LUND UNIVERSITY

The Lay of the Land: Information Capacity and the Modern State

Brambor, Thomas; Goenaga, Agustín; Lindvall, Johannes; Teorell, Jan

2016

Document Version:
Other version

Link to publication

Citation for published version (APA):
Brambor, T., Goenaga, A., Lindvall, J., & Teorell, J. (2016). *The Lay of the Land: Information Capacity and the Modern State*. (STANCE Working Paper Series; Vol. 2016, No. 2). Department of Political Science, Lund University.

Total number of authors:
4

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

STANCE

State-Making and the Origins of Global Order
in the Long Nineteenth Century and Beyond

The Lay of the Land:
Information Capacity and the Modern State

Thomas Brambor, Agustín Goenaga,
Johannes Lindvall, and Jan Teorell

Working Paper Series, 2016:2
STANCE, Lund University

LUND
UNIVERSITY

STANCE is a six-year research program at the Department of Political Science at Lund University, Sweden. The program, consisting of several separate but connected research projects, aims to answer the question of how state-making and the international system co-evolved in the long 19th century (1789-1914) and beyond. The program is constructed around three research themes: (1) How did the different dimensions of state-making evolve? What actors and organized interests supported or put up resistance to these processes?; (2) How were these dimensions of state-making affected by geopolitical competition, warfare and the diffusion of novel political technologies?; and (3) What were the consequences for the international system, both with respect to the type of state that emerged and what entities were granted membership in the state system? The program aims to bridge the gaps between comparative politics and IR, as well as those between the study of political thought and positive empirical political science. The research has been made possible by the Bank of Sweden Tercentenary Foundation (Riksbankens Jubileumsfond). Visit the research program's website at www.stanceatlund.org

Please address comments and/or queries for information to:
Email address: info@stanceatlund.org
Mailing address: STANCE
Department of Political Science
Lund University
Box 52, SE-221 00 Lund, Sweden

In Series 2016:

1. "STATE CAPACITY AS POWER: A CONCEPTUAL FRAMEWORK", Johannes Lindvall and Jan Teorell
2. "THE LAY OF THE LAND: INFORMATION CAPACITY AND THE MODERN STATE", Thomas Brambor, Agustín Goenaga, Johannes Lindvall, and Jan Teorell

STANCE working papers are available in electronic format at www.stanceatlund.org

COPYRIGHT © 2016 by authors. All rights reserved.

The Lay of the Land:

Information Capacity and the Modern State[1]

Thomas Brambor
Post Doc in Political Science
Lund University

Agustín Goenaga
Post Doc in Political Science
Lund University

Johannes Lindvall
Professor of Political Science
Lund University

Jan Teorell
Professor of Political Science
Lund University

Abstract

Relying on three new indicators of the information capacity of states, this paper provides new evidence on the ability of states to collect and process information about the territories and populations that they govern. The three indicators are (a) the availability of a reliable census, (b) the establishment of a permanent government agency tasked with processing statistical information about the territory and the population, and (c) the regular release of statistical yearbooks. We find, as expected, that there has been a secular increase in information capacity over time. We also investigate salient differences among countries from the early 1800s onward.

[1] This paper was prepared for presentation at the STANCE conference in Lund on May 19–20, 2016. The data and computer code that are required to replicate the tables and figures in this paper will be made publicly available via the websites of the authors once the paper is published. Stata was the statistical package used in this study. All four authors are affiliated with the Department of Political Science at Lund University and can be contacted at thomas.brambor@svet.lu.se, agustin.goenaga@svet.lu.se, johannes.lindvall@svet.lu.se, and jan.teorell@svet.lu.se. We are grateful to Moa Olin for her excellent work on data collection.

I Introduction

It is now widely accepted among political scientists, economists, and sociologists that if we wish to explain enduring differences in economic and political development among countries—and within countries over time—we first need to understand how states attain the ability to carry out government policies effectively. This ability, which is commonly referred to as "state capacity," has been defined as "a government's ability to make and enforce rules, and to deliver services" (Fukuyama 2013, 350), the "institutional capability of the state to carry out various policies" (Besley and Persson 2011, 6) and 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). State capacity is typically treated as "a function of state bureaucracy, the state's relations with social actors, and its spatial and societal reach" (Soifer and vom Hau 2008, 220).

Of all the resources that states use when they "make and enforce" rules, "carry out" their policies, and "control" persons, activities, and resources, the most important is information. On the basis of a theoretical argument about the relationship between state capacity and information (Section 2), this paper develops a new measure of the ability of states to collect, retrieve, and process information. We rely on systematic comparative and historical data on three information- gathering activities: the introduction of a national census (Section 3), the establishment of a permanent state agency tasked with processing statistical information about the territory and the population, and the publication of statistical yearbooks (Section 4). We then analyze the empirical relationships between these three different components, and present evidence regarding the economic and political correlates of "information capacity" (Section 5).

II Information and State Capacity

Measuring State Capacity. The main limitation of most existing measures of state capacity is that these measures risk conflating the *capacity* to "make and enforce" rules, "carry out" policies, and "control" populations and territories with *these outcomes themselves*.

For example, Soifer (2012) has proposed a measure of state capacity that is based on the idea that state capacity has three constitutive dimensions: coercive, fiscal, and administrative capabilities. He uses a combination of surveys and social statistics to measure likely *outcomes* of these capabilities: violent-crime rates, lynching rates, and

private- security expenditures for coercive capabilities; census administration, national-identity-card registration and vaccination rates for administrative capabilities; and direct taxes as a proportion of taxes, direct taxes per capita and the share of the population working in the formal sector for fiscal capabilities.

Like Soifer, Hanson and Sigman (2013), distinguish between "extractive," "coercive," and "administrative" capacity, but unlike Soifer's, Hanson and Sigman (2013)'s empirical strategy—which is more inductive—mixes outcome measures with survey-based measures of the efficiency and effectiveness of bureaucratic institutions (as well as some measures that are closely related to the measures of "information capacity" that we develop in this paper). However, Hanson and Sigman's aggregated measure of state capacity does rely in part on measures of outcomes, such as tax evasion.

In fact, there are very few available measures of state capacity that do *not* rely, at least partly, on data on policy outputs and outcomes. Expert-survey-based indicators of the quality of political institutions and state bureaucracies are probably not an exception: although such measures are typically not presented as output- or outcome-based, it seems highly likely that the expert respondents that provide the country-level data make inferences from outcomes when they give their answers to questions about capacities.

Information as a Resource. We concentrate on "inputs," not "out- puts." At the most general level, the term state capacity refers to the extent to which a state is able to achieve the outcomes that the government that controls the state attempts to achieve (whether it is to "make and enforce" rules, "carry out" policies, and "control" populations and territories). To fix ideas, we can refer to these "attempts to achieve" something as a vector of policies, $\mathbf{p}$, and the intended "out- comes" as changes in a vector of individual preferences or behavior, $\mathbf{y}$ (we use boldface to denote vectors). Following Lindvall and Teorell (2016), we can think of state capacity as the strength of the causal relationship between $\mathbf{p}$ and $\mathbf{y}$: when the government of a high-capacity state decides to adopt the policy p in order to achieve the outcome y, it is more likely to be successful than the government of a low-capacity state would be, if it adopted the same policy.

The reason that high-capacity states are more likely to produce the outcomes that the government wants to achieve is that they deploy *resources*—which we denote $\mathbf{r}$—in order to increase the likelihood that government policies ($\mathbf{p}$), achieve the intended outcomes ($\mathbf{y}$). In other words, the causal relationship between $\mathbf{p}$ and $\mathbf{y}$ is conditioned by $\mathbf{r}$.

Since state capacity, in this framework, is a *causal effect*, it cannot be measured (it can only be estimated). But the resources that states deploy—the elements of $\mathbf{r}$—can be measured. (The analogy with the general literature on political power is obvious: power cannot be measured directly; power resources can.) In our view, the way

forward for the literature on state capacity is therefore to measure resources, or "state capacities" (Skocpol 1985).

The resources that we measure in order to understand the main sources of state capacity must be carefully selected. Most importantly, they should be as fungible as possible (as transferable as possible from one area of state action to another), and their use should be as independent of a particular state leader's preferences as possible. The ability to collect, process and publish reliable *information* about "persons, activities, and resources within their government's territorial jurisdiction" (McAdam, Tarrow, and Tilly 2001, 78) is, in our view, the most important and general resource that modern states rely on (although certainly not the only one).

Most of what states do requires information of some sort. Start with what Scott (1998, 2) calls the "classic state functions of taxation, conscription and prevention of rebellion." As argued by Soifer (2013, 9), the power needed to "assess and collect taxes" depends on information about the population, such as that collected through a census or cadastral maps (D'Arcy and Nistotskaya 2015). Similarly, to effectively implement the draft, authorities need information on what parts of the population are eligible, and where to find them. According to Fearon and Laitin (2003, 80), moreover, a state's capacity to avoid civil war de- pends on an effective counter-insurgency strategy, a critical component of which is information about "goings-on at the local level." But the key role played by information in explaining the success of state policies goes beyond these three core functions. States may help promote growth by providing reliable information, for example about possible business partners (Evans and Rauch 1999). Similarly, successful public service delivery depends, in part, on reliable information about needs, and on monitoring when these needs are not met in the general population (Hanson 2015; Lee and Zhang 2016).

