



UNIVERSITY OF
GOTHENBURG

CONCEPTUALIZING AND MEASURING STATE CAPACITY:

Testing the Validity of Tax Compliance as a Measure of State
Capacity

MATTIAS OTTERVIK

WORKING PAPER SERIES 2013:20

QOG THE QUALITY OF GOVERNMENT INSTITUTE
Department of Political Science
University of Gothenburg
Box 711, SE 405 30 GÖTEBORG
December 2013
ISSN 1653-8919
© 2013 by Mattias Ottervik. All rights reserved.

Conceptualizing and Measuring State Capacity: Testing the Validity of Tax Compliance as Measure of State Capacity
Mattias Ottervik
QoG Working Paper Series 2013:20
December 2013
ISSN 1653-8919

ABSTRACT

This paper proposes and quantitatively tests a measurement of state capacity using Robert Adcock and David Collier's four-step framework. Drawing from the work of state-centered structuralists, rational choice-inspired theorists, as well as studies of the relationship between state and extractive capacity, state capacity is defined as *the ability of the state to dominate, i.e. coax compliant behavior from, the individuals of a given territory* and operationalized as tax compliance (as measured by the size of the shadow economy relative all legal economic activity). Large-n correlations as well as regression tests show that this operationalization has convergent and discriminant validity, as well as explanatory power. A paired comparison between China and India suggests that this operationalization is able to account for the differences in apparent governance capacity and development performance that other governance measurements cannot.

Keywords:

Governance; state capacity; human development; China; India

Mattias Ottervik

Department of Political Science
University of Gothenburg
mattias.ottervik@gu.se

INTRODUCTION

State capacity is a critical concept in political science, especially important perhaps within the good governance/quality of government literature. Yet there are few, if any, good measures of it. Reflecting on this situation, Francis Fukuyama in a commentary in *Governance* called for a better conceptualization of governance in order to remedy the “poor state of empirical measures of the quality of states, that is, executive branches and their bureaucracies” (Fukuyama 2013: 347).

Conceptualizing governance as a whole, which touches on all aspects of political power (cf. Peters and Pierre 2006: 209; Fukuyama 2013: 350), is an awesome task. One way to approach this is to break governance into its constituent parts, the making and enforcing of rules. This paper focuses on the latter of these, the ability of a political system to enforce rules and to deliver services, i.e. state capacity.

State capacity, the “degree of control that state agents exercise over persons, activities, and resources within their government’s territorial jurisdiction” (McAdam, Tarrow and Tilly 2001: 78), is critical for the performance of a political system. High capacity states are able to provide public goods such as human security, medical and health care, and the social and physical infrastructure that promote human development (Rotberg 2003: 2-4). Low capacity states are limited in their ability to provide these goods, leading to low social trust (Rothstein and Stolle 2008), low development levels, regime, or even state failure (Rotberg 2004; Skocpol 1979). For democracy to be consolidated and successful over time a high capacity state is critical (Wang 2003; Fukuyama 2005; Cf. Diamond 2000; Carothers 2002).

The concept of state capacity is theoretically well developed. From Weber’s original definition of the state and discussion of domination (Gerth and Mills 1946: 78; Weber 1968: 53-54, 212-216), the theoretical nature of both have been extensively developed (cf. Mann 1984; Migdal 2001). However, given the importance of the concept for studies of development performance, regime-, and state stability, it is noteworthy that there are few, if any, cross-country measures of state capacity (Fukuyama 2013:353-355).

Part of the reason why practical quantitative measurements of state capacity are few, if not non-existent, is likely that while the background concept of state capacity is widely accepted, scholars working in different traditions use different terminology that reflect subtle conceptual differences.

For example, state power (Skocpol 1979; Mann 1984; Tilly 1991; Mann 1984), state strength (Migdal 2004), state capacity (Hui 2005; Fukuyama 2013), and (government) effectiveness (Rothstein 2011) all describe the same background concept, i.e. the ability of states to implement policy, but reflect differences in systematized conceptualization.

Measurement validity, "whether operationalization and the scoring of cases adequately reflect the concept the researcher seeks to measure" (Adcock and Collier 2001: 529), is dependent on the systematized concept, which is derived from the background concept. Scholars whose work is largely qualitative and theoretical tend to have more clearly defined systematized concepts, whereas more quantitatively focused scholars use or create measurements or approximations of state capacity that are based on unclear systematized conceptualizations. The disjuncture between qualitative and theoretical on the one hand and quantitative on the other is not conducive for the creation of indicators with measurement validity.

The lack of a valid measurement of state capacity is especially problematic for the good governance/quality of government literature that is marked by quantitative analysis. One of the most common indicators of the state's capacity to implement policy, the Government Effectiveness (GE), one of the six World Bank Governance Indicators (WGI) which have received considerable scholarly criticism (Kurtz and Schrank 2007; Arndt 2008; Andrews 2008; Pollitt 2008; Thomas 2010; Langbein and Knack 2010).¹ The most powerful criticism of the WGI is that they are atheoretical (Andrews 2008; Arndt 2008), and lack measurement validity (Thomas 2010; cf. Adcock and Collier 2001). In a defense of the WGI the principal authors argue that a lack of definitional consensus means that they are free to propose their own definitions (Kaufmann, Kraay, and Mastruzzi 2010a: 56). This defense however misses the point. Aggregating existing indicators and then theoretically defining the new index precludes the possibility of measurement validity; for the WGI to have measurement validity, the process would have to be reversed. Without measurement validity the analyses using GE, or any other WGI, are problematic because it is unclear what a key indicator is actually a measure of.

The purpose of this paper is to test an indicator of governance, state capacity, that has measurement validity. Drawing from the work of state-centered structuralists (Tilly 1992; Hui 2005), ra-

¹ For summaries of criticisms against the WGI, cf. Pollitt 2009 and Apaza 2009.

² In later writings Easton replaced 'government' with 'political authorities' (1965;1975), which in recent literature seems

tional choice-inspired theorists (Levi 1981; 1988), as well as studies of the relationship between state and extractive capacity (Wang and Hu 2001; cf. Wang 2001), state capacity is defined as *the ability of the state to dominate, i.e. coax compliant behavior from, the individuals of a given territory* and operationalized as tax compliance (as measured by the size of the shadow economy relative all legal economic activity). Large-n correlations as well as regression tests show that this operationalization has convergent and discriminant validity, as well as explanatory power. A paired comparison between China and India suggests that this operationalization is able to account for some of the differences in development performance that other measurements can not (cf. Fukuyama 2013: Appendix). In other words, state capacity as measured by tax compliance has measurement validity and is analytically useful.

The structure of this paper follows the Adcock and Collier framework (Adcock and Collier 2001: 531) for testing indicators. First, the basic concept, the "broad constellation of meanings and understandings associated with" the state, is discussed. Second, the concept is systematized, i.e. the background concept is distilled down to "a specific formulation" and defined explicitly. Third, the systematized concept is operationalized. Finally, the scores produced by the operationalization are tested on a number of cases. This process is followed to ensure that the proposed indicator has measurement validity, which is "achieved when scores (including the results of qualitative classification) meaningfully capture the ideas contained in the corresponding concept" (Adcock and Collier 2010:530).

BACKGROUND CONCEPT

Before state capacity can be systematically conceptualized in a definition, the basis for the operationalization, the background concept must be reviewed. A comprehensive state of the field that accounts for all approaches to the study of the state is beyond the scope of this paper so the focus of this section will be on three core issues. The first issue is the nature of the state and social domination. The second issue is the state's relationship to other forces. This third issue is how to understand the role of the state in a political system, i.e. does it describe the whole or is merely one part? While a treatment of these issues might seem an unnecessary excursus for what is ultimately a quantitative study, the background concept forms the basis for the systematized concept. Omitting or

glossing over this part in favor of quantitative tests would not produce an improvement over present indicators.

This first issue is the nature of the state and social domination. The classical definition of the state is Weber's, that it is "a human community that (successfully) claims the monopoly of the legitimate use of physical force within a given territory" (Gerth and Mills 1946: 78). This pithy formulation is however the introduction to a more nuanced discussion wherein the state is described as a "relation of men dominating men ... [I]f a state is to exist, the dominated must obey the authority claimed by the powers that be." (Gerth and Mills 1946: 78). While this relation is supported by violence, what characterizes it is social domination not violence. Domination (*Herrschaft*) is "the probability that a command with a given specific content will be obeyed by a given group of persons" (Weber 1968: 53). The measure of domination is compliance; the nature of the state is a domination that produces compliance within a given territory.

According to Weber domination rests on two pillars, the inner justification (i.e. legitimacy) as well as the external means. These two pillars have inspired two complementing lines of research. Using a rational choice approach focused on the individual citizen (cf. Levi 1988:185-204), Margaret Levi has studied the intimately related quasi-voluntary compliance, consent, and legitimacy (Levi 1988; 1997; 2009). By contrast, the scholars associated with the 'Bringing the State Back In' (BTSBI) movement (cf. Evans, Rueschemeyer, and Skocpol 1985) have generally focused on the means by which the state sustains itself (cf. Skocpol 1979: 29; Tilly 1992). These lines of research intersect the most clearly in the focus on the state as an aggregator of resources. Levi's point of departure is the "major limitation on rule is revenue" (Levi 1988: 1). Charles Tilly's argument that "war made the state, and the state made war (Tilly 1975: 42) is less about war and more how the efforts of European rulers to mobilize more and more resources fundamentally transformed their polities. The social domination that the state represents is based on and manifested in the mobilization of men and money.

