

WIDER Working Paper 2021/177

No taxation without informational foundation

On the role of legibility in tax state development

Matthias vom Hau,¹ José Peres-Cajías,² and Hillel David Soifer³

December 2021

Abstract: This paper combines cross-national statistical analysis and in-depth historical case studies of Argentina and Chile to explore the relationship between two crucial dimensions of state capacity. We show that information capacity contributes to the development of fiscal capacity. States require accurate information about their subject populations, territories, and economies in order to effectively mobilize revenues. In developing this argument this paper makes three broader contributions. First, while existing scholarship either treats distinct dimensions of capacity as separate entities, or simply assumes that they complement each other, our findings urge scholars to treat state development as sequential and to further investigate how multiple dimensions of state capacity are interrelated. Second, the paper suggests a broader underlying set of mechanisms—economies of scope—which connect these dimensions, and explores them in the specific context of the dependency of fiscal capacity on information capacity. Third, we join scholarship on the importance of societal compliance in the creation of the fiscal state, but with a focus on elite compliance with the state’s information collection efforts, which was crucial to tax state development.

Key words: state capacity, information capacity, fiscal capacity, taxation, state formation, mixed methods, Latin America

JEL classification: C23, H20, H83, N46

Acknowledgements: We are grateful to James Robinson, Antonio Savoia, Kunal Sen, and the participants of the 2021 UNU-WIDER Fiscal States Workshop for their constructive feedback. Guillem Amatller provided excellent research assistance. Matthias vom Hau’s contribution to this paper received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (agreement 864333—ETHNICGOODS).

¹ Institut Barcelona d’Estudis Internacionals (IBEI), Barcelona, Spain, corresponding author: mvomhau@ibei.org; ² Universitat de Barcelona (UB), Barcelona, Spain; ³ Temple University Philadelphia, United States

This study is published within the UNU-WIDER project [Fiscal states—the origins and developmental implications](#), which is part of the [Domestic Revenue Mobilization](#) programme. The programme is financed through specific contributions by the Norwegian Agency for Development Cooperation (Norad).

Copyright © UNU-WIDER 2021

UNU-WIDER employs a fair use policy for reasonable reproduction of UNU-WIDER copyrighted content—such as the reproduction of a table or a figure, and/or text not exceeding 400 words—with due acknowledgement of the original source, without requiring explicit permission from the copyright holder.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-117-4

<https://doi.org/10.35188/UNU-WIDER/2021/117-4>

Typescript prepared by Joseph Laredo.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

1 Introduction

States shape economic development. There is a broad agreement in the literature that state institutions and their actions have major implications for economic growth and transformation (Leftwich 2008; Rueschemeyer 2005). But not all states are equally equipped to foster economic development because of differences in *state strength*.¹ Indeed, scholars have shown that the capacity of states to secure property rights, steer economic activities, transform class relations, and engage in redistribution often varies dramatically both across and within countries, and over time (Evans 1995; Sandbrook et al. 2007).² It is therefore not surprising that state capacity has become a widely accepted analytical tool for studying the role of states in economic development.

Its background conceptualization is largely uncontroversial. Scholars of various theoretical and disciplinary backgrounds agree that state capacity refers to the capability of states to effectively implement their initiatives throughout the territory they claim to govern (Saylor 2013; Soifer and vom Hau 2008). There is also a broad-based consensus that state capacity should be understood as something that states have, independent of the specific goals it is used for (Centeno et al. 2017).

But how to approach state capacity as a systematized concept (Adcock and Collier 2001) that can be clearly operationalized remains a matter of intense debate (Hendrix 2010; Savoia and Sen 2015; vom Hau 2012). While scholars agree that state capacity is a multi-dimensional concept, they focus on different sets of dimensions (e.g. Hanson and Sigman 2021; Soifer 2008) and there is no consensus on which aspect of state capacity should be considered most fundamental and therefore privileged when developing a plausible measurement strategy.