Another reason to believe that the collection of information is one of the most fundamental and fungible resources that the state commands is that successful information-gathering depends on geographical penetration. Unlike other features that mainly pertain to the central state in the capital—such as the financial resources that are available to state leaders or the number of civil servants working for the central state bureaucracy—the ability to gather reliable information requires spatial reach (Soifer 2008, 2012; Herbst 2014).

In sum, we argue that the collection and processing of reliable information on population and territory—including economic production, consumption, and social organization—are key aspects of modern state administration, and that "state capacity" more generally can in several important ways be linked to the ability of the state to obtain and use high-quality information.

Related Literatures. Our project is most closely related to that of Lee and Zhang (2016), who develop a theoretical analysis of the relation- ship between "legibility" and state capacity, and who introduce a new measure of legibility that is based on the accuracy of age data in national population censuses during the period between 1960

and 2014. Lee and Zhang (2016) draw on the pioneering work of Scott (1998), who introduced the concept of "legibility" and who identified the relationship between the growth of the modern state and attempts by state officials to render societies "legible" by producing standardized information about territories and populations. Our contribution, in relation to Lee and Zhang's work, is twofold: we examine several different types of information (not just the quality of the census), and we cover a much longer time period: in addition to the first part of the post-war period), we cover the 150-year period between the French Revolution in 1789 and the Second World War.

Many other scholars, in addition to Scott, have noted that the capacity to process information is a key driver in the increase in the power of the modern state in the late-modern period. Laurence Whitehead (1995) has written about the importance of information in the context of Latin American states. He defines "cognitive capacity" as the "sustained organization to collect, process, analyse and deliver the types of information about society needed for a modern state to monitor and interpret the impact of its measures, and to adjust them or reformulate them when they prove ineffective or counter-productive" (Whitehead 1995, 46–47). And in the prolegomena to his great study of social power, Mann (1984) noted that one of the most important elements of the state's infrastructural power is information: "it [the state] stores and can recall immediately a massive amount of information about all of us" (189).

When it comes to other attempts to measure information capacity directly, we should mention that the World Bank—as a part of the World Development Indicators project—has gathered data on the "statistical capacity" of contemporary states from 2004 to the present. The World Bank defines statistical capacity as "a nation's ability to collect, analyze, and disseminate high-quality data about its population and economy," noting that high-quality statistics "are essential for all stages of evidence-based decision-making, including: monitoring social and economic indicators, allocating political representation and government resources, guiding private sector investment, informing the international donor community for program design and policy formulation" (datatopics.worldbank.org/statisticalcapacity). The World Bank's Statistical Capacity Indicator (SCI) is based on a diagnostic framework that assesses the methodology, data sources, and periodicity of official figures.

The index that we present in this paper is an attempt to develop a similar measure that spans a much longer time period, in order to asses the historical evolution of information capacity.

III The Census

Before we discuss the first information-gathering activity that we consider, the census, we would like to discuss the basic design of our study; in particular, we will explain how we have selected the countries that are included in our main sample. Our sample includes 85 countries chosen so as to include all those that were (a) sizeable (> 250, 000), (b) sovereign in the pre-1900 era (either in the formal-juridical sense or the *de facto* sense), and that (c) match present-day state units, using Gledistch & Ward (1999) as our main point of departure.[2]

Population counts are the oldest and simplest technology of information-gathering available to rulers. They do not require major infrastructural resources, technical sophistication, or political coordination. Archaeological evidence of population counts for fiscal or military purposes goes as far back as Babylon (3,800 B.C.), and has been found in ancient Egypt, Israel, China, Greece, and Rome (Alterman 1969, 17–35). During the Middle Ages in Europe, parishes typically kept track of the population, and as a result population counts rarely went beyond the local level. Nevertheless, there are examples of monarchs who ordered population counts, almost as inventories of the physical and human assets they commanded, such as Charlemagne's Capitolorum in France (808 A.D.) or William the Conqueror's *Domesday Book* in England (1085) (Alterman 1969, 42-43).

By contrast, "modern" censuses represent a highly sophisticated technology that only emerged by the turn of the nineteenth century. A census can be defined as "a direct enumeration, preferably on a set date and by name, of each individual in a census area" (Ventresca 1995). In other words, seven features distinguish modern censuses from other forms of population accounting.

---

[2] This is the sample of the "Historical Varieties of Democracy (V-Dem)" project, which with some exceptions plus additions comprises the set of 74 states (counting Great Colombia and Colombia as one) that gained independence prior to 1900 according to the authoritative list provided by Gledistch & Ward (1999). Five Gledistch & Ward (1999) states were omitted, two because they have hitherto not been funded for data collection by the twentieth-century V-Dem project (Luxembourg and Oman), and three because they do not match any contemporary state entities (Orange Free State, Transvaal, and Zanzibar). In turn, Historical V-Dem covers an additional 16 state entities not covered by Gledistch & Ward (1999). The inclusion of these additional countries is based on three different criteria. First, we include seven states that match contemporary states and that wielded enough *de facto* domestic sovereignty to be treated as a state unit despite the fact that their international sovereignty, and hence status a "independent state," was compromised prior to 1900 (Norway, Finland, Australia, New Zealand, Poland, Hungary and Yemen). This group also includes two "precursor" states of contemporary states whose borders do not quite fit the latter (Nejd/Saudi Arabia, Bukhara/Uzbekistan). Second, we include the two largest colonies by population: British India and the Dutch West Indies (Indonesia), both of which had considerable domestic autonomy. Third, due to a particular extra funding source, we include five pre-unification German principalities that did not meet Gleditsch & Ward's (1999) population criterion (Brunswick, Hamburg, Oldenburg, Nassau and Saxe-Weimar). In total, this means we are covering 85 state units (74 - 5 + 16 = 85).

(1) *Sponsorship*. In a modern census, an administrative body is granted legal authority to conduct the census, which means that the population is legally required to respond to the questions of the enumerators.

(2) *Defined territory*. The enumeration corresponds to a predefined geographical territory, and any changes in the territory covered by subsequent census iterations need to be clearly stated.

(3) *Universality*. The census does not produce an estimated value based on a sample of the population, but includes every person (however defined) in the territory, without omission or duplication.

(4) *Individual enumeration.* The characteristics of each individual are recorded separately; in other words, the census must generate a horizontal data set where rows represent individuals and columns represent individual-level characteristics.

(5) *Simultaneity and specified time*. The enumeration occurs with-in a bounded time-period, where each individual's characteristics are recorded as simultaneously as possible.

(6) *Periodicity*. The census is conducted at regular intervals in order to produce a dynamic picture of demographic changes.

(7) *Compilation and publication*. The census results are made public as soon as possible, not as raw data but through summary statistics (Goyer and Domschke 1983, 1–3).

This new technology for "reading" populations was a product of the transformation in the relationship between rulers and subjects that characterized the emergence of nation states. During the nineteenth century, states went beyond providing national defense and public safety, and became concerned with managing populations and fostering economic growth. In contrast to the extensive power of empires, nation-states sought to exert more intensive power on their subjects, regulating their behavior in a growing number of areas of social life (Mann 2012, 6, 81, 489). In this context of political transformation, modern censuses were both a consequence of the intensification of political power and an instrument that states could use to further enhance their power over their subjects.

Population censuses are a crucial resource that empowers states to carry out other functions. The ways in which modern censuses contributed to the expansion of the coercive and fiscal dimensions of state capacity are self-evident. First, the census offered state actors a clear sense of the distribution of different social groups in the territory, and this was essential to estimate the time and resources that it would take to mobilize that manpower in the case of war. More importantly, modern enumerations gave states knowledge about the general location of most individuals. This information, which in the past only local authorities possessed (by virtue of being immersed in those communities), was now accessible to national governments, contributing to the process of centralization of political authority.

Similarly, manufacturing and occupational censuses that sought to render economic production legible to fiscal eyes quickly followed the implementation of

modern population censuses (Fishbein 1973). These economic censuses made possible the modernization of fiscal systems through the introduction of income and sales taxes, and created the opportunity for states to influence the economic structure of their societies, to accelerate the process of industrialization, and to identify and address market failures.

The ability to obtain individualized data on a number of categories gave the state access to some of the most intimate areas of human activity. Michael Mann notes how questions about female fertility, illegitimacy rates or age of marriage could then be used to inform policy on family arrangements, parental responsibilities, and sexual morality (2012, 488-489). Managing populations became part of the purview of states, which were now concerned with the "health" of the population (Foucault 1978). Eugenics ultimately represented the most extreme manifestation of these heightened capabilities (Hansen and King 2013). In this way, modern censuses not only empowered the state to directly intervene in the bodies of its subjects, but they also created sharp distinctions between normal and deviant identities and behavior (Curtis 2002).

Finally, the modern census not only contributed to the rationalization of the administrative apparatus of the state, but also contributed to the rationalization of political authority by becoming the main point of reference for the apportionment of public resources. In other words, censuses not only offered new tools to enhance the infrastructural power of the state, but could also be used by society to tame arbitrariness and despotism. According to Kathrin Levitan, for instance, many proponents of the British census in the early nineteenth century "were skeptical of government motives and saw statistics as a useful safe- guard against governmental corruption" (Levitan 2011, 5). Indeed, the motivation behind the American census of 1790 was not to increase the degree of surveillance and control of the federal government over society but to determine the fair distribution of state representatives in Congress (Anderson 1988).