The second issue is the state's relationship to other forces. There are two aspects of this issue. The first aspect is how autonomous the state is. The second aspect is what actors the state interacts with. With regards to the autonomy of the state Michael Mann has argued that the power of the state and its elites is not unidimensional. According to Mann despotic power, "the range of actions which the [state] elite is empowered to undertake without routine, institutionalized negotiation with civil soci-

ety groups”, is distinct from infrastructural power, “the capacity of the state actually to penetrate civil society, and to implement logistically political decisions throughout the realm” (Mann 1984: 113). Despotic power, the power of the state elite *over* society is shallow, and allows for autonomy but is not able to transform society. Infrastructural power by contrast is deep, transformative, but also limiting as it creates or strengthens countervailing civil societies group. While some of the historical examples of despotic power are arguable, considered as ideal types the two are analytically useful. The idea of infrastructural power agrees with the idea of the state-society bargain, wherein states must engage with society in return for invasive mobilization of its resources (Tilly 1992: 99-103; Hui 2005: 170-177). The state is not autonomous, but of rather than above society.

The second aspect of the state’s relationship to other forces is what actors the state interacts with. Scholars working from Weber’s definition of the state tend to conceive it as distinct and separate from an often undifferentiated civil society. Based on studies of state-society relations in the (so-called) Third World Joel Migdal argues against this conceptualization (1988), proposing in its stead the ‘state-in-society model’ (Migdal 2004). The state is one actor among the many social forces that seek to “mobilize followers and exercise power” (Migdal 1994: 20-21; Cf. Migdal 2004, 49-50). The interaction between state and social forces can according to Migdal produce four outcomes. The first is a total transformation of society, as the state’s penetration leads to the subjugation of social forces and the state’s domination. The second is state incorporation of existing social forces, as the state appropriates existing social forces, but is in the act also transformed by them. The third outcome is the incorporation of existing social forces into the state, as existing dominant social forces adapt to the state’s presence and thereby prevents radical changes in the pattern of domination. Finally, the state might altogether fail in penetrating society, with little transformative effect upon either society or itself (Migdal 2004: 126-128). The state must compete and contend with a multitude of forces that seek to affect it or achieve independent domination; the domination of the state is neither complete not a given but should be considered on a scale.

The third issue is how to understand the role of the state in a political system, i.e. does it describe the whole or is it merely one part. Most studies of the state, especially those by the state-centered structuralists associated with the BTSBI-movement, tend to describe the state holistically, e.g. as the “set of organizations invested with the authority to make binding decisions for people and organizations juridically located in a particular territory and to implement these decisions using, if necessary, force” (Rueschemeyer and Evans 1985: 46-47). This description is not dissimilar to how the

word 'government' is sometimes used in other literatures to describe a whole political system (e.g. the good governance/quality of government literature, cf. Rothstein 2009). These all-encompassing terms are analytically problematic. For it to be studied the state (i.e. the political system) has to be disaggregated (Migdal 1994: 15).

Migdal proposes a four-level taxonomy of the state, but there is already a widely accepted alternative. Though it is some ways removed from the study of the state, David Easton's tripartite division of the political system into government, regime, and the political community is applicable and analytically useful (Easton 1957).² A government is the group of leaders responsible for making political decisions, i.e. a presidential administration, the party or coalition of parties controlling the prime ministry, or the party leadership in a one-party system. Regime describes “the formal and informal organization of the centre of political power “ (Fishman 1990: 428; cf. Easton 1957: 392). Political community is sometimes interpreted as nation (Norris 1999: 10; Dalton 2004: 5) but state seems more accurate.³ State indicates the entity that holds a monopoly on the legitimate violence and administers a territory.

In this taxonomy government is the least permanent feature of the political system and in democracies changes as a matter of course. Regimes may change without a complete breakdown of social order whereas state failure is associated with widespread violence and unlawful or lawless behavior (Lawson 1993; Rotberg 2004:2-3). This does not contradict the description of the state as a set of organizations. Rather, the organizations and institutions in a political system should be grouped according to function for the sake of analytical clarity.

SYSTEMATIZED CONCEPT

Based on the background concept in the previous section – the nature of the state, its relationship to other social actors, and the part it plays in the political system – state capacity is defined as *the ability of the state to dominate, i.e. coax compliant behavior from, the individuals of a given territory*. This is in line with McAdam, Tarrow, and Tilly's definition (2001: 78). Domination as such is neutral as to wheth-

² In later writings Easton replaced 'government' with 'political authorities' (1965;1975), which in recent literature seems to describe what Easton originally meant by the word government (e.g. Dalton 2004:6-7).

³ As an imagined community (cf. Anderson 2006), not seldom the product of the state (cf. Weber 1976), nation is a problematic interpretation of political community. Interpreting it as the state (as Pippa Norris also seems to do; cf. Weber 1968:901ff) has better support.

er compliance is voluntary or non-voluntary and the extent to which citizen compliance is voluntary is outside of the scope of this paper. However, the preponderance of theory, as well as history, would suggest that voluntary compliance (such as produced by democracy in the West, cf. Tilly 1992; Hui 2005) is more efficient, and more effective (Levi et al. 2009: 354-5).

While this domination might ultimately rest on the capacity for violence, in practice it is not the product of force but some generally accepted claim to legitimacy that promotes voluntary compliance. This legitimacy could be described in the form of one of Weber's ideal types, dependent on contingent consent (Levi 1997), the product of a state-society bargain, an expression of infrastructural power, or "an acceptance of the state's rules of the game, its social control, as true and right" (Migdal 2004: 52).

The domination of the state is limited in that the state does not make policy. Rather, it is at least formally the tool by which government implements the policies it formulates.⁴ While the ability to dominate the individuals of a given territory might sound ominous, if not contradictory to the idea of democracy, a consolidated democratic regime requires it (Wang 2003; Fukuyama 2005; Cf. Diamond 2000; Carothers 2002). Through policy successive governments can strengthen the state, or weaken it (cf. Tilly 1992; Hui 2005).

OPERATIONALIZATION

From the systematized conceptualization of state capacity, compliance is the most straightforward indicator of that capacity. High capacity states should be able to ensure higher levels of compliance than low capacity states. There are many forms of compliant behaviors that could be used as an indicator of state capacity, but compliance with state efforts to mobilize resources stands out in the literature on the state. Domination is an expensive enterprise that requires men and money, and the level of resource mobilization by the state determines its capacity for it (Fukuyama 2011: 470; cf. Levi 1988: 1).

⁴ The fact that some bureaucracies, such as MITI (Johnson 1982), are able to formulate policy does not mean they are formally superior to government.

A common approach to measuring the capacity of a state is to look at the amount of taxes it collects (Fukuyama 2013: 353; Lieberman 2002; Cf. Persson 2008: 30-33). This approach has merit. The history of state formation in Europe and China is the history of the state increasing its scope to more effectively mobilize resources. Put differently, states can be broken down into a number of logical functions that are mutually supportive. For an effective monopoly of violence (coercive function) the military and police need to have resources extracted from society (extractive function), be supported by effective bureaucracies that coordinate with other elements of the state that also constrain them (control function), and operate in a society with some level of consent (legitimation function); “the state’s capacity to mobilize and extract financial resources is the core of state capacity and the foundation for the state’s ability to realize its other capacities” (Wang and Hu 2001: 27; Cf. Levi 1988:1). Along the same line, Schumpeter famously argued that the “public finances are one of the best starting points for an investigation of society, especially though not exclusively of its political life” (Schumpeter 1991: 101).

Though measuring state capacity by taxation as a percentage of the economy is perhaps the most straightforward approach, it nevertheless has several caveats (Cf. Fukuyama 2013: 353-355). First, it measures not only the capacity to tax, but also the willingness. As such it captures policy as much as capacity to implement the policy, and is therefore not a measure of state capacity.

Second, natural resource rents makes taxation too easy in some countries. This breaks the theoretical relationship between the different state functions as elites in such countries need only enough coercive capacity to guard the resources, which reduces the need for a well-rounded state that generates consent through social control and performance. It is not easy to compensate for the existence of such resource rents in global taxation statistics, which reduces their explanatory potential.

Finally, taxation as a percentage of the economy is ultimately a measure of the comparatively easy collection of taxes, not the more difficult registration and regulation of the economy that requires social control and the provision of public goods. Between 1999 and 2007 the average size of the shadow economy, the legal production of goods and services that are concealed from authorities, across 162 countries was about one third of the official economy (Schneider, Buehn and Montenegro 2010). This discrepancy between the official economy, which is used for tax assessment, and the much larger actual economy makes taxation as a measure of state capacity problematic.

As an operationalization of state capacity tax compliance has more merit than simple extraction. First, it measures extractive capacity only, not tax policy. Second, it removes the problem of having to compensate for resource rents. Finally, it measures the capacity of the state to penetrate society and regulate the economy. It is a measure of the ability of the state to dominate, i.e. coax compliant behavior from, the individuals of a given territory.