A significant body of work focuses on *fiscal capacity*, the state's ability to mobilize revenues, as a shorthand for overall state strength (e.g. Bräutigam et al. 2008; Slater 2010; Tilly 1990). The basic logic of this approach is that states require resources to exercise control and pursue their projects. Without revenues, it is difficult to build and maintain effective bureaucracies that can plan and execute policy objectives. Thus, many studies assess variations in state capacity by focusing on (income) tax ratios or comparable indicators of state revenues during historical periods (e.g. Queralt 2019; Thies 2005).

Yet the focus on fiscal capacity as a shorthand for state strength has been challenged. A growing body of work suggests that another aspect of state capacity—*information capacity*—should be given conceptual priority. This aspect of the state is defined as the basic knowledge states hold about the make-up of their populations and territories. Making societies 'legible' (Scott 1998) has long been a focus of state efforts. States frequently count, measure, and otherwise record detailed information about the make-up of their populations, territories, and economies, whether through population censuses or cadastral records, and rely on a diverse set of tools, including specialized statistical agencies and statistical yearbooks, to do so. Some scholars argue that this information capacity lies at the core of the state's overall strength (Lee 2020; Scott 1998).

In this paper, we build on, yet also move beyond, both of these bodies of work. We begin with the presumption that both fiscal capacity and information capacity are central aspects of the state and important determinants of development outcomes. But whereas existing scholarship has treated these two dimensions separately, we explore the relationship between them. Our paper

¹ We use the terms 'state strength' and 'state capacity' interchangeably.

² High-capacity states might of course use their capacity to hinder economic transformation and generate dramatic failures.

tests the hypothesis that information capacity is a necessary (but not sufficient) determinant of fiscal capacity—that the modern tax state depends in part on the state’s prior development of information capacity. We do so through a mixed-methods approach that combines cross-national statistical analysis with in-depth historical case studies of Chile and Argentina. This allow us to move beyond the better-known experience of developed countries and investigate the historical development of information and tax capacity in less well-studied middle-income states. Both components of our empirical analysis find strong support for the hypothesis, and we therefore conclude that information capacity should be seen as one of the causes of the development of the fiscal state.

In the statistical analysis, which is based on the most cutting-edge cross-national dataset on information capacity currently available (Lee 202; Lee and Zhang 2017), the paper shows a significant association between legibility and various measures of taxation for 1980 to 2010, even when controlling for other theoretically relevant factors. During this period, information capacity appears to underlie extractive capacity. Information capacity and taxation have also been relatively stable over time, with high-performing states maintaining their respective position, and low-performing states only rarely catching up.

Because tax state development is so path dependent, we turn to its formative moments to explore the contributions of information capacity. Our case studies of this crucial phase of state building in 19th-century Argentina and Chile show that taxation developed in the wake of successful initiatives to centralize and systematize information collection, most importantly through the establishment of a statistics office, a regular population census, and a land cadastre. Our case studies demonstrate that the construction of information capacity, itself driven in part by elite compliance, provided human capital and legibility, which made tax state development possible.

This paper thus makes three contributions to scholarship on the fiscal state and state capacity more broadly. First, while existing scholarship either treats distinct dimensions of capacity as separate entities, or simply assumes that they complement each other, our findings urge scholars to treat state development as sequential and further investigate how multiple dimensions of state capacity are interrelated. Second, the paper suggests a broader underlying set of mechanisms—economies of scope—which connect these dimensions, and explores them in the specific context of the dependency of fiscal capacity on information capacity. Third, we join scholarship on the importance of societal compliance in the creation of the fiscal state, but with a focus on elite compliance with the state’s information collection efforts, which was crucial to tax state development.

The next sections develop these theoretical arguments in greater detail, by theorizing the links between information capacity and extraction and then evaluating them in the light of available evidence. Our empirical analysis first presents the cross-national statistical analysis, moving from measurement issues to our research strategy to the main results. The following sections present the two case studies, while the conclusion identifies some open threads and suggests directions for future research.