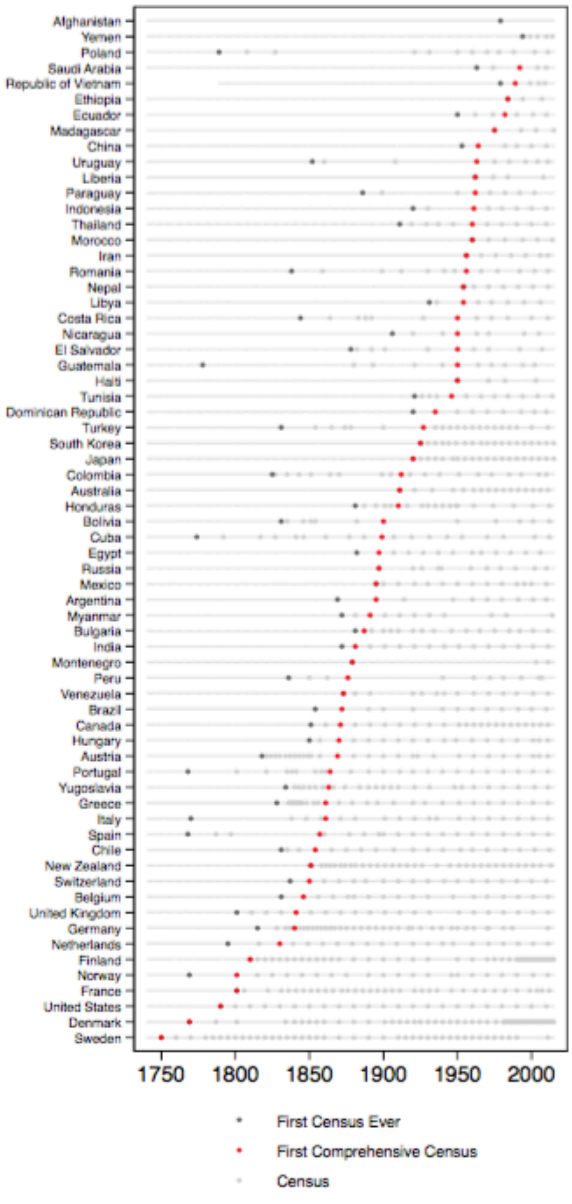
Figure 1 presents a timeline of the adoption of the modern census in our sample of countries. As mentioned before, modern censuses were characterized by the individual enumeration of every person in the territory, and this made them qualitatively different from earlier forms of counting populations. The enumeration of New France in 1666 and the Icelandic census of 1703 used canvassing methods to record every individual in the territory over a relatively short period of time. Nevertheless, since they covered fairly small jurisdictions (3,215 in New France, around 50,000 in Iceland) that were under the authority of foreign powers (France and Denmark, respectively), these enumerations are not typically counted as "national" censuses (Goyer and Domschke 1983, 5).

Sweden's census of 1749 represents the first modern census of a populous nation. There is some controversy on this point, however, since it was not based on door-to-door canvassing or self-enumeration methods, but rather relied on ecclesiastical registers of vital statics (births, deaths, marriages) (Goyer and Domschke 1983, 5).

Nevertheless, we treat it as a "modern" census since parishes produced tabular compilations that followed the principles of individual enumeration and universality (Centralbyrån 1969). On the other hand, the proto-census of the Kingdom of Denmark-Norway of 1769 mirrored the decentralized administrative strategy of its Swedish predecessor, but parishes did not produce tabular compilations, only summary data that did not record individual characteristics (Goyer and Draaijer 1992, 361). It was only with the census of 1787 that both principles of individual enumeration and universality were met in Denmark.

By the end of the eighteenth century, the modern census gained popularity through the diffusion of democratic institutions in the context of the American and French revolutions. As mentioned before, the first census in the United States was implemented in 1790 with the goal of defining the apportionment of political representation. France and England carried out comprehensive population counts in 1801, and produced individual enumerations shortly after.

FIGURE 1. Introduction, Census



The first non-European state, other than the United States, to implement a modern census was Chile in 1854. After two failed attempts in 1831 and 1843 that yielded incomplete results, the census of 1854 achieved a fairly complete coverage of the population (although indigenous peoples and inhabitants of remote areas were excluded) (Goyer and Domschke 1983, 115-117). Several non-European countries also began to implement modern censuses during the second half of the nineteenth century: Canada, Brazil, Venezuela, Peru, India, Burma, Argentina, Mexico, Russia, Egypt, etc.

This diffusion seems to have been the result of two interrelated processes that took place at the time. First, the internationalization of standards in the implementation of censuses began in 1853 with the First International Statistical Congress and gained traction in subsequent meetings of the International Statistical Institute in 1872 and 1897 (Ventresca 1995). This allowed late adopters to import the know- how of census implementation from abroad.

Second, the political leaders of the newly independent countries of Latin America were enthusiastic adopters of these new information technologies as they allowed them to consolidate their internal and external sovereignty. On the one hand, investing in the information capacity of the state offered them a valuable resource to establish political order and impose their rule over the population. On the other hand, the ability to carry out modern censuses suggested a family resemblance with the modern nations of Europe and North America, giving credibility to Latin American leaders in the eyes of the international community (Loveman in Centeno and Ferraro 2013).

IV Statistical Agencies and Yearbooks

To collect, organize, and preserve, and ultimately publish official statistics, states needed to develop dedicated bureaucratic structures to execute these tasks. The establishment of an official statistical agency is perhaps the most directly observable outward sign of the state's investment in "information capacity." Of course, some numerical in- formation about people, land, and production were often collected be- fore these statistical offices were established, as discussed in the previous section on the efforts to obtain information about the population through censuses. But the very fact that a specialized agency with the sole task of performing statistical activities is created certainly signifies the increased effort in expanding the ability of the state to gather statistical information and its raised significance for government activity more generally.

For the purposes of this paper, we code the establishment of the first independent statistical agency—that is, an official government organization exclusively devoted to gathering numerical information in a variety of subjects about the country — in each

country in our sample. Figure 2 presents an overview of the timing of the introduction of statistical agencies as well as the publication of official statistical yearbooks (which are discussed further below).

The Swedish *Tabellverket* was probably the first such official statistics institution in the world, created in 1756 by Pehr Wilhelm Wargentin (Arosenius 1918, 742), only a few years after the first nationwide census was conducted in Sweden in 1749.[3] But Sweden was an early pioneer in this regard, albeit with a meager budget for its statistical office. The proliferation of lasting statistical agencies—along with ef- forts to specialize, professionalize, and autonomize their services—did not start in earnest until the 1830s and 1840s.

A few attempts to follow Sweden's lead were made in several countries a bit earlier than that, but they fizzled out for various reasons. Denmark and Norway—at the time two kingdoms in personal union— had established a Tabulating Office in 1797 with the hope of improving on the disappointing quality of its earlier censuses as well the revision and analysis of its public accounts. The poorly staffed office was unable to live up to these high expectations, in particular with the failure to complete the enumeration of 1801, and it was abolished in 1819. In 1834, Denmark set up a Tabulating Commission, consisting of high- ranking administrators from various departments, which helped to establish a respected office and facilitated cooperation across agencies.

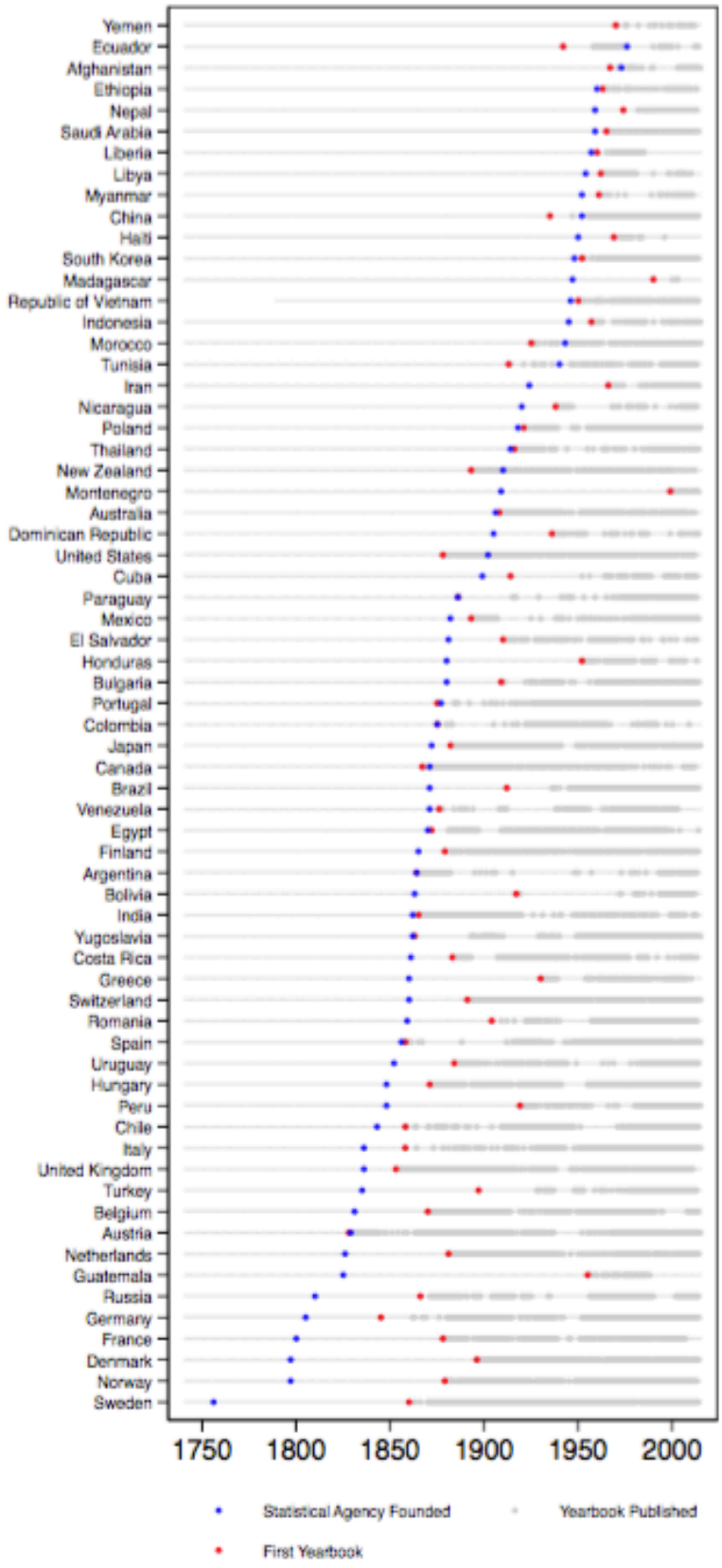
The commission was replaced in 1850 by an independent central Statistical Bureau, further professionalizing its services (Jensen 1918). In Norway, which Denmark ceded to the Swedish kingdom in 1814, an independent Statistical Bureau was set up in 1837 within the finance ministry (Kiær 1918).