There are a number of approaches to measuring the tax compliance. One would be to aggregate national statistics on tax compliance, which is often collected by national tax agencies. There are two caveats to this. First, tax evasion is not uniformly defined. Second, it is not clear that the organization whose job it is to assess economic activity is any better at assessing its failure to do so. Another approach would be to use a uniform methodology to evaluate tax compliance in all countries, and hope that one could do a better job than the individual tax agencies. Finally, one could use existing research on the shadow economy.

The shadow economy is the *aggregate of the legal economic activities that should be included in the National Income Accounts, but because of evasive strategies are not*; “these market economies are connected with the evasion of taxes and social security contributions, as well as the avoidance of regulations (official working time, safety regulations), and social security fraud” (Schneider and Enste 2013: 9-10). The shadow economy excludes financial transaction tax evasion, which is often hard to distinguish from legal tax avoidance, and the informal sector that is often not taxed (Schneider and Enste 2013: 9-10). Legal economic activities are distinct from illegal activities in that the former are sanctioned by law when reported and regulated, whereas the latter are not sanctioned under any circumstance. Selling food or painting someone’s home would be examples of legal economic activities, whereas selling drugs or engaging in prostitution are (generally) illegal economic activity.

Measuring the size of the shadow economy is in some ways similar to measuring a black hole in that its existence and size must imputed from what is measurable. There are three approaches to doing this: the direct, the indirect, and the model (Schneider and Hofreither 1986: 18-19; Schneider and Enste 2013).

The direct approach, using surveys and point (tax) audits to infer the size of the shadow economy, has a number of drawbacks, such as the representativeness of the surveyed population being unclear and that it is very costly. It is not something that can be funded on any larger scale by anything

else than a (truly dedicated) national state. For large-n global comparisons this approach is of limited usefulness.

The indirect approach, comparing economic indicators to estimate the shadow economy, is a collection of approaches that all have their strengths and weaknesses. A common weakness is that it is a comparison of two or more possible incorrect measurements to determine a third. This adds a layer of complexity that is potentially problematic.

The model approach uses a model to calculate the (unobservable) size of the shadow economy based on a number of causes and indicators. This is the model favored by Friedrich Schneider, Andreas Buehn, and Claudio Montenegro (2010) and Friedrich Schneider and Dominik Enste (2013).⁵ Based on 25-years of research, in Schneider et al.'s model the causes are the size of actual and perceived direct and indirect taxation, regulatory burden, tax morale, and income. The indicators in the model are monetary indicators, labor market development, and production market development.

Schneider et al. (2010; 2013) calculate the size of shadow economy as a percentage of the official economy, i.e. the legal economic activity accounted for in national statistics. For example, the shadow economy was in the five-year period between 2003 and 2007 on average 34.3 percent of the official economy in Albania. This data makes it possible to calculate the tax compliance of citizens engaging in legal economic activity. Between 2003 and 2007 the Albanian state was only able to regulate, and thereby tax, about 75 percent of legal economic activity. On a scale from 0 to 1, the tax compliance in Albania was 0.748.

The resultant index of state capacity is *prima facie* reasonable (appendix A). The top-10 countries are Switzerland, United States, Luxembourg, Austria, Japan, Macao (China), New Zealand, United Kingdom, China, and Singapore. With the possible exception of China, which is discussed in greater depth below, these are all polities that could be expected to top an index of state capacity. Some countries such as Sweden or Denmark are lower on the index than might be expected given their scores on other governance indices. This is cause for continued study, but for now the validity tests of the index suggests that it is generally correct. Furthermore, given that apparent overall validity of tax compliance as a measure of state capacity, it is possible that some of the discrepancy

⁵ For an in-depth discussion, cf. Schneider, Buehn and Montenegro 2010 and Schneider and Enste 2013.

between this index and other indices reflects the fact that most governance indicators are either very broad and/or perception-based.

The systematized conceptualization of state capacity - *the ability of the state to dominate, i.e. coax compliant behavior from, the individuals of a given territory* - is operationalized as tax compliance. Tax compliance is calculated from the size of the shadow economy as a percentage of all legal economic activity.

VALIDITY TESTS

Testing Design and Data Sources

Having created a systematized concept from the background concept, and then operationalized the systematized concept, the validity of the resultant indicators needs to be tested. The purpose of the validity testing is to verify that “the indicator employed produces scores that can be interpreted as adequately capturing the systematized concept” (Adcock and Collier 2001: 533). Testing the validity of an indicator meant to capture something as intangible as state capacity presents a challenge. In this paper validity is tested three ways. The first two tests are quantitative, and the final is qualitative.

First, convergent validity and discriminant validity are tested. Because there is no “true” measure against which to assess validity, a number of related indicators are used (cf. Adcock and Collier 2001: 541). Governance indicators which were designed to capture concepts theoretically close to state capacity, an indicator that measures state provision of public goods, as well as an indicator (level of democracy) that is weakly related to state capacity are correlated against the proposed indicator.⁶ For the proposed indicator of state capacity to be valid, the governance-related indicators should be highly correlated, the outcome-related variable should be somewhat less highly correlated (as it also captures policy), and the democracy measure should be weakly related. This outcome would demonstrate convergent and discriminant validity.

Second, the proposed indicator is used in a regression analysis as one of several independent variables. The dependent variable is development as measured by the UN’s Human Development Index

⁶ The relationship between democracy and state capacity is weak in the sense that while state capacity is a prerequisite for consolidated democracy, non-democratic regimes can be consolidated and democracy *per se* does not appear to produce state capacity.

(HDI). Human development, as measured by the HDI, is a product of the willingness of government and the capacity of the state to provide public goods. Research on good governance/quality of government is generally focused on explaining and finding the factors that drive development. As such this is a test of the usefulness of the state capacity indicator to such analysis. The indicator should ideally have some explanatory power when other indicators that are known to promote development are controlled for. The level of explanatory power is not as important as the existence of a statistically significant effect; the value of the proposed indicator lies in its measurement validity, not correlation with outcome.

Finally, the indicator is used in the critical case of China and India's development performance. There is an incongruence between the development performance, the hall-mark of 'good governance', in China and the country's low scores on various governance measures (Mahbunani 2013; Rothstein 2012: 3; Cf. Fukuyama 2013: 366). China often scores worse than India on these measures, even though its development performance has been consistently better since the founding of the respective countries. This makes China along with India a critical case in the negative sense; the proposed operationalization should ideally be able to account for some China's apparent capacity to provide public goods.

With the exception of the shadow economy all variables used in the quantitative tests are from the University of Gothenburg Quality of Government Institute Standard Dataset (Teorell et al. 2013). The institute is an independent research institute within the Department of Political Science at the University of Gothenburg and it maintains a number of datasets on various aspects of governance. All variables in the quantitative analysis are from the May 2013 QoG dataset. Except for the shadow economy and the historical HDI control variable, the data is generally for 2009-2010. For the qualitative test data is drawn mainly from UN sources, such as the 2010 Human Development Report.

Because the West tends to be a statistical outlier the quantitative tests break out results for the West and the world outside of the West.

Indicators

Tax compliance rates are calculated from the size of the shadow economy in different countries (Schneider, Buehn and Montenegro 2010). While newer data can be found in Schneider and Enste

2013, the country coverage is better in Schneider, Buehn and Montenegro 2010. The methodology appears unchanged between the two, and there does not appear to be significant differences between them ($r=0.971$). The shadow economy is measured as a percentage of the official economy, and this is used to calculate the tax compliance rates of all legal economic activity. The resultant index is from 0 to 1 where higher numbers mean greater tax compliance. The mean value is 0.760 for the world ($n=160$),⁷ and 0.859 for the West ($n=23$).

The governance indicators are drawn from the ICRG Quality of Government (QoG), TI's Corruption Perception Index (CPI), as well as the WGI. These are designed to measure aspects of governance that should correlate highly with high levels of state capacity, such as low levels of corruption, rule of law, and effective bureaucracy. Of the tested indicators GE, which captures "perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies", is perhaps conceptually the closest to state capacity. It is often used as a stand-in for state capacity, and it could be expected to have one of the highest correlations with state capacity operationalized as tax compliance.

QoG is a composite of three dimensions -- corruption, law and order, and the quality of the bureaucracy -- that capture the functioning of a state. Scores are on a scale from 0 to 1, where higher values indicate higher quality of government. Globally the mean is 0.53 ($n=139$), and 0.86 ($n=23$) in the West.

TI's Corruption Perception Index (CPI) captures perceptions of corruption in the public sector where corruption is defined as the abuse of public office for private gain. Scores range between 0 and 10, where higher numbers indicate lower perceptions of corruption. The global mean score is 3.98 ($n=181$), and the Western mean is 7.65 ($n=23$).