2 How information capacity underpins fiscal capacity

At the most basic level, fiscal capacity captures the ability of states to mobilize revenues. Those resources might be gathered through different methods of extraction, ranging from easier-to-collect taxes on trade and natural resource wealth to highly complex income taxes. Information capacity refers to the ability of states to gather and analyse reliable information about their subject

populations and territories (Brambor et al. 2020). Commonly-used means of information collection include population censuses and cadastral maps, while specialized statistical agencies are critical for information analysis.

In the existing scholarship, these two dimensions of state capacity tend to be treated as radically distinct from each other. In fact, each of them is often thought of as the very foundation of other state functions, and thus as indicative of overall levels of state strength. A significant body of work going back to Schumpeter (1954) treats extraction as the basis for all other state activities. According to this line of thought, states can do little without having the necessary resources (Tilly 1990). Conversely, another literature puts information capacity at the core of the state's capacity. In this perspective, all the other 'classic' state functions of extraction, coercion, and public goods provision require reliable information in order to be pursued effectively (Lee 2020; Scott 1998).

Our paper explores how these two dimensions of the state might be interrelated and develops a more specific argument about the relationship between fiscal capacity and information capacity. Most existing scholarship on fiscal capacity focuses on the *motives* that underpin efforts to build the tax state. These motives range from war and external threats (Downing 1992; Ertman 1997) to self-enrichment (Olson 1993; Tilly 1985). But this scholarship assumes that motivation is sufficient for tax state development to succeed. It fails to problematize the state's ability to acquire information on the whereabouts and assets of its subjects that will enable effective taxation.

By contrast, the recent 'informational turn' in the study of state building explores precisely this issue. This scholarship (Brambor et al. 2020; D'Arcy and Nistotskaya 2017, 2018; Lee and Zhang 2017) identifies the state's collection and analysis of information about its citizens as a fundamental component of state capacity. Its starting point is the observation: 'States that cannot gather accurate information about their populations are likely doing little else effectively' (Lee 2020: 22).

Based on this line of argument, we hypothesize that information capacity is necessary for fiscal capacity. Some basic knowledge about the demographic characteristics, wealth, and whereabouts of individuals and their households and businesses is required in order to tax effectively. Even the calibration of tariffs and customs collected at ports benefits from the availability of systematic information about the prevailing economic activities in society. Knowledge of the whereabouts and assets of their subjects enables rulers to tax more effectively and to move from taxing production (e.g. rents derived from profitable resources) to more sophisticated and potentially more profitable forms of extraction that revolve around taxing people and their consumption and wealth. States with information capacity are expected to be capable of handling even sophisticated taxation systems such as income or consumption taxes, whereas states without the necessary informational foundations struggle even with comparatively less complex taxes such as customs. The remainder of the paper explores the insights and limitations of this hypothesis empirically. In so doing, we contribute to the analysis of the temporal order in which state development unfolds, and thus improve on existing scholarship that either treats distinct types of capacity as separate entities, or simply assumes that they 'travel together' and complement each other, without accounting for the specific links and dependencies among them. To the best of our knowledge, this paper is the first study of how fiscal capacity depends on information capacity.

Another issue that has been little explored in existing scholarship is the set of mechanisms by which different dimensions of state capacity relate to and depend on each other. The concept of economies of scope—the effect of production of a single output on the marginal cost of producing additional outputs—provides some useful intuition to underpin these dependencies. The kinds of information the state collects about its population and territory, and the human and technological infrastructure built to systematize this information, can be put to multiple uses by the state. Thus, once the state collects demographic information from individuals via the census, it can more