In France, the revolution of 1789 had swept away estates and privileges and with it some of the mental constricts of the *ancien régime*. Napoleon's subsequent drive to rationalize government and social processes within the ideal of an authoritarian-centralist but modern and efficient state was a perfect moment for the development of statistical capacity. With the world of subjects and estates disintegrating, the counting and adding up of individuals and their characteristics across administrative regions started to make much more sense (Prévost and Beaud 2015, 68). In this era of rationalization, France established a *Bureau de Statistique* in 1800. But with the start of the Napoleonic wars and the French invasion of Russia in 1812, the appetite for statistical analysis vanished. That year, the French statistical agency was closed down; it would not be re-established until 1833 (Westergaard 1932, 114).

---

[3] These first tabulations were initially kept secret to prevent other countries from gaining knowledge that was potentially dangerous from the perspective of national security, especially about the perceived military disadvantage of Sweden's low population count. The first official report of the tabulation commission was issued in 1761 (Arosenius 1918, 744).

FIGURE 2. Introduction, Agencies and Yearbooks



In Prussia, the king urged the establishment of a statistical office in 1805, the first in the German Empire. After a crushing defeat against Napoleon's forces at the Battle of Jena in 1806, however, the office was shut down and statistical publications banned (Wu ¨rzburger 1918; Westergaard 1932). The central statistical office was re-established in 1810, with the explicit aim to aid the reconstruction of the shattered Prussian state and to provide statistical analyses that could be "put to immediate use by the highest administrators in the land" (Hacking 1990, 30). In Russia, the first statistical office was founded in 1811 as a division within the newly created Ministry of Police, itself a conspicuous choice for the Tsarist autocracy (Kaufmann 1918).

Despite its short-lived existence, the French model of a statistical office had kick-started the desire of governments to expand their ability to collect information about the populations and territories that they governed. Both the territories touched directly by the influence of the French administrative model through military subjugation and coalition forces that would eventually derail Napoleon's plans learned from the French model. Around the 1830s and 1840s—sometimes called the "era of enthusiasm" (Westergaard 1932, 136–171)—many official statistical agencies were re-established or newly founded in several countries. Their work was further buttressed by new statistical societies. Following new statistical offices in the Netherlands (1826), Austria (1829), and Belgium (1831), a major development was the establishment of a statistical department at the Board of Trade in the United Kingdom in 1832. Though substantial statistics had been collected previously within the organization, much of it remained within the office without being utilized for administrative purposes.

Returning to figure 2, we see that the early establishment of independent statistical offices was not limited to Europe. Early followers of the trend to investing in information capacity can also be found in Latin America with Guatemala (1825), Chile (1843), Peru (1848), and Uruguay (1852) establishing their statistical offices.

Official statistical agencies are commonly meant to serve a dual purpose: to provide statistical information for the state administration and to allow the public to access information collected by government agencies. Most countries in the world (and all countries in our sample) have at one time or another achieved part of the latter purpose by publishing an annual statistical yearbook. In general, statistical yearbooks compile available statistics on demographic, social, and economic conditions and activities in a country, often including information on the regional as well as the national level.

The publication of a statistical yearbook requires a high level of sophistication in several government agencies, since more than one agency typically contributes to the collection, processing, and aggregation of statistical information. As discussed above, many countries established a dedicated statistical agency precisely to aid with the task of combining statistical information across government agencies. We argue, therefore, that the ability to produce a statistical yearbook is a further, useful measure of a government's "information capacity".

Beyond the *ability* to produce a statistical yearbook—that is, to collect the necessary information and process it—governments need to have the *willingness* to disseminate such information publicly. Thus, unlike the establishment of the census and a statistical agency, pro- viding statistical accounts to the public is not unequivocally beneficial for rulers: governments may have well have an interest in obtaining and systematizing such statistical information for internal use, but still be hesitant to provide such information transparently to the public— including to domestic and foreign adversaries.

Figure 2 also provides the dates of publication of all yearbooks of the countries in our sample, with the first yearbook publication highlighted in red. The first country to publish an official statistical yearbook was the Austro-Hungarian empire with its 1828 publication *Tafeln zur Statistik der österreichischen Monarchie*, which was followed a year later by the establishment of a central statistical office (*Statistisches Bureau*, later *Statistisches Zentralamt*) as part of the general auditing oce. In most countries, we unsurprisingly observe the establishment of an independent statistical agency before the first publication of a statistical yearbook.

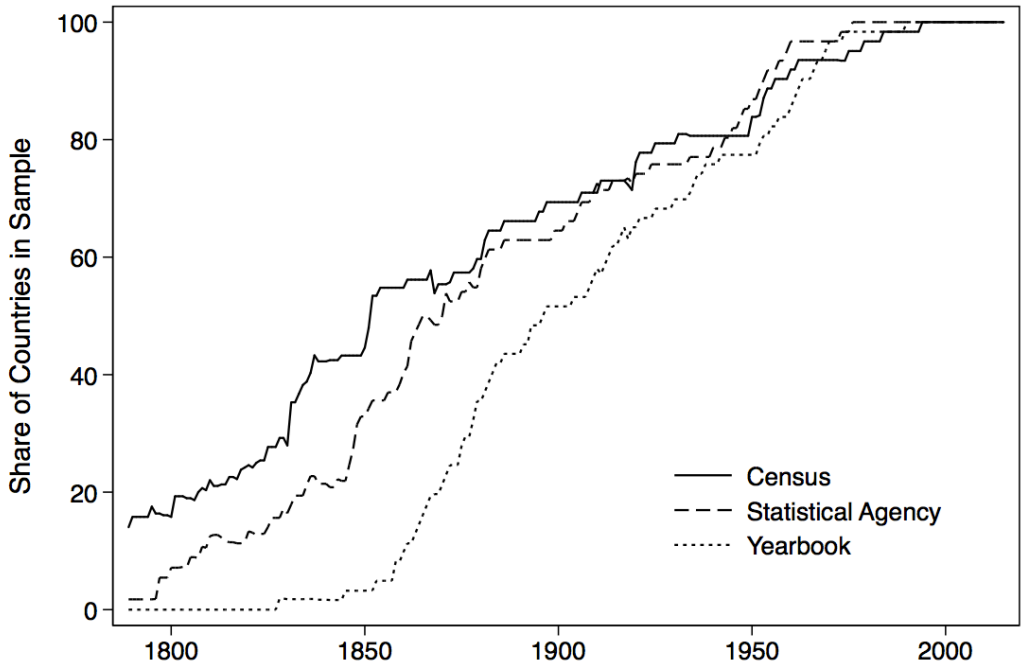
V Information Capacity Since the French Revolution

Introduction of the Census, *Agencies*, *and Yearbooks*. Figure 3 summarizes the information in Figure 1 and Figure 2 in a single graph, describing the share of countries in our sample (in percent) that had introduced the census, that had established a statistical agency, and that released statistical yearbooks. As this figure shows, across countries, these three information-gathering and information-organizing institutions were introduced in that order: the census came first, agencies came later, and yearbooks came last.

There is an interesting over-time pattern, however, that becomes apparent in Figure 4. This figure combines some of the country-specific information in Figure 1 and Figure 2 in a single graph, showing when the first census ever was conducted in each country (the red C's), when each country's national statistical agency was introduced (the blue S's), and when a statistical yearbook was first released in each country (the black Y's). Among early adopters—countries in which the first census was conducted in the eighteenth or early nineteenth century—there was typically a big gap between the first census and the establishment of a statistical agency. Among states that had not introduced any of these things until the middle of the nineteenth century or later, however, those gaps were smaller: they often started to do all of these things at once. It seems likely that the intensification of collaboration among statisticians and administrators in the middle of the nineteenth century contributed to this development (Randeraad 2010): over time, it became the norm in the

international system that modern states should have both censuses and statistical agencies.