The six WGI capture perceptions of fundamental governance concepts that the authors argue are related to development outcomes, "such as higher per capita incomes, lower infant mortality, and higher literacy" (Kaufmann, Kraay and Zoido-Lobaton 1999: 4; Cf. Kaufmann, Kraay, and Mastruzzi 2010b: 20). They should therefore be directly related to development outcome as measured by the HDI, and indeed the average of the WGI correlates highly with human development

⁷ Hong Kong and Macao are not sovereign states, and therefore not included in the world average.

level ($r=0.787$). Using a survey-of-surveys approach, the indicators are aggregates of various perception-based indices and do not measure governance directly. The scale is from about -2.5 to 2.5 and higher values indicate better performance. The six indicators of the WGI measure three aspects of governance (Kaufmann, Kraay, and Mastruzzi 2010b: 4):

- a) The process by which governments are selected, monitored and replaced:
 - Voice and accountability
 - Political Stability and Absence of Violence/Terrorism
- b) The capacity of the government to effectively formulate and implement sound policies:
 - Government Effectiveness
 - Regulatory Quality
- c) The respect of citizens and the state for the institutions that govern economic and social interactions among them:
 - Rule of Law
 - Control of Corruption

Empirically, there seems to be little real-world difference between the four of the six indicators as the correlation between GE, Regulatory Quality, Rule of Law and Control of Corruption is very high at $r>0.92$. It stands to reason that perceptions of a polity's ability to formulate and implement sound policy should correlate highly with the respect of the citizenry for the institutions that govern economic and social interactions. However, the high correlation has drawbacks and not all WGI are used in the validity tests. The mean of the WGI Average is -0.07 for the world ($n=191$), and 1.41 for the West ($n=25$).

The HDI uses a simple model of human development based on three components: longevity, education and a decent living standard (UNDP 1990: 12; UNDP 2010: 13-15). It is a well-rounded measure of a country's development level, reflecting both willingness of government as well as the ability of the state to provide public goods. It is on a scale from 0 to 1, where higher is better. The HDI mean is 0.66 for the world ($n=185$), and 0.88 for the West ($n=25$).

There appears to be a level of path dependency to human development trends, at least on a regional level. Development performance over time shows that relative development levels of regions are, excepting East Asia, stable (UNDP 2010: 56). There is a natural lag in the dimensions that constitute the index as the average life expectancy, education level, and income of a population cannot be changed overnight. To account for the importance of historic development levels the 1970-1974 average HDI score is used as a control variable in the regression.

Freedom House/Imputed Polity is used as a measure of democracy. The scale is from 0 to 10, where 10 is the most democratic. The mean is 6.29 for the world (n=162), and 9.92 for the West (n=27).

Finally, in the quantitative analyses the Teorell and Hadenius 2005 politico-geographical classification of world regions is used to identify Western countries. The West, category 5, includes Western Europe, North America, as well as Australia and New Zealand (but not Cyprus), 27 countries in all (Teorell and Hadenius 2005; Teorell et al 2013).

Some indicators have also been normalized for the regression analysis. The descriptive statistics of all variables are in table 1.

	Descriptive Statistics				
	N	Minimum	Maximum	Mean	Std. Deviation
State Capacity (Tax Compliance)	160	0.61	0.92	0.7601	0.07184
ICRG Indicator of Quality of Government	139	0.08	1.00	0.5323	.20281
TI Corruption Perceptions Index	181	1.01	9.40	3.9829	2.08954
WGI Average	191	-2.41	1.86	-0.0714	.90112
WGI Average (Normalized)	191	0.00	1.00	0.5477	.21103
WGI Control of Corruption	191	-1.73	2.48	-0.0648	1.00182
WGI Government Effectiveness	191	-2.27	2.29	-0.0625	.99270
WGI Government Effectiveness (Normalized)	191	0.00	1.00	0.4841	.21770
WGI Political Stability	193	-3.32	1.57	-0.0596	1.00197
WGI Rule of Law	193	-2.49	1.97	-0.0724	.99283
WGI Regulatory Quality	191	-2.56	1.91	-0.0750	.99079
WGI Voice and Accountability	193	-2.24	1.57	-0.0575	1.01309
Freedom House/Imputed Polity	193	0.00	10.00	6.6731	3.10514
Freedom House/Imputed Polity (Normalized)	193	0.00	1.00	0.6673	.31051
Human Development Index	185	0.28	0.94	0.6555	.17333
Historical HDI (1970-74)	134	0.15	0.81	0.5435	.18466
Valid N (listwise)	113				

Table 1. Descriptive statistics of all variables.
Source: Teorell et al. 2013; Schneider, Buehn and Montenegro 2010.

Quantitative Tests

Convergent and Discriminant Validity Test

For the operationalization of state capacity as tax compliance to be valid it should correlate highly with related governance variables, somewhat weaker with development performance, and weakly with democracy level. The four governance variables that are included in the correlation test are the ICRG QoG, the TI Corruptions Index, the average of the six WGI, as well as the WGI GE. The outcome variable is the HDI. Freedom House/Imputed Polity measures level of democracy.

The results are largely as theory would suggest, and the proposed indicator appears to have both convergent as well as discriminant validity (table 2).

	Correlations between State Capacity and other Indicators		
	World State Capacity I	World Excl. West State Capacity II	West Only State Capacity III
ICRG Indicator of Quality of Government	.773**	.638**	.629**
TI Corruption Perceptions Index	.722**	.536**	.704**
WGI Average	.693**	.495**	.613**
WGI Government Effectiveness	.737**	.581**	.595**
Human Development Index	.629**	.474**	.445*
Freedom House/Imputed Polity	.271**	-.011	.517*
	.002	.912	.012
	N=133	N=110	N=23

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Table 2. The correlations are in the directions that theory suggests.
Source: Teorell et al. 2013; Schneider, Buehn and Montenegro 2010.

For the world (column I) as a whole the correlation between tax compliance and the governance variables are between .773 for QoG, and .693 for the WGI Average. That the weakest relationship is between tax compliance and the WGI Average stands to reason as it is conceptually the most confused. The relationship between tax compliance and development level is somewhat weaker than that between tax compliance and various governance indicators (.629).

At .271 the weakest relationship is between tax compliance and democracy level. The correlation between the other governance indicators and democracy level is .493 (QoG) between .661 (WGI Average). Tax compliance seems to capture state capacity more narrowly than the other governance

indicators that capture also the distinct concept of democracy. Tax compliance thus appears to have discriminant validity.

Outside of the West (column II) the general strength of the relationships hold true. Governance indicators correlate the strongest with the shadow economy, and human development level is somewhat weaker. The relationship between democracy level and state capacity as measured by the shadow economy breaks down completely outside of the West, and is in fact no longer significant. A case could be made for a quadratic, as opposed to linear, relationship between democracy level and state capacity where consolidated democracies *and* non-democracies have high state capacity. Regime consolidation is difficult in weak states, and the countries where neither form of regime is consolidated (i.e. the countries which exist in what Thomas Carothers (2001) calls the Gray Zone) are consequently those with the lowest tax compliance.

Focusing only on Western countries (column III) the relationship between tax compliance and related governance indicators is strong as expected. The relationship between tax compliance and human development level is somewhat weaker, also as expected. The comparatively strong relationship between tax compliance and democracy level can be explained by the outliers. The only Western countries not to score a 10 on the Freedom House/Imputed Polity index are France, Belgium, Italy, and Greece. As a group these countries have somewhat lower tax compliance than the majority of the rest of the countries. That tax compliance correlates with level of democracy could be seen as support of, or at least in line with, the European state-society bargain whereby citizens were offered democracy in return for more invasive resource mobilization (cf. Tilly 1992).

Across the world tax compliance correlates highest with the measurements that are closest to the ability of the state to coax compliance, i.e. the QoG (which should capture level of corruption, law and order, and the quality of the bureaucracy), corruption, and government effectiveness, and somewhat weaker with the more all-encompassing WGI average. The convergent validity of tax compliance as an operationalization of state capacity appears to be good. Without more precise alternative measures of state capacity, discriminant validity is harder to conclusively establish. However, higher correlations for QoG and GE than the more expansive WGI, lower correlation with HDI, and quite low correlation with democracy level in the world (especially compared to other governance indicators) support the discriminant validity of tax compliance.

Explanatory Power

For the operationalization of state capacity as tax compliance to be useful, it should have some explanatory power when other indicators that are known to promote development are controlled for. In theory state capacity determines the ability of the state to produce public goods, such security, education, and health care. The capacity to provide these goods is however distinct from the political willingness to do so. As an ultimate rather than proximate determinant of human development the effect of state capacity on development should be significant, but the relationship should not be perfectly linear.

State capacity is used as an independent variable in a correlation and regression where development is the dependent variable. The other independent variables are democracy, government effectiveness (GE), and good governance (here operationalized using the average of the WGI) - all common factors in models of development. Historical development levels are also included because of their influence on present-day development levels.

Tax compliance, HDI, and the historical HDI are all on a scale of 0 to 1, where higher values are better. GE, the average of the WGI, and the Freedom House/Imputed Polity democracy measure were normalized so that the scale is the same for all variables.

The chosen indicators are, with the exception of Freedom House/Imputed Polity, strongly correlated with human development (table 3). In the world as a whole and the world excluding the west (column I and II) the relationships are generally the same. Historical HDI level is highly correlated

	Correlations between Human Development and Significant Indicators		
	World	World Excl. West	West Only
	Human Development	Human Development	Human Development
	I	II	III
State Capacity	.616**	.455**	.458*
	.000	.000	.032
Avg. 1970-1974 Hybrid HDI	.942**	.920**	.823**
	.000	.000	.000
WGI GE Normalized	.807**	.736**	.451*
	.000	.000	.035
WGI Avg Normalized	.778**	.683**	.477*
	.000	.000	.025
Freedom House/Imputed Polity	.475**	.294**	.164
	.000	.002	.465
	N=129	N=107	N=22

** . Correlation is significant at the 0.01 level (2-tailed).
 * . Correlation is significant at the 0.05 level (2-tailed).

