., SHLEIFER, A. AND R. VISHNY [2002], "Investor Protection a,nd Corporate Valuation, " *Journal of Finance*, 57, 3, 1147-1170.
LA PORTA, R., LOPEZ-DE-SILANES, F., SHLEIFER, A. AND R. VISHNY [1997], "Legal Determinants of External Finance", *Journal of Finance*, 52, 3, 1131-1150.
LA PORTA, R., LOPEZ-DE-SILANES, F., SHLEIFER, A. AND R. VISHNY [1998], "Law and Finance," *Journal of Political Economy*, 106, 6, pp. 1113-1155.
LA PORTA, R., LOPEZ-DE-SILANES, F., SHLEIFER, A. AND R. VISHNY [1999] "Corporate Ownership Around the World, " *Journal of Finance*, 1999, 54, 2, 471-517.
Lavoie, M., 1992, *The Foundations of Post-keynesian Economic Analysis*, Edwar Elgar
LEE, J-Y. AND E. MANSFIELD [1996], "Intellectual Property Protection and U.S. Foreign Direct Investment," *The Review of Economics and Statistics*, 78, 181-186.
LEE, S. AND D. MCCANN [2006], "Measuring Working Time Laws: Texts, Observance, and Effective Regulation, " University of Melbourne Centre for Employment and Labour relations, *Law Labour Law Seminar Series*.
LEGRAND, P. [1997], "The Impossibility of Legal Transplants," *Maastricht Journal of*

European and Comparative Law, 4, pp. 11-124.
LEONTIEF, W.W. [1947A], "Introduction to a Theory of the Internal Structure of Functional Relationships," *Econometrica*, 15, 4, 361-73.
LEONTIEF, W.W., [1947B], "A Note on the Interrelationship of Subsets of Independent Variables of a Continuous Function with Continuous First Derivatives", *Bulletin of the American Mathematical Society*, 53, 343-50.
LEWBEL, A. [1996] "Aggregation Without Separability: A Generalized Composite Commodity", The American Economic Review, 86, 3, 524-543
MAGNUS J.R. AND A. D. WOODLAND [1990], "Separability and Aggregation", *Economica*, New Series, 57, 226, 239-247.
MALIK, A. [2002], "State of the Art in Governance Indicators," Occasional Paper. Human Development Report Office.
MOSCHINI, G., MORO, D., AND R.D. GREEN [1994], "Maintaining and Testing Separability in Demand Systems," *American Journal of Agricultural Economics*, 76,1, 1994, 61-73.
MUNCK, J.L AND J. VERKUILEN [2002], "Conceptualizing and evaluating democracy, " *Comparative Political Studies*, 35, 2, 5-34.
NARDO, M., SAISANA, M., SALTELLI A. AND S. TARANTOLA, A. HOFFMAN AND E. GIOVANNINI, [2004], *Handbook On Constructing Composite Indicators: Methodology And User Guide*, OECD Statistics Working Paper.
NORSWORTHY, J.R., AND D. MALMQUIST [1983], "Input Measurement and Productivity Growth in Japanese and U.S. Manufacturing," *American Economic Review*, 73, 947-967.
NORTH, D. [1991], "Institutions," *Journal of Economic Perspectives*, 5, 1, 97-112.
OAXACA, R.L. [1973], "Male-female wage differentials in urban labor markets," *International Economic Review*, 14,693-709.
OECD [2004], *2004 Employment Outlook* ; Available at : http://www.oecd.org/dataoecd/8/4/34846856.pdf
OECD [2005], *Management in Government: Feasibility report on the Development of Comparative Data;* Available at : http://appli1.oecd.org/olis/2005doc.nsf/linkto/GOV-PGC(2005)10
OLSON, M., [1993], "Dictatorship, Democracy, and Development, "*American Political Science Review,* 87, 567-76.
OLSON, M., [2000] *Power and Prosperity: Outgrowing Communist and Capitalist Dictatorships*, Oxford, Oxford University Press.
PESARAN, M.H., PIERSE, R.G., AND M.S. KUMAR [1989], "Econometric Analysis of Aggregation in the Context of Linear Prediction Models", *Econometrica*, 57, 4, 861-888.
PISTOR, K., KEINAN, Y. , KLEINHEISTERKAMP J. AND M. WEST, [2003], "Innovation in Corporate Law", *Journal of Comparative Economics*, 31, 4, 676-694
POPE, R.D. AND A. HALLAM [1988], Separability Testing in Production Economics, *American Journal of Agricultural Economics*, 70, 1, 142-152.
POSNER, E. [2000], *Law and Social Norms*, Harvard University Press.
PRZEWORSKI, A. [2004], "Some Historical, Theoretical, and Methodological Issues in Identifying Effects of Political Institutions ".; Available at: http://www.econ.nyu.edu/cvstarr/seminars/archive/2004_fall/africanaintl1012_przeworski.pdf
PRZEWORSKI, A. [2007], "Is the Science of Comparative Politics Possible?," In Carles Boix and Susan C. Stokes (eds.), *Oxford Handbook of Comparative Politics*. New York: Oxford University Press.
RODRIK, D. [2000], "Institutions for high-quality growth: what they are and how to acquire them, " *Studies in Comparative International Development*, 35, 3-10.

Samuelson, P., 1962, Parable and realism in capital theory: the surrogate production function, Review of Economic Studies, 29, 3, 193-206
SIEMS, M, [2005], "Numerical Comparative Law - Do We Need Statistical Evidence in Order to Reduce Complexity?," *Cardozo Journal of International and Comparative Law*, 13, 521-540.
SMELSER, N. AND R. SWEDBERG (eds) [1994], "Introduction," in *The Handbook of Economic Sociology,* Princeton, Princeton University Press.
Solow, R., 1963, *Capital Theory and the Rate of Return*, Amsterdam: North-Holland
SONO, M. [1945], "The effect of prices changes on the demand and supply of separable goods," *Kokumin Keizai*, 74, 1-51. Reprinted in the *International Economic Review*, 1961, 2, 239-275.
Swan ,T., 1956,Economic growth and capital aggregation, Economic Record, 32,2, 343-361
THOMAS, M.A. [2007], "What do the worldwide governance indicators measure?" ; Available at SSRN: http://ssrn.com/abstract=1007527
THORNDIKE, E.L. [1920], "A constant error in psychological ratings", *Journal of Applied Psychology,* 4, 469-477.
UNITED NATIONS [2007], *Public Governance Indicators: A literature Review, Department of Economic and Social Affairs,* New-York.
Usher, D., 1981, Introduction, in D. Usher (ed.) *The Measurement of Capital* University of Chicago Press 1-21,
VRIES, [2007], How does your country rank? The discussion about international data on good governance, *Proceedings from the International Conference on Government Performance Management.*
WILLIAMSON O.E., [xxx] « The New Institutional Economics: Taking Stock, Looking Ahead », *Journal of Economic Literature,* 38, 3, pp. 595-613.

Eric Brousseau
EconomiX
Université de Paris X
200 avenue de la république
F-92001 Nanterre Cedex, France
E-mail: eric@brousseau.info

Sophie Harnay
EconomiX
Université de Paris X
200 avenue de la république
F-92001 Nanterre Cedex, France
E-mail: sophie.harnay@u-paris10.fr

Jean-François Sattin
Larime
Université de Valenciennes
Site Les Tertiales
Rue des Cent Têtes
59300 Valenciennes, France
E-mail: jean-francois.sattin@univ-valenciennes.fr