FIGURE 3. Order of Introduction, All Countries

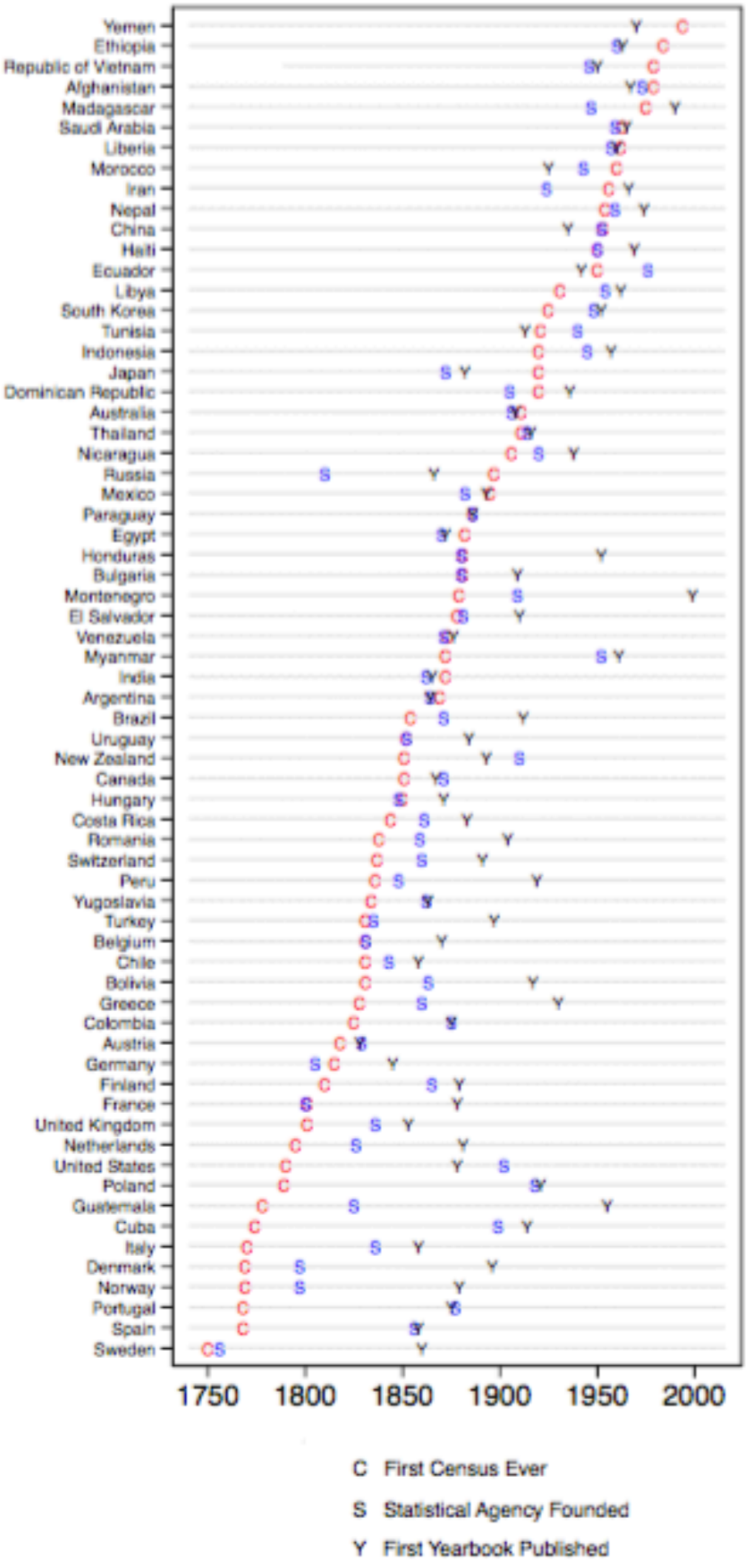


An Index of Information Capacity. While the individual indicators of information capacity we have identified—the introduction of the census, the establishment of a statistical agency, and the publication of a yearbook—are of interest in their own right, we are ultimately interested in estimating the underlying *level of information capacity* in any given country-year. To be able to compare the level of information capacity over time and across countries, we thus need to aggregate these three binary indicators.

Simply summing the binary indicators, the simplest method of aggregation, is unsatisfactory, since it introduces strong assumptions that are unlikely to be met: ad hoc scaling through simple summation assumes that each of the three binary variables is an equally good indicator of a single underlying unidimensional continuum, or, in other words, that they have equal weight when it comes to determining information capacity.

Instead, we rely on a simple unidimensional item-response-theory (IRT) framework that helps to relax that assumption (we regard the assumption of unidimensionality as relatively unproblematic given our theoretical setup and the measures that we have obtained).

FIGURE 4. Order of Introduction, By Country



We want to estimate the latent level of *information capacity* θ_i for all countries *i* in our sample, where *i* = 1, . . . , n. The latent level of θ_i is unobserved, but we observe the responses for individual items *j*, where *j* = 1, . . . , *J*. Given that our three indicators are binary responses, we start with the simplest one-parameter logistic IRT model (1pl), which is known in the literature as the Rasch model (Rasch 1960). In the Rasch model, the probability of a positive response for item *j* given a level of information capacity θ_i is defined as

$$p_j(\theta_i) = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}}, \text{ i=1,} \ldots, n, \text{ j=1,} \ldots, J,$$

where β_j is a measure of *difficulty* for item *j*. Importantly, both θ_i and β_j are measured on the same scale, allowing a direct comparison between information capacity and the difficulty of the items.

All IRT models discussed in this paper make the following three assumptions (see Bartolucci, Bacci, and Gnaldi 2015, 66-67 for details): (1) Unidimensionality. The latent trait *information capacity* θ_i is assumed to be unidimensional. In other words, we assume that for each country *i*, the responses to the *J* items depend on the same latent trait level θ_i, which is unidimensional and belongs to R. (2) Local independence. For each country *i*, the responses to the *J* items are independent given the latent level of information capacity θ_i. (3) Monotonicity. The conditional probability of responding correctly to item *j*, denoted by $p_j(\theta_i) = p(Y_{ij} = 1 | \theta_I)$, and known as ICC or item characteristic curve, is a monotonic non-decreasing function of θ_i. Neither of these assumptions is particularly problematic in our context. However, the one-parameter IRT model makes one additional assumption: since the Rasch model only distinguishes items by their difficulty, β_j, this simple model further assumes that all items are equally discriminating. This is an assumption that we will discuss and relax in subsequent analyses.

TABLE 1. IRT Models of Information Capacity (1750-1970)

	Model 1	Model 2
Difficulty (β)		
Census	0.42***	0.47***
	(0.10)	(0.09)
Statistical Agency	0.37***	0.28**
	(0.07)	(0.11)
Yearbook Coverage	0.86***	0.87***
	(0.07)	(0.07)
Discrimination (λ)		
Common Coefficient	3.42***	
	(0.19)	
Census		2.57***
		(0.27)
Statistical Agency		5.43***
		(1.63)
Yearbook Coverage		3.83***
		(0.39)

*$p < 0.10$, **$p < 0.05$, *** $p < 0.01$ (two-tailed). Robust standard errors in parentheses (clustered by country).

The results for the one-parameter logistic model are presented in Table 1. Model 1 presents the results for the Rasch model for the time period from 1750 to 1970. The common discrimination coefficients λ of the models are presented in the bottom panel of the table.[4]

We find that all three difficulty parameters are significantly different from zero, and thus contribute to distinguishing countries with respect to their underlying information capacity. The fact that all the thresh- old parameter estimates are larger than zero also suggests that all three items are relatively difficult, which means that they are most useful for discriminating countries who have somewhat higher information capacity abilities. Establishing a statistical agency and running a census are, perhaps not surprisingly, found to be "easier" than the publication of a yearbook.

While appealing for its simplicity, the Rasch model, as we discussed earlier, makes the strong assumption that each of the items can dis- criminate between high- and low-capacity states across the same range of information capacity. Model 2 in Table 1 relaxes that assumption, by allowing each item to have its own λ_i rather than constraining them to a single, common discrimination coefficient. The two-parameter model fits the data significantly better.