Table 3. With the exception of democracy level, the independent variables correlate highly with human development. Source: Teorell et al. 2013; UNDP 2010; Schneider, Buehn and Montenegro 2010.

with present-day development. In the world as a whole this relationship is almost perfectly linear. Conversely, the relationship between level of democracy and human development is comparatively weak. The average of the WGI and the WGI are highly correlated with human development. Given the origin of the WGI this stands to reason. State capacity is also highly correlated with human development, but not as highly as the conceptually more expansive WGI.

Looking at the relationships between the various indicators and human development in the West (column III) relationships are weaker, and the significance is for all indicators, outside of historical HDI level, lower. The small number of cases, only 22 for the West, limits the conclusions that can be drawn about the difference between the West and the rest of the world. What is important is that the general relationship appears to hold true even in the West.

Using the same indicators in a regression analysis (except for GE whose correlation with tax compliance is too high), where the effect of the individual indicators on human development are controlled for, changes the results somewhat (table 4).⁸ The effect of historical development levels is

	Human Development		
	World	World Excl. West	West Only
	I	II	III
State Capacity	.273*** (.086)	.348*** (.094)	-.140 (.137)
Avg. 1970-1974 Hybrid HDI	.789*** (-.004)	.783*** (.036)	.741*** (.145)
WGI Avg Normalized	.132*** (.044)	.185*** (.051)	0.031 (.063)
Freedom House/Imputed Polity	-.004* (.002)	-.004* (.002)	0.031 (.035)
Intercept	-0.027 (-.027)	-0.103 (.064)	0.099 (.334)
Adjusted R2	.924	.905	.668
N	129	107	22

***. Correlation is significant at the 0.01 level (2-tailed).
 **. Correlation is significant at the 0.05 level (2-tailed).
 *. Correlation is significant at the 0.1 level (2-tailed).

Table 4. When the WGI and historical HDI have been controlled for the shadow economy variable has an impact on the development level of a country.

Source: Teorell et al. 2013; UNDP 2010; Schneider, Buehn and Montenegro 2010.

strong, and significant across the world, whether the West is included or not. Level of democracy,

⁸ The very high R²-value is because of the inclusion of historical HDI-data in the regression analysis. Removing historical HDI-data reduces the R² to 0.631. It was included because of the high correlation between present-day and historical development-level.

at least as measured by the Freedom House/Imputed Polity index, loses in significance and its effect is (negative and) negligible when other factors are controlled for.

In the world as a whole the effect of state capacity on human development is strong, .273 and significant (column I). Removing the West increases the impact of state capacity, to .348, and the result is still significant at the 0.01 level (column II). The effect of WGI also increases somewhat with the West removed from the universe of analyzed cases.

Looking specifically at the West (column III) all indicators except for historical development level lose significance completely. The why of these results are outside the scope of this study, but they are suggestive. A possible explanation for this is twofold. First, looking specifically at the West reduces the number of cases significantly. Second, the West attained high development levels many decades ago, and it is possible that the institutional features that made it possible to achieve those levels are different from those that have at least in the short term maintained the West's lead in development levels.

The results of the analysis suggest one general and one tax compliance-specific conclusion. The general conclusion that can be drawn is that historical development levels have a significant impact on present-day development level when other factors are controlled for. This suggests that it is probably the periods of rapid development in more developed countries that could serve as a model for institutional reform in less developed countries, not the current institutional make-up of the currently developed countries whose current development levels undoubtedly reflect past development performance and governance. Not even Denmark knows how to 'get to Denmark' today (cf. Fukuyama 2011: 14).

With regards to state capacity as measured by tax compliance it has an effect on human development across the world, especially outside of the West, in the manner suggested by theory. While tax compliance might not correlate as highly with human development as other governance indicators, its parsimony and theoretical grounding does make it more clear what it measures. This suggests that beyond appearing to be valid, the indicator is analytically useful.

Qualitative Test: China and India Development

The proposed indicator for state capacity has been tested for validity and explanatory power in a quantitative test. Here it is tested on a critical case: China's development performance. It should

ideally be able to better account for China's development than other governance indicators, according to which China's development performance is inexplicable.

Historical human development levels have a considerable effect on present-day development levels. This means that they should ideally be controlled for in studies of development performance. From this perspective China and India is uniquely interesting in a comparative study. When China and India were founded in 1949 and 1947 their development levels were almost the same. Life expectancy was slightly longer in China, while literacy was slightly higher in India. Thirty years later, in 1979, China's HDI was 12% higher than India's, and by 2010 it was 27% higher (figure 1).

There is every reason to be skeptical of historical statistics such as these, especially in developing countries where less-than-stellar results were not always appreciated. However, speaking in favor of the approximate correctness of these numbers is that China's performance in the three dimensions of the HDI is steady across the decades for which the UN has statistics. This steady, continued development trend across four decades lends credence to the direction and relative difference in

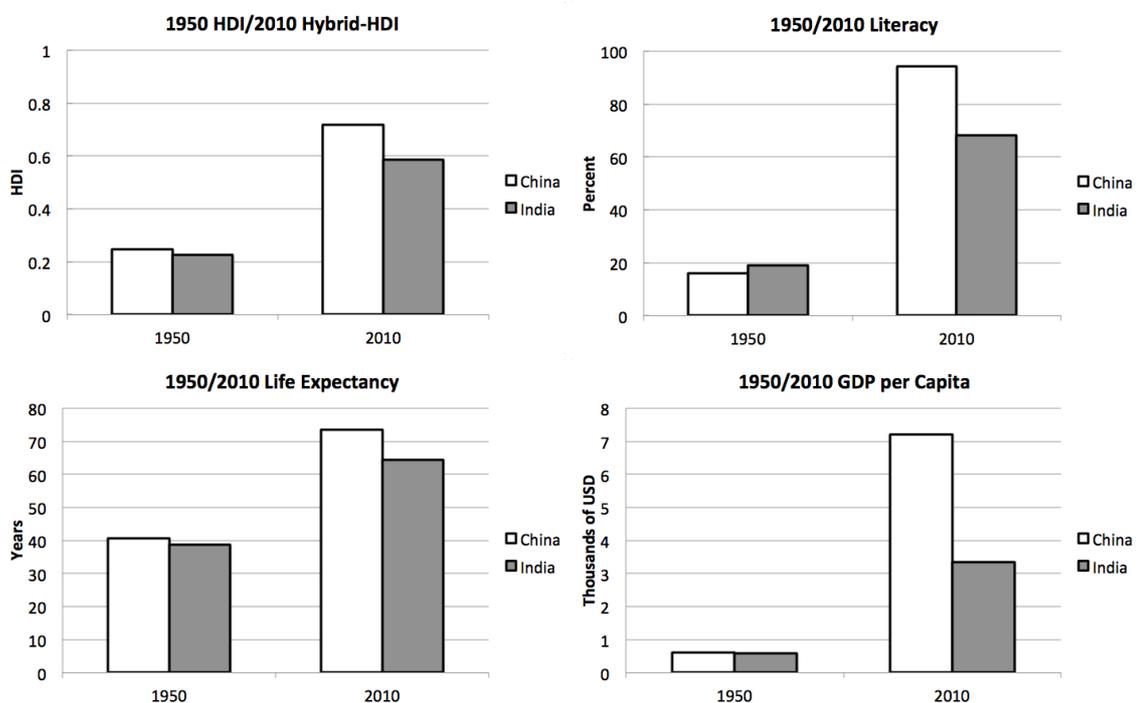


Figure 1. From nearly the same starting point in 1950, China has been able to realize a higher development than India. Note: 1950 HDI is calculated using the pre-2010 methodology while the 2010 Hybrid HDI is calculated using the new methodology but the 1950 GDP per Capita is in 1990 Int. USD, while the 2009 data is PPP-adjusted. Source: Crafts 1996; UNDP 2010.

development performance if not the absolute values of any one year (figure 2).

Because the two countries are largely similar - similar size, similar population levels, suffered similar trauma at their founding as constituent regions sought independence, and the respective governments having initially pursued similar state-led socialist development policies - this diverging development performance is noteworthy. Furthermore, regional and interpersonal inequalities appear to be comparable if not similar. Where the two differ, especially in terms of development policy, it is theoretically in India's favor as it pursued democracy and a mixed-market model rather than authoritarian communist policies.

Qualitative studies suggest that one of the critical differences between the two countries is the ability of the respective central government to mobilize resources to meet policy goals. At the founding the leaders of the respective central governments had similar development ambitions, if anything India's were more focused on development outcome, which suggests that the divergence in development performance is likely best explained by capacity not ambition (Cf. Nehru 1973: 397; Mao 1949).