efficiently design public health policies, conscript, and collect direct taxes. Once it collects information about property ownership, it can more effectively design trade taxes. And once individuals are trained and deployed across the territory to collect and systematize these kinds of information, they can also more easily collect and systematize other kinds of information that underpin a range of other state functions. The development of information capacity therefore has spillover effects or complementarities with other state functions; these cost savings and gains in efficiency, or economies of scope, represent the mechanisms by which tax state development is dependent on information capacity. In our case studies, we show these links and dependencies empirically in order to demonstrate how information capacity contributes to tax state development. Our paper, then, aligns itself with a growing body of work on the intra-state determinants of state development (Soifer 2015; vom Hau 2008). While much of the existing scholarship focuses on external causes of fiscal capacity—most prominently geography, geopolitical conflict, and the power configurations between different societal actors—it is important also to explore the evolution of human capital and technologies of data collection internal to the state that facilitate state development.

Our focus on the interdependencies between different dimensions of state capacity also provides a fresh perspective on the origins of the fiscal state, and state development more broadly. Whereas it is intuitive to see extraction as something imposed on society, Levi (1988) convincingly shows that societal compliance facilitates tax state construction. Our case studies reinforce this line of argument, but with an important twist. We find that quasi-voluntary compliance with the state's information-gathering efforts facilitated tax state development. In both Chile and Argentina, we find evidence that the state's endeavours to collect demographic and economic information—the kinds of knowledge crucial to the creation of fiscal capacity—have been facilitated by support from economic elites, who saw the state's information-gathering interventions as useful for their own interests in multiple ways. For example, regular population censuses, and knowledge about the whereabouts and economic assets of the population generated by them, enabled economic elites to more accurately plan their own economic activities. For the same reasons, elites would cooperate with the state's efforts to gather and systematize information about the national economy (e.g. on prices and production levels) and conduct land cadastres.

Information capacity (like many other aspects of the state) should thus be seen as something that emerges from a combination of state initiative and societal collaboration, and the interests of both sets of actors play roles in its development. We therefore echo Levi (1988) in arguing that the creation of a tax state requires societal compliance, but we depart from her argument by emphasizing societal compliance with the state's information capacity development.

3 The informational foundations of fiscal capacity: a cross-national statistical analysis

In this section we test the main hypothesis developed above using cross-national statistical analyses. We begin by discussing how we operationalize information capacity and extractive capacity, and the trade-offs implied by each of the existing indicators. We then discuss the statistical modelling choices and present our results.

3.1 Methods and data

The operationalization of information capacity puts the analytical spotlight on the main information-gathering technologies employed by states. A particularly central role is played by the administration of a national census, which reflects the ability of the states to generate basic knowledge about society within their borders (Lee and Zhang 2017; Soifer 2013). We draw on the

State Capacity Scores, an original longitudinal dataset recently assembled by Melissa Lee and Nan Zhang (2017 and Lee 2020) for more than 120 countries, as our main measure of information capacity. The State Capacity Scores trace variations in the *legibility* of citizens and their activities by states by concentrating on the quality of population censuses. Specifically, Lee and Zhang draw on demographic techniques to calculate the extent to which the age information included in a population census follows a smooth curve and thus is accurate, or whether the census is characterized by age ‘heaping’ around certain numbers (e.g. 0 and 5). The latter pattern is indicative of limited legibility either because a state does not issue birth certificates and therefore citizens do not know their exact date of birth, or because census-takers face severe constraints in administering the census.

For robustness checks we draw on two other plausible measures of information capacity. The first one is the *information capacity index* developed by the research group State-Making and the Origins of the Global Order in the Long 19th Century and Beyond (STANCE) at the University of Lund (Brambor et al. 2020). This index identifies when a state first had a statistical agency (even if it was later abolished) and when it first conducted a modern population census.³ It also indicates whether a civil and a population register were present in a given year, and calculates graded indexes of census ability and yearbook ability. Second, we use the *cadastre* indicator developed by Michelle D’Arcy and Marina Nistotskaya (2017, 2018). This cross-sectional measure sums the number of years a state has experience in administering a land cadastre, weighted by the quality of the