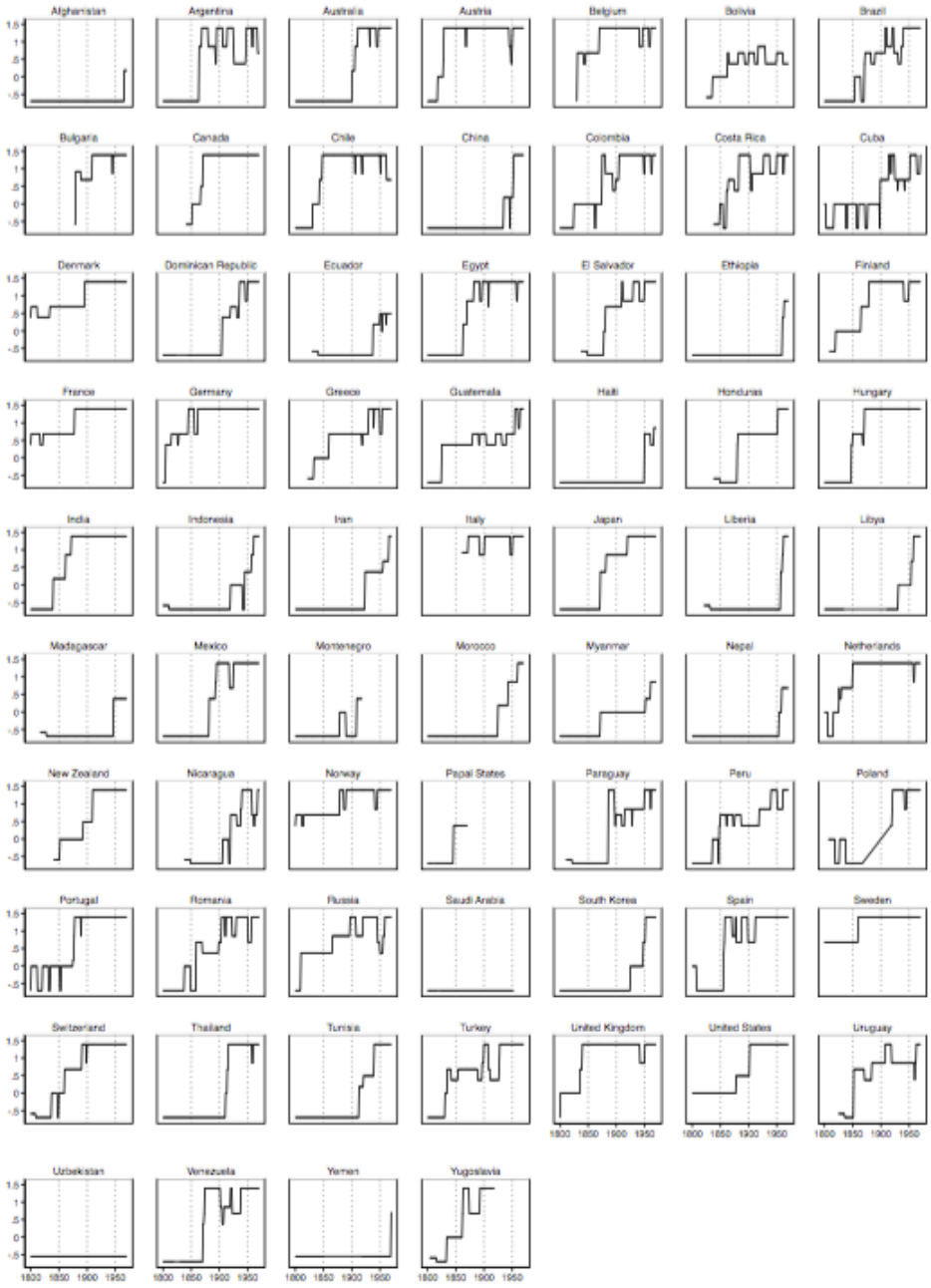
Figure 5 uses the estimates of Model 2 in Table 1 to generate predictions for the estimated level of information capacity. In accordance with our historical descriptive sections above, we find that countries such as Sweden, Austria, France, Germany, and the United Kingdom developed high information capacity relatively early in the nineteenth century. Interestingly, we find a drop in the estimated information capacity around the time of the Second World War in many countries in our sample, since censuses were often postponed and yearbook publication sometimes ceased.

Correlates of Information Capacity. Figure 6 describes the relation- ship between (the natural log of) GDP per capita (using data from Maddison (2011)), political regimes (using data from Boix, Miller, and Rosato (2012), where black dots are democracies and hollow dots are non-democracies), and our measure of information capacity in four moments in time: 1800 (the first year for which data on regimes are avail- able), 1850, 1900, and 1950. The grey lines are LOWESS smoothing lines that describe the estimated bivariate relationship between GDP per capita and information capacity.

---

[4] Since very few countries in our sample extended their information capacity on the three selected measures after 1970, the ability of the model to distinguish levels of information capacity of individual countries (and thus the model fit) decreases for the time period after 1970. Accordingly, in the post-1970 period all three difficulty coefficients are insignificant. Since the model is identified by setting the average level of information capacity equal to zero, the results imply that the items are no longer able to differentiate low- and high-capacity states when the most recent data is included.

FIGURE 5. Estimated Information Capacity By Country

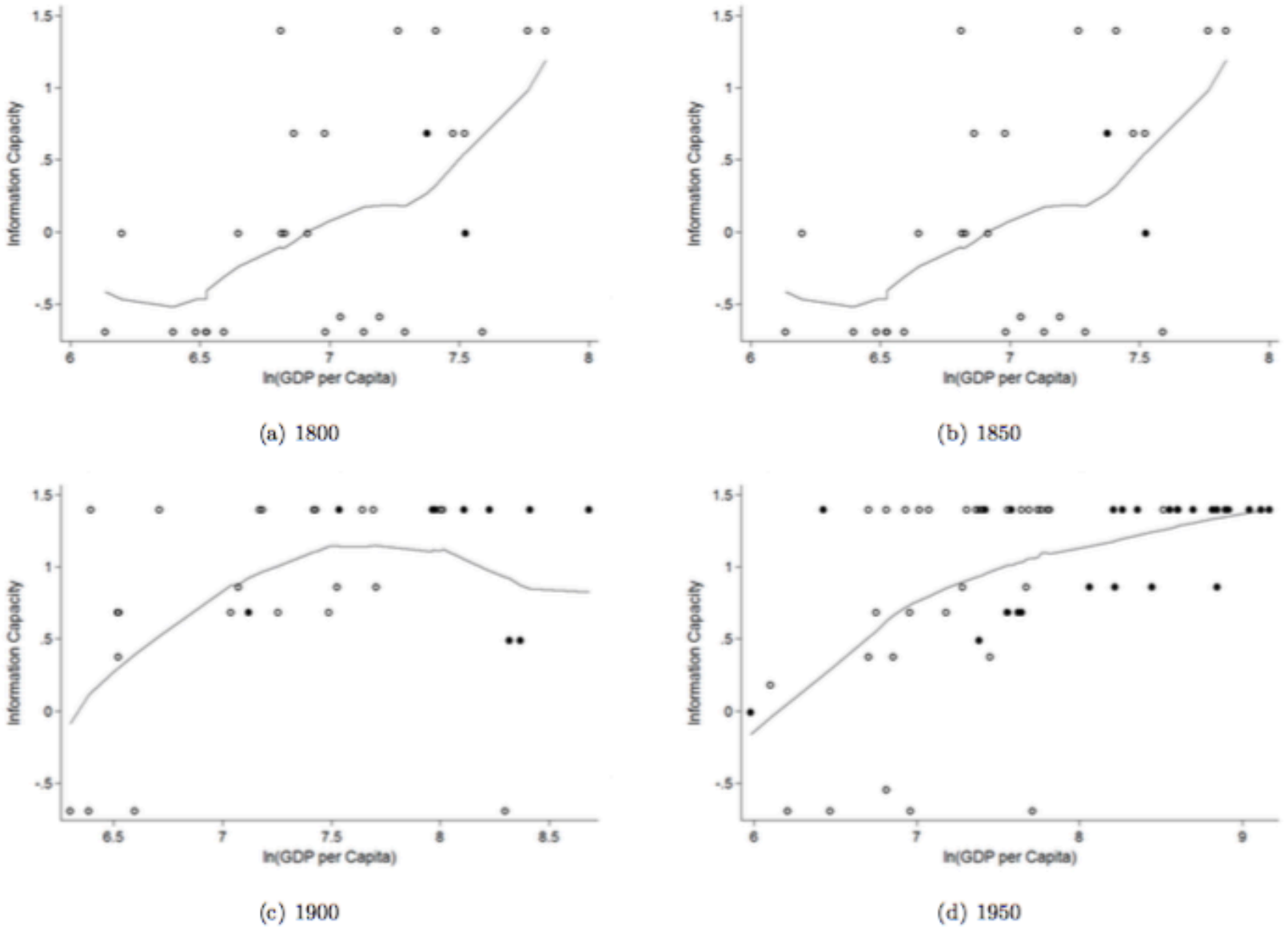


Based on a 2 parameter IRT model, including information on the census, the statistical agency, and the yearbook.

Clearly—and as one would expect—relatively richer countries have been more likely than relatively poorer countries to have high information capacity in both the nineteenth and the twentieth centuries.

The relationship between political regimes and information capacity is much less clear-cut. By the end of the period that we examine, in the middle of the twentieth century, most of the democracies in our sample had relatively high information capacity (and were typically rich, which means that they cluster in the top right-hand corner of Figure 6d. In the nineteenth century, however, all the states that had the maximum level of information capacity were *not* democracies. We expect to be exploring these interesting patterns in future work.

FIGURE 6. Prosperity, Regimes, and Information Capacity



Data sources: Boix, Miller, and Rosato (2012) (democracy), Maddison (2011) (GDP per capita), own data collection and estimation (information capacity). Hollow dots are non-democracies. Solid dots are democracies. GDP per capita (the x-axis) is logged. The grey lines are LOWESS smoothers.

VI Conclusion

In this paper, we have argued that all the resources that states use when they implement their policies—that is, when they "make and enforce" rules, "carry out" policies, and "control" persons, activities, and resources—information is the most important. We have there- fore developed a new, quantitative measure of the ability of states to collect, retrieve, and process information, relying on systematic comparative and historical data on three information-gathering activities: the introduction of a national census, the establishment of a permanent statistical agency, and the publication of a statistical yearbook. We have also analyzed the empirical relationships between these three different components and presented evidence regarding the economic and political correlates of information capacity. The next step will be to add indicators of other widely used information-gathering activities, to broaden the basis for our estimation of information capacity. [5]

[5] See the Appendix.