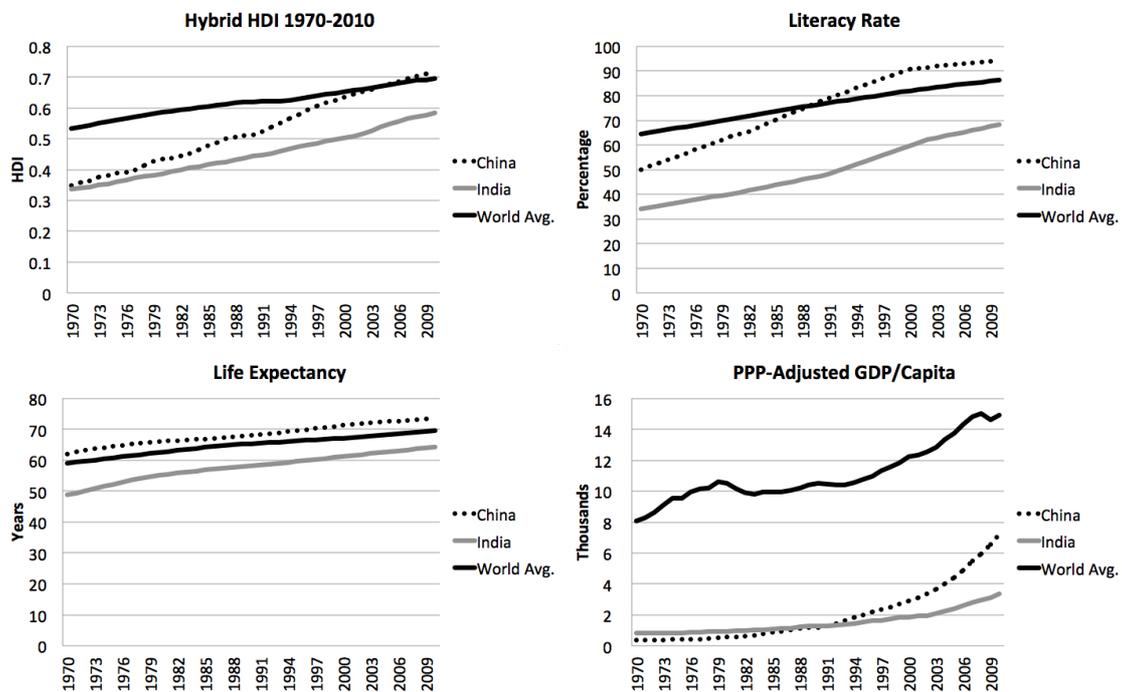


Figure 2. By 1970 China and India's human development level had diverged. This divergence has continued over the following 40 years. Source: UNDP 2010.

The Indian National Congress, the INC, the party which for decades dominated the central government under Gandhi and then Nehru and his successors, was from the founding of the country onwards a blunt instrument for the social reform its successive leaders sought (Kohli 1989: 57-58):

The ideology and the organization of a weak nationalist movement [such as the INC] precludes the use of an essential political resource, compulsion, either to resist concerted opposition from strategic social groups or to implement its redistributive goals. ... Enmeshed into the existing social structure, the INC was incapable of generating an autonomous political force to confront and reform this social structure.

In a study on India's economic development, Vivek Chibber seems to echo Kohli's evaluation of the INC in his own assessment of the Indian state (Chibber 2003: 162):

...the overall weakness of [Indian] industrial policy was generated by a lack of capacity in the two tasks that are central to a developmental state: the capacity to impose discipline on state agencies around a coherent project and the capacity to discipline private capital into abiding by that project.

China, by contrast, had only years after its modern founding fully penetrated and reorganized society. A major land reform that removed a traditional locus of power was quickly carried out, women were given the right to initiate divorce and hold land, and the remnants of the old *Guomindang* administration were integrated into the CCP state (Spence 1991: 516-519). The new state instituted extensive social control, with street-committees composed of neighbors responsible for social services, such as street-cleaning, health and vaccination programs, night schools, as well as public security (Spence 1991: 518). To fully penetrate society and administer the new state the CCP needed new members, but it wanted to ensure that they would be effective and reliable. Mass campaigns were used to identify true believers with an ability to lead, and undesirables were identified by cataloging all citizens into different social categories (Spence 1991: 533-540). Though the decades of chaos that culminated in the Cultural Revolution severely damaged the capacity of the party-state apparatus, the damage does not seem to have been irreparable. The generation of leaders after Mao were able to reorganize and arguably reinvigorate the party as well as the state (Shue 1994: 73; Spence 1991: 691-692).

Given China and India's diverging development performance and the qualitative evidence that since the respective founding China's political system has to a greater extent been able to formulate and implement development policy than India's, it could be expected that China's scores on various

governance indicators would be higher than India's. This is however not the case. Indeed, set against China's development performance its scores are inexplicably low "in all of the available measures of quality of government" (Rothstein 2012: 3). Compared to India's, China's scores are consistently worse on almost any given governance indicator even as its development performance is, and has been, consistently better (table 5).

	Absolute Scores		Scale	Higher Score is...	Rel. World Average	
	China	India			China	India
Human Development Index	0.66	0.52	0-1	Better	6%	-17%
BTI Management Index	5.00	6.60	1-10	Better	2%	34%
Failed State Index	84.60	77.80	0-120	Worse	-17%	-8%
ICRG Quality of Government	0.56	0.61	0-1	Better	5%	15%
TI Corruption Perceptions Index	3.60	3.40	0-10	Better	10%	15%
WGI Government Effectiveness	0.13	-0.02	-2.5-2.5	Better	8%	2%
WGI Average	-0.52	-0.30	-2.5-2.5	Better	-19%	-9%
State Capacity	0.89	0.82	0-1	Better	17%	8%

Table 5. According to most indicators there is little difference between China and India's governance, the difference being generally in India's favor. State capacity measured by tax compliance tells another story. In the same way that China's development level is decisively higher than India's, so is its tax compliance greater.

Source: Teorell et al. 2013; UNDP 2010.

Two conclusions can be drawn from China's performance on governance indicators. The first is that governance in China is not good. This conclusion is, however, highly problematic (Mahbubani 2013):

[China's government] is not perfect but it has lifted more people out of poverty, educated more people, increased their lifespans and generated the world's largest middle class. No other society in human history has improved human welfare as much as the Chinese government. It would be insane to deny that China has enjoyed "good governance".

The second, more defensible, conclusion that could be drawn is that these indicators are not able to capture some critical aspect of governance in general, and the Chinese political system in particular. Matt Andrews in an article about the GE argued that "the good governance agenda suggests a one-best-way model, ostensibly of an idyllic, developed country government: Sweden or Denmark on a good day, perhaps" (Andrews 2008: 379). From this perspective China's poor scores relative India's are understandable; the Indian political system has more in common with that of 'Sweden or Denmark on a good day' than the Chinese.

Of the governance indicators surveyed in table 5, only state capacity operationalized as tax compliance seems to show the sort of difference in capacity to provide public services that area studies and development performance suggest should exist. Normalizing all indicators, China decisively outperforms India on human development and tax compliance (figure 3). While the GE does show an appreciable difference in China's favor, the problems with measurement validity means that it is not clear what this represents. Furthermore, once the indicators have been normalized China's performance on GE relative that of India's is not nearly as large as the difference in development performance.

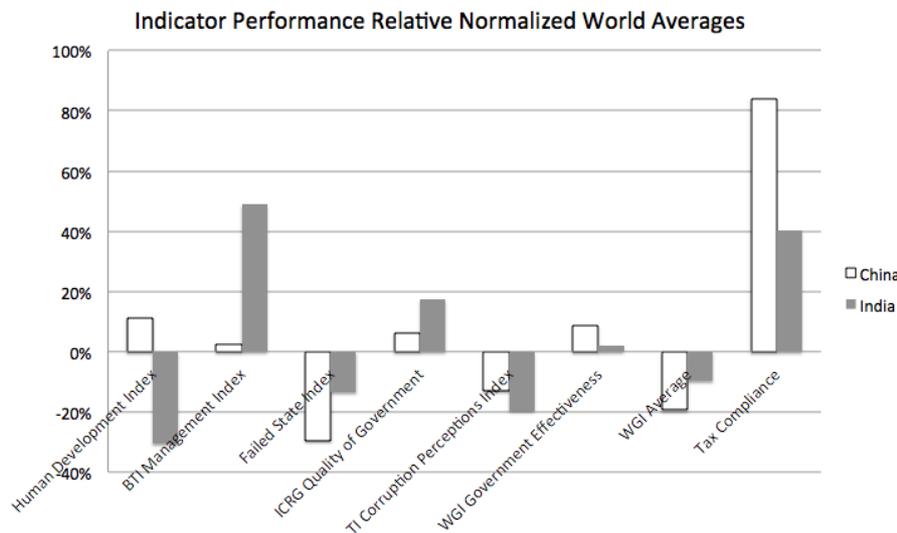


Figure 3. Tax compliance is unique among governance indicators in that it shows the significant governance capacity edge for China over India that qualitative evidence and development performance suggests.
Source: Teorell et al. 2013; UNDP 2010.

The size and complexity of China and India is such that this short overview clearly cannot do justice to the state-building or human development of the two countries. Regional, and ideally sub-regional, state capacity as well as development performance needs to be accounted for in a more comprehensive analysis. What this limited comparative case study can do, and seems to accomplish, is to corroborate a level of difference in the state capacity of the two countries that development performance and more qualitative studies suggest. Of the governance indicators shown, only tax compliance is able to account for this apparent difference in state capacity, and resultant development performance.