References

Alterman, Hyman. 1969. *Counting people : the census in history.* 1969.

Anderson, Margo J. 1988. *The American census: a social history.* New Haven: Yale University Press.

Arosenius, Edvard. 1918. "The History and Organization of Swedish Official Statistics." In The *History of Statistics, Their Development and Progress in Many Countries*, edited by John Koren. New York: MacMillan, 537–569.

Bartolucci, Francesco, Silvia Bacci, and Michela Gnaldi. 2015. *Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata.* Boca Raton: Chapman and Hall/CRC.

Besley, Tim and Torsten Persson. 2011. *Pillars of Prosperity*. Princeton: Princeton University Press.

Boix, Carles, Michael Miller, and Sebastian Rosato. 2012. "A Complete Dataset of Political Regimes, 1800–2007." *Comparative Political Studies*.

Centeno, Miguel Angel and Agust´ın Ferraro, editors. 2013. *State and nation making in Latin America and Spain: republics of the possible*. New York: Cambridge University Press.

Centralbyrån, Statistiska. 1969. *Historisk statistik för Sverige Del 1. Befolkning 1720-1967.* Stockholm: AB Allmänna förlaget, 2nd ed.

Curtis, Bruce. 2002. *The Politics of Population: State Formation, Statistics, and the Census of Canada, 1840-1875*. Toronto, Canada: University of Toronto Press.

D'Arcy, Michelle and Marina Nistotskaya. 2015. "State First, Then Democracy: Using Cadastral Records to Explain Governmental Performance." QoG Working Paper Series 2015:11 (Gothenburg: the Quality of Government Institute).

Evans, Peter and James E. Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of "Weberian" State Structures on Economic Growth." *American Sociological Review* 64 (5):748–765.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97 (1):75–90.

Fishbein, Meyer H. 1973. *The censuses of manufactures, 1810-1890.* Washington: National Archives and Records Service, General Services Administration.

Foucault, Michel. 1978. *The History of Sexuality: An introduction*. Vintage.

Fukuyama, Francis. 2013. "What Is Governance?" *Governance* 26 (3):347–368.

Goyer, Doreen S. and Eliane Domschke. 1983. *The handbook of national population censuses: Latin America and the Caribbean, North America, and Oceania.* Westport, Conn: Greenwood Press.

Goyer, Doreen S. and Gera E. Draaijer. 1992. *The handbook of national population censuses. Europe.* New York: Greenwood Press.

Hacking, Ian. 1990. *The Taming of Chance*. Cambridge University Press.

Hansen, Randall and Desmond King. 2013. *Sterilized by the State: Eugenics, Race, and the Population Scare in Twentieth-Century North America*. Cambridge University Press.

Hanson, Jonathan K. 2015. "Democracy and State Capacity: Complements or Substitutes?" *Studies in Comparative International Development* 50 (3):304–330.

Hanson, Jonathan K. and Rachel Sigman. 2013. "Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Re- search." SSRN Scholarly Paper ID 1899933, Social Science Research Network, Rochester, NY.

Herbst, Jeffrey. 2014. *States and Power in Africa: Comparative Lessons in Authority and Control.* Princeton University Press.

Jensen, Adolf. 1918. "The History and Development of Statistics in Denmark." In *The History of Statistics, Their Development and Progress in Many Countries*, edited by John Koren. New York: MacMillan, 201–214.

Kaufmann, A. 1918. "The History and Development of the Official Russian Statistics." In *The History of Statistics, Their Development and Progress in Many Countries*, edited by John Koren. New York: MacMillan, 469–534.

Kiær, Anders Nicolai. 1918. "The History and Development of Statistics in Norway." In *The History of Statistics, Their Development and Progress in Many Countries*, edited by John Koren. New York: MacMillan, 447–65.

Lee, Melissa and Nan Zhang. 2016. "Bringing Legibility Back In: The Informational Foundations of State Capacity." Unpublished manuscript.

Levitan, Kathrin. 2011. *A Cultural History of the British Census.* Basingstoke: Palgrave Macmillan.

Lindvall, Johannes and Jan Teorell. 2016. "State Capacity as Power: A Conceptual Framework." Unpublished manuscript, Lund University.

Maddison, Angus. 2011. "Statistics on World Population, GDP and Per Capita GDP, 1–2008 AD." University of Groningen, available on http://www.ggdc.nl/maddison, accessed on September 13, 2011.

Mann, Michael. 1984. "The Autonomous Power of the State: Its Origins, Mechanisms and Results." *European Journal of Sociology / Archives Européennes de Sociologie* 25 (02):185–213.

Mann, Michael. 2012. *The Sources of Social Power: Volume 2, The Rise of Classes and Nation-States, 1760-1914*. Cambridge University Press.

McAdam, Doug, Sidney Tarrow, and Charles Tilly. 2001. *Dynamics of Contention*. Cambridge: Cambridge University Press.

Prévost, Jean-Guy and Jean-Pierre Beaud. 2015. *Statistics, Public Debate and the State, 1800–1945: A Social, Political and Intellectual History of Numbers.* Routledge.

Randeraad, Nico. 2010. *States and statistics in the nineteenth century: Europe by numbers.* Manchester: Manchester University Press.

Rasch, Georg. 1960. “Probabilistic models for some intelligence and achievement tests.” Copenhagen: Danish Institute for Educational Research.

Scott, James. 1998. *Seeing Like A State*. New Haven: Yale University Press.

Skocpol, Theda. 1985. “Bringing the State Back In: Strategies of Analysis in Current Research.” In *Bringing the State Back In*, edited by Theda Skocpol, Peter B. Evans, and Dietrich Rueschemeyer. Cam- bridge: Cambridge University Press, 3–38.

Soifer, Hillel. 2008. “State Infrastructural Power: Approaches to Conceptualization and Measurement.” *Studies in Comparative International Development* 43 (3-4):231–251.

Soifer, Hillel. 2012. “Measuring State Capacity in Contemporary Latin America.” *Revista de Ciencia Política* 32 (3):585–98.

Soifer, Hillel. 2013. “State Power and the Economic Origins of Democracy.” *Studies in Comparative International Development* 48 (1):1–22.

Soifer, Hillel and Mathias vom Hau. 2008. “Unpacking the Strength of the State.” *Studies in Comparative International Development* 43 (3):219–230.

Ventresca, Marc Joseph. 1995. *When states count: Institutional and political dynamics in modern census establishment, 1800-1993*. Ph.D., Stanford University, California, United States of America.

Westergaard, Harald. 1932. *Contributions to the History of Statistics.* P.S. King.

Westfall, Gloria. 1986. *Bibliography of Official Statistical Yearbooks and Bulletins*. Alexandria, VA: Chadwyck-Healy.

Whitehead, Laurence. 1995. “State organization in Latin America since 1930.” In *The Cambridge History of Latin America*, edited by Leslie Bethell. Cambridge: Cambridge University Press, 1–96.

Würzburger, Eugene. 1918. “The History and Development of Official Statistics in the German Empire.” In *The History of Statistics, Their Development and Progress in Many Countries*, edited by John Koren. New York: MacMillan, 333–62.

Appendix A. Complexity of Information

In addition to our general strategy for measuring information capacity through the indicators of censuses, statistical agencies and year- books, we have also experimented with a measure of *yearbook complexity* for a smaller pilot sample of countries. This is work in progress, so any comments or suggestions would be most helpful.

The pilot sample was primarily selected on the basis of having access to digitized yearbooks in a language that was intelligible to our research assistant, which in this case meant English, German, Dutch, Danish, Norwegian, Swedish, Portuguese, Spanish or French (in Russia and Finland requiring a French translation). Accordingly, we collected information from a total of 113 yearbooks in the following 16 countries and time periods: Austria (8 in 1829–1900), Belgium (7 in 1870–1930), Canada (8 in 1867–1930), Chile (7 in 1863–1930), Denmark (5 in 1896–1930), Finland (6 in 1879–1930), France (7 in 1878–1930), Germany (9 in 1845–1930), Italy (8 in 1858–1930), Netherlands (9 in 1851–1930), Norway (6 in 1879–1930), Portugal (7 in 1875–1930), Russia (3 in 1904– 1916), Sweden (7 in 1870–1930), the United Kingdom (9 in 1853–1929) and the United States (7 in 1878–1930).

According to Westfall (1986), the themes typically covered by statistical yearbooks can be divided into five different sections: A. Physical environment, B. Demography, C. Economics, D. Politics, and E. Society and culture. Based on this categorization, we decided to code the presence (1) or not (0) of information within each yearbook on at least two indicators within each of these themes—one that we deemed more difficult for the statistical agency to collect information on, and one that we deemed would be easier to collect information on. Since information on themes B, C, and, in particular, D are particularly prevalent, we decided to over-sample these categories, ending up with 18 selected indicators. The idea behind this measure is that the higher the information capacity of the state in question, the larger the share of these indicators that are included in a yearbook should be, taking the estimated relative "difficulty" of each indicator into account.