CONCLUSION

Using Adcock and Collier's 4-step framework as a guide to ensure measurement validity state capacity a systematized conceptualization of state capacity was derived from the background concept of the state. This systematized concept, with the state defined as *the ability of the state to dominate, i.e. coax compliant behavior from, the individuals of a given territory*, was then operationalized as tax compliance. Using the size of the shadow economy in 160 countries, the tax compliance was calculated. This operationalization was then tested in three ways. First, the convergent and discriminant validity of the state capacity indicator was tested. Second, the usefulness of state capacity in large-n output analyses, such as those in the good governance/quality of governance literature, was tested. Finally, the new indicator was used in a qualitative comparison between China and India where existing governance indicators are unable to capture what qualitative evidence and development performance suggests is a difference in state capacity in China's favor.

The first test suggests that state capacity as tax compliance has both convergent and discriminant validity. It correlates highly with existing measures of effective government, especially, GE, suggesting convergent validity. The correlation with output measures, such as human development, is weaker in the way theory would suggest it should be. The correlation with democracy is quite low outside of the West, which is also in accordance with theory. Furthermore, the relationship between state capacity and democracy is lower than for any of the other governance indicators. The lower relationships between tax compliance and an output measure and democracy supports the discriminant validity of the proposed indicator.

In the second test, wherein state capacity is used as an independent variable, the indicator also behaves as theory would suggest it should. It has a significant effect on development levels, especially as other, more diffuse governance indicators as well as historical development levels are controlled for. Furthermore, because of the high content validity, it is clear exactly what the state capacity indicator is a measure of. This separates it from other, more nebulous governance indicators such as the WGI's.

Finally, applied to the critical case of China, the proposed indicator is able to substantiate a difference in governance capacity between China and India that qualitative evidence and development performance suggest should be there, but existing governance indicators do not capture. While this

is of course not the conclusive solution to the 'Puzzle with China' (Rothstein 2012), it is suggestive of a solution.

The proposed indicator of state capacity, defined narrowly as dominance and operationalized as tax compliance, seems promising. It appears to have measurement validity, as well as being analytically useful. While this does not fully answer Fukuyama's call for a conceptualization of governance as a whole, it does, perhaps, offer a piece of the puzzle.

REFERENCES

- Adcock, Robert and David Collier. 2001. Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review* 95(3): 529-546.
- Apaza, Carmen R. 2009. Measuring Governance and Corruption through the Worldwide Governance Indicators: Critiques, Responses, and Ongoing Scholarly Discussion. *PS: Political Science & Politics* 42 (1): 139-143.
- Anderson, Benedict. 2006. *Imagined communities: reflections on the origin and spread of nationalism*. London: Verso.
- Andrews, Matt. 2007. The good governance agenda: beyond indicators without theory. *Oxford Development Studies* 36 (4): 379-407.
- Arndt, Christian. 2008. The politics of governance ratings. *International Public Management Journal* 11 (3): 275-297.
- Carothers, Thomas. 2002. The End of the Transition Paradigm. *Journal of Democracy* 13 (1): 5-21.
- Crafts, Nicholas. 1996. The Human Development Index: Some Historical Comparisons. *London School of Economics & Political Science: Working Papers in Economic History*. London: London School of Economics.
- Dalton, Russell J. 2004. *Democratic Challenges, Democratic Choices: The Erosion of Political Support in Advanced Industrial Democracies*. Oxford: Oxford University Press.
- Diamond, Larry Jay. 2000. Is Pakistan the (Reverse) Wave of the Future? *Journal of Democracy* 11(3): 91-106.
- Easton, David. 1957. An Approach to the Analysis of Political Systems. *World Politics* 9(3): 383-400.
- Evans, Peter, Dietrich Rueschemeyer and Theda Skocpol. 1985. *Bringing the State Back In*. Cambridge: Cambridge University Press.
- Fishman, Robert M. 1990. Review: Rethinking State and Regime: Southern Europe's Transition to Democracy. *World Politics* 42(3): 422-440.
- Fukuyama, Francis. 2005. Stateness First. *Journal of Democracy* 16(1): 84-88.
- Fukuyama, Francis. 2011. *The Origins of Political Order: From Prehuman Times to the French Revolution*. Farrar, Straus and Giroux. Kindle Edition.
- Fukuyama, Francis. 2013. What is Governance?. *Governance: An International Journal of Policy, Administration, and Institutions* 26(2): 347-362.
- Gerth, Hans and C. Wright Mills. 1946. *From Max Weber: Essays in Sociology*. New York : Oxford University Press.
- Hui, Victoria Tin-bor. 2005. *War and State-Formation in Ancient China and Early Modern Europe*. Oxford: Oxford University Press.

- Johnson, Chalmers. 1982. *MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925-1975*. Stanford: Stanford University Press.
- Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobaton. 1999. *Governance Matters*. World Bank Policy Research Working Paper No. 2196. Washington: The World Bank. <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=188568> (Accessed May 1, 2013)
- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010a. Response to 'What do the Worldwide Governance Indicators Measure?'. *European Journal of Development Research* 22 (1): 35–58.
- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010b. *The Worldwide Governance Indicators: Methodology and Analytical Issues*. World Bank Policy Research Working Paper No. 5430. Washington: The World Bank. <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1682130> (Accessed May 9, 2013)
- Kurtz, Marcus J. and Andrew Schrank. 2007. Growth and governance: Models, measures, and mechanisms. *Journal of Politics* 69 (2): 538-554.
- Langbein, Laura and Stephen Knack. 2010. The Worldwide Governance Indicators: Six, One, or None?. *Journal of Development Studies* 46 (2): 350-370.
- Lawson, Stephanie. 1993. Conceptual Issues in the Comparative Study of Regime Change and Democratization. *Comparative Politics* 25(2): 183-205.
- Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley: University of California Press.
- Levi, Margaret. 1997. *Consent, Dissent, and Patriotism*. Cambridge: Cambridge University Press.
- Levi, Margaret, Audrey Sacks and Tom Tyler. 2009. Conceptualizing Legitimacy, Measuring Legitimizing Beliefs. *American Behavioral Scientist* 53(3): 354-375.
- Lieberman, Evan S.. 2002. Taxation data as indicators of state-society relations: possibilities and pitfalls in cross-national research. *Studies in Comparative International Development (SCID)* 36 (4): 89–115.
- Mann, Michael. 1984. The Autonomous Power of the State: Its Origins, Mechanisms, and Results. *European Journal of Sociology* 25 (2): 185–213.
- McAdam, Doug, Sidney Tarrow, and Charles Tilly. 2001. *Dynamics of Contention*. Cambridge: Cambridge University Press.
- Migdal, Joel S. 1988. *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World*. Princeton, NJ: Princeton University Press.
- Migdal, Joel S.. 1994. The state in society. In *State Power and Social Forces*. Edited by Joel S. Migdal, Atul Kohli, and Vivienne Shue. Cambridge: Cambridge University Press.
- Migdal, Joel S. 2001. *State in Society: Studying How States and Societies Transform and Constitute One Another*. Cambridge: Cambridge University Press.
- Norris, Pippa. 1999. Introduction: The Growth of Critical Citizens?. In *Critical Citizens*, edited by Pippa Norris. Oxford: Oxford University Press

- Persson, Anna. 2008. *The Institutional Sources of Statehood: Assimilation, Multiculturalism, and Taxation in Sub-Saharan Africa*. Gothenburg: University of Gothenburg.
- Peters, B. Guy, and Jon Pierre. 2006. Governance, Government and the State. In *The State: Theories and Issues*, edited by Colin Hay, Michael Lister, and David Marsh. New York: Palgrave Macmillan.
- Pollitt Christopher. 2009. 'Simply the best? The international benchmarking of reform and good governance'. In *Public sector administrative reform and the challenges of effective change*, edited by Jon Pierre and Patricia W. Ingraham. Montreal and Kingston: McGill-Queens University Press.
- Rotberg, Robert I.. 2003. Failed States, Collapsed States, Weak States: Causes and Indicators. In *State Failure and State Weakness in a Time of Terror*, edited by Robert I. Rotberg. Washington D.C.: Brookings Institution Press.
- Rotberg, Robert I.. 2004. The Failure and Collapse of Nation-States: Breakdown, Prevention, and Repair. In *When States Fail: Causes and Consequence*, edited by Robert I. Rotberg. Princeton: Princeton University Press.
- Rothstein, Bo. 2009. Creating Political Legitimacy: Electoral Democracy Versus Quality of Government. *American Behavioral Scientist* 53(3): 311–30.
- Rothstein, Bo. 2011. *The Quality of Government: Corruption, Social Trust, and Inequality in International Perspective*. Kindle Edition.
- Rothstein, Bo. 2012. *Understanding the Quality of Government in China: The Cadre Administration Hypothesis*. Working Paper Series 2012:17. University of Gothenburg: The Quality of Government Institute.
- Rothstein, Bo, and Dietlind Stolle. 2008. The State and Social Capital: an Institutional Theory of Generalized Trust. *Comparative Politics*.
- Rueschemeyer, Dietrich and Peter Evans. 1985. The State and Economic Transformation: Toward an Analysis of the Conditions Underlying Effective Intervention. In *Bringing the State Back In*, Edited by Peter Evans, Dietrich Rueschemeyer and Theda Skocpol. Cambridge: Cambridge University Press.
- Schneider, Friedrich and Markus Hofreither. 1986. Measuring the Size of the Shadow Economy: Can the Obstacles Be Overcome? *Economic Affairs* 7(2): 18–23.
- Schneider, Friedrich, Andreas Buehn, and Claudio E. Montenegro. 2010. New Estimates for the Shadow Economies all Over the World. *International Economic Journal* 24(4): 443–461.
- Schneider, Friedrich and Dominik H. Enste. 2013. *The Shadow Economy: An International Survey*. Cambridge: Cambridge University Press.
- Schumpeter, Joseph. 1991. The Crisis of the Tax State. In *The Economics and Sociology of Capitalism*. Edited by Richard Swedberg. Princeton: Princeton University Press.
- Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.