The result of a Rasch model (see section 4 above) applied to these 18 indicators is presented in Figure A.1. Since we are here assuming that each indicator has the same discrimination parameter, the curves only differ in terms of their θ_i score, indicating the degree of difficulty of each information task. What should be obvious from this graph is that most indicators cluster toward the left-hand, or, in other words, the "easier," side of scale. The two major exceptions are the political indicators of total number of personnel in central government administration and the number of government personnel subdivided my ministry (a third partial exception is the number of prisons or penitentiaries in the country). Overall, however, most yearbooks in most countries appear to be covering most of these indicators. This could be the result of our se- lection of (a) high information-capacity countries (by

and large), (b) indicators of too simple data-collection tasks for statistical agencies, or both.

We have nevertheless estimated the latent levels of information complexity based on these 18 indicators. Results by country, ordered from the highest to the lowest average level of yearbook complexity, are presented in Figure A.1. These estimates are both promising and discouraging. On the one hand, several within-country trajectories make substantial sense, such as the progressive increase in information complexity in Italy after unification in 1860, the slumps in Chile during the War of the Pacific in 1879–1883, in France, Russia and Belgium after or during the ravages of the First World War 1914–1918, the slump in Canada possibly caused by territorial expansion after independence in 1867, and perhaps also the secular rising trend after the civil war in the United States. The overall lower averages in countries such as Portugal, Chile and Russia as compared to countries such as Denmark and Sweden also conforms to conventional wisdom.

Other patterns make less sense. Most worrisome are the overall very low scores for Britain, which seem to be a reflection of whims or parochial preferences for what should be included in a yearbook rather than of a lack of information capacity. The low average scores for Germany also seem out of line (although they may in part be explained by potentially missing parts of their digitized yearbooks and the use of fonts that were hard to interpret). There also seems to be no natural explanation for the early rise but subsequent decline in the information capacity of Austria and the Netherlands.

FIGURE A.1. Item Response Model of Information Complexity

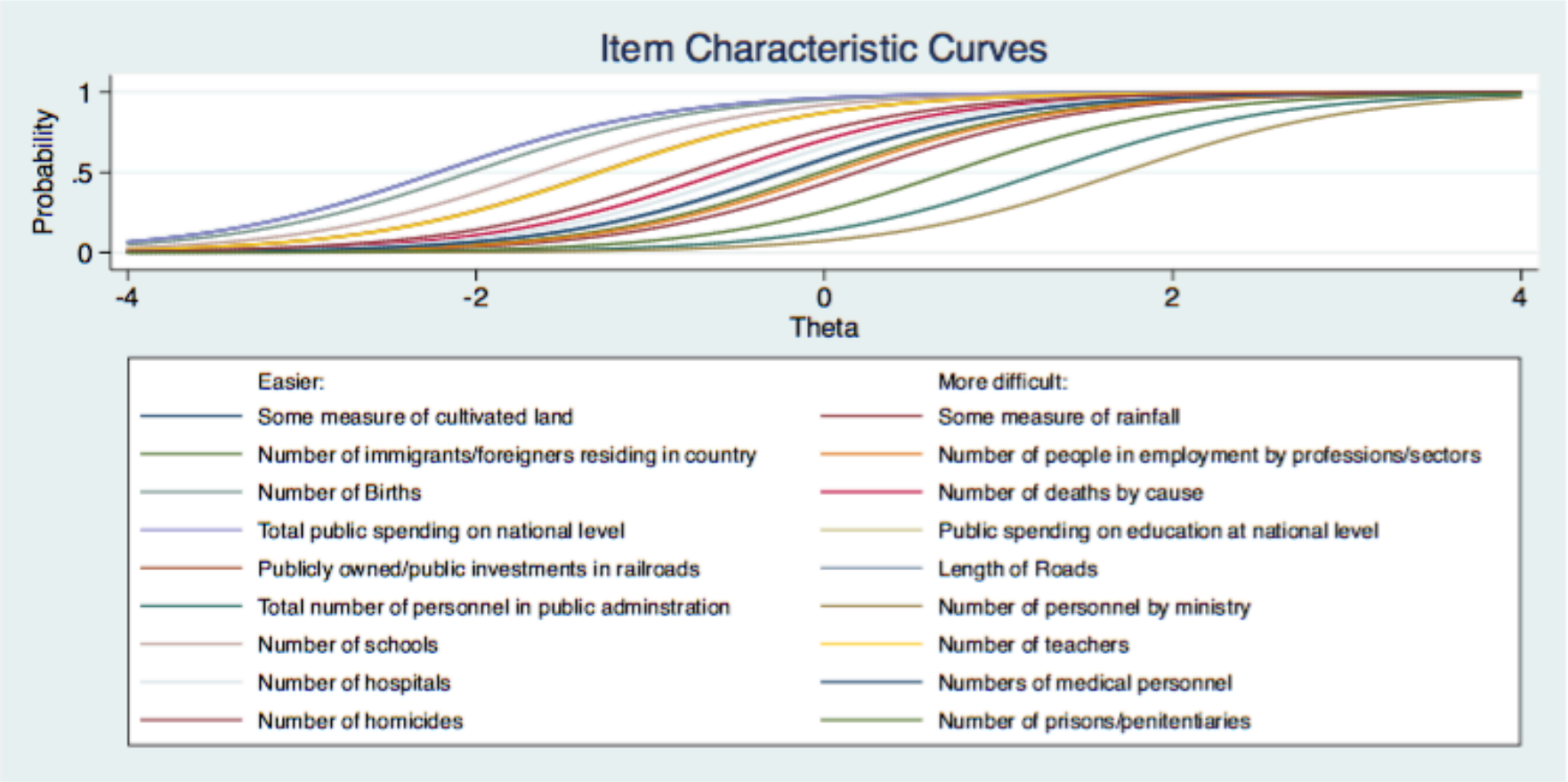
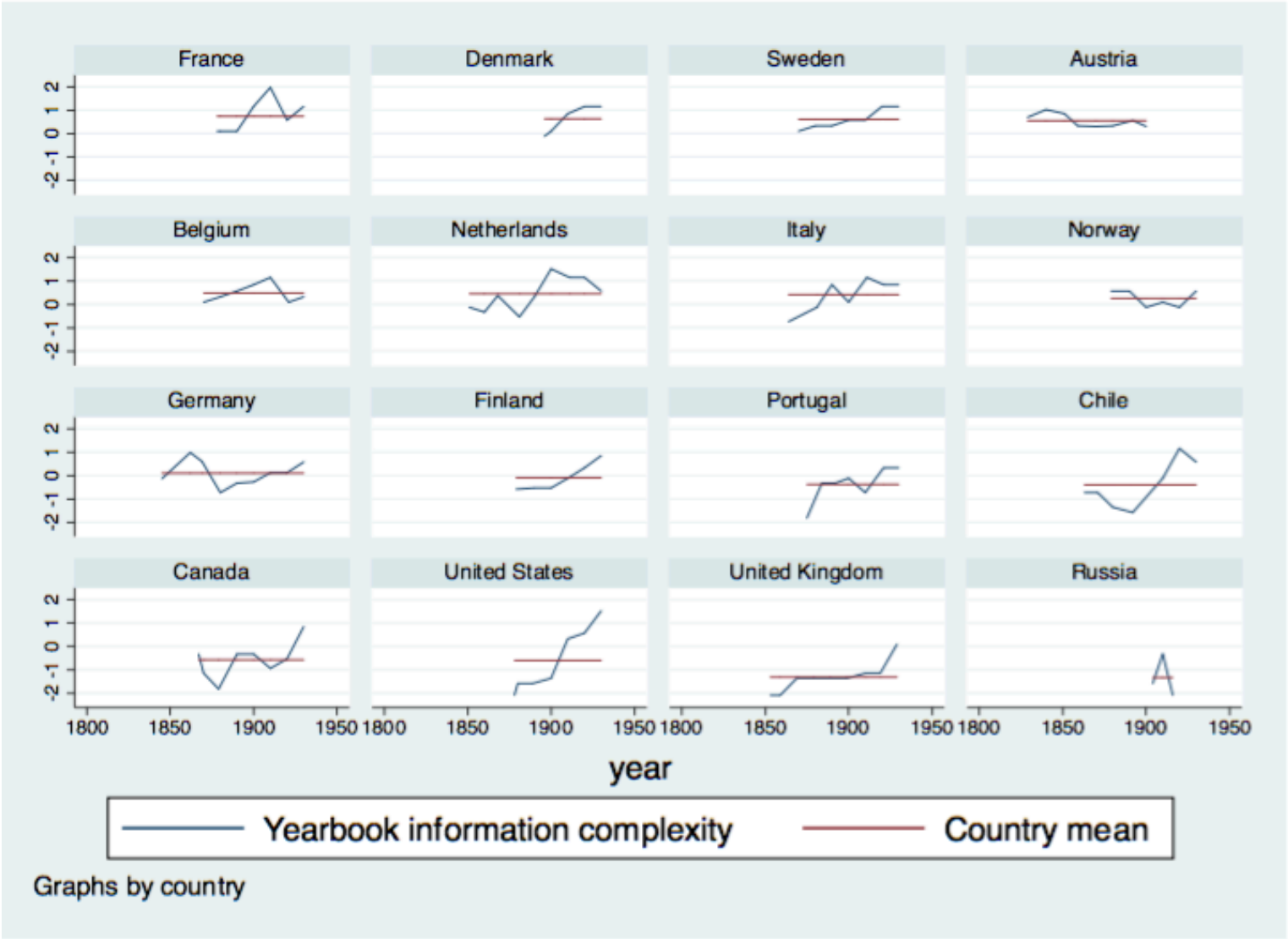



FIGURE A.2. Estimates of Information Complexity