- Teorell, Jan and Axel Hadenius. 2005. *Determinants of Democratization: Taking Stock of the Large-N Evidence*. mimeo. Uppsala University: Department of Government.
- Teorell, Jan, Nicholas Charron, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Petrus Sundin & Richard Svensson. 2013. The Quality of Government Dataset, version 15May13. University of Gothenburg: The Quality of Government Institute, <http://www.qog.pol.gu.se>.
- Thomas, M.A.. 2010. What Do the Worldwide Governance Indicators Measure? *European Journal of Development Research* 22 (1): 31–54.
- Tilly, Charles. 1975. Reflections on the History of European State-Making. In *The Formation of National States in Western Europe*, edited by Charles Tilly. Princeton University Press: Princeton.
- Tilly, Charles. 1992. *Capital, Coercion, and European States*. Oxford: Blackwell Publishing.
- UNDP. 1990. *Human Development Report 1990*. New York: United Nations Development Programme. <<http://hdr.undp.org/en/reports/global/hdr1990/chapters>> (Accessed Feb. 14, 2013)
- UNDP. 2010. *Human Development Report 2010*. New York: United Nations Development Programme. <<http://hdr.undp.org/en/reports/global/hdr2010/chapters/>> (Accessed Mar. 24, 2013)
- Wang, Shaoguang. 2001. The Construction of State Extractive Capacity: Wuhan 1949-1953. *Modern China* 27 (2): 229-261.
- Wang, Shaoguang. 2003. The Problem of State Weakness. *Journal of Democracy* 14(1): 36-42.
- Wang, Shaoguang and Hu Angang. 2001. *The Chinese Economy in Crisis: State Capacity and Tax Reform*. New York: M.E. Sharpe.
- Weber, Max. 1968. *Economy and Society: An Outline of Interpretive Sociology*. Edited by Guenther Roth and Claus Wittich. Berkeley: University of California Press.
- Weber, Eugen. 1976. *Peasants into Frenchmen : the Modernization of Rural France 1870-1914*. Stanford: Stanford University Press.

Appendix A: 2003-2007 Average of Shadow Economy and Tax Compliance

No.	Country	Shadow Economy	Tax Compliance
1	Switzerland	8.460	0.922
2	United States	8.520	0.921
3	Luxembourg	9.660	0.912
4	Austria	9.700	0.912
5	Japan	10.700	0.903
6	Macao, China	11.860	0.894
7	New Zealand	12.080	0.892
8	United Kingdom	12.360	0.890
9	China	12.400	0.890
10	Singapore	12.640	0.888
11	Netherlands	13.180	0.884
12	Australia	13.700	0.880
13	Vietnam	14.800	0.871
14	France	14.840	0.871
15	Iceland	15.300	0.867
16	Canada	15.480	0.866
17	Hong Kong, China	15.500	0.866
18	Ireland	15.660	0.865
19	Germany	15.860	0.863
20	Mongolia	17.060	0.854
21	Finland	17.360	0.852
22	Bahrain	17.433	0.852
23	Denmark	17.460	0.851
24	Saudi Arabia	17.520	0.851
25	Slovak Republic	17.600	0.850
26	Czech Republic	17.840	0.849
27	Iran, Islamic Rep.	17.840	0.849
28	Jordan	17.940	0.848

29	Oman	18.075	0.847
30	Sweden	18.380	0.845
31	Norway	18.440	0.844
32	Quatar	18.467	0.844
33	Kuwait	18.525	0.844
34	Indonesia	18.540	0.844
35	Chile	18.920	0.841
36	Syrian Arab Republic	18.920	0.841
37	India	21.580	0.823
38	Belgium	21.660	0.822
39	Israel	21.700	0.822
40	Mauritius	22.320	0.818
41	Spain	22.380	0.817
42	Portugal	23.120	0.812
43	Hungary	23.980	0.807
44	Taiwan	24.500	0.803
45	Argentina	24.680	0.802
46	United Arab Emirates	25.000	0.800
47	Costa Rica	25.320	0.798
48	Slovenia	25.680	0.796
49	Korea, Rep.	26.220	0.792
50	South Africa	26.520	0.790
51	Bahamas, The	26.700	0.789
52	Poland	26.820	0.789
53	Yemen, Rep.	26.840	0.788
54	Greece	26.860	0.788
55	Italy	26.960	0.788
56	Malta	27.180	0.786
57	Cyprus	27.560	0.784
58	Bhutan	28.320	0.779
59	Latvia	28.340	0.779
60	Lao PDR	28.900	0.776

61	Maldives	29.120	0.774
62	Namibia	29.460	0.772
63	Mexico	29.700	0.771
64	Lesotho	29.820	0.770
65	Turkey	30.280	0.768
66	Malaysia	30.380	0.767
67	Estonia	30.460	0.767
68	Equatorial Guinea	30.640	0.765
69	Lithuania	30.960	0.764
70	Brunei Darussalam	30.980	0.763
	World Average	32.488	0.762
71	Ecuador	31.200	0.762
72	Croatia	31.260	0.762
73	Romania	31.480	0.761
74	Algeria	31.500	0.760
75	Cameroon	31.540	0.760
76	Dominican Republic	31.540	0.760
77	Fiji	31.880	0.758
78	Kenya	32.320	0.756
79	Botswana	32.540	0.754
80	Lebanon	32.560	0.754
81	Trinidad and Tobago	32.560	0.754
82	Libya	32.960	0.752
83	Bosnia & Herzegovina	33.200	0.751
84	Jamaica	33.540	0.749
85	Solomon Islands	33.560	0.749
86	Albania	33.640	0.748
87	Venezuela, RB	33.640	0.748
88	Guyana	33.880	0.747
89	Morocco	34.060	0.746
90	Sudan	34.100	0.746
91	Bulgaria	34.160	0.745

92	Mauritania	34.250	0.745
93	Egypt, Arab Rep.	34.480	0.744
94	Pakistan	34.760	0.742
95	Togo	34.875	0.741
96	Cape Verde	34.880	0.741
97	Bangladesh	34.960	0.741
98	Colombia	35.940	0.736
99	Tunisia	36.460	0.733
100	Suriname	36.500	0.733
101	Nepal	36.540	0.732
102	Macedonia	36.720	0.731
103	Papua New Guinea	36.980	0.730
104	Ethiopia	37.560	0.727
105	Brazil	38.200	0.724
106	Paraguay	38.250	0.723
107	Comoros	38.480	0.722
108	Guinea	38.760	0.721
109	Mozambique	39.250	0.718
110	Burundi	39.620	0.716
111	Swaziland	39.625	0.716
112	Niger	39.675	0.716
113	Ghana	39.720	0.716
114	Kazakhstan	39.760	0.716
115	Kyrgyz Republic	39.800	0.715
116	Rwanda	39.825	0.715
117	Burkina Faso	39.880	0.715
118	Mali	40.080	0.714
119	Philippines	40.300	0.713
120	Eritrea	40.800	0.710
121	Madagascar	40.820	0.710
122	Tajikistan	41.500	0.707
123	GuineaBissau	41.640	0.706

124	Uganda	41.680	0.706
125	Malawi	41.840	0.705
126	Belize	42.020	0.704
127	Chad	42.140	0.704
128	Russian Federation	42.260	0.703
129	Armenia	42.680	0.701
130	Senegal	42.800	0.700
131	Gambia, The	43.220	0.698
132	Sri Lanka	43.240	0.698
133	Nicaragua	43.920	0.695
134	Sierra Leone	44.000	0.694
135	Moldova	44.075	0.694
136	El Salvador	44.280	0.693
137	Liberia	44.800	0.691
138	Angola	45.100	0.689
139	Congo, Rep.	45.120	0.689
140	Belarus	45.160	0.689
141	Zambia	45.900	0.685
142	Central African Republic	46.420	0.683
143	Cote d'Ivoire	46.440	0.683
144	Congo, Dem. Rep.	46.860	0.681
145	Honduras	47.140	0.680
146	Gabon	47.700	0.677
147	Cambodia	47.720	0.677
148	Ukraine	48.080	0.675
149	Myanmar	48.633	0.673
150	Thailand	49.100	0.671
151	Benin	49.460	0.669
152	Guatemala	49.660	0.668
153	Uruguay	49.700	0.668
154	Nigeria	54.550	0.647
155	Tanzania	55.280	0.644

156	Azerbaijan	56.080	0.641
157	Peru	56.620	0.638
158	Haiti	57.000	0.637
159	Panama	62.400	0.616
160	Zimbabwe	62.600	0.615
161	Georgia	64.440	0.608
162	Bolivia	65.040	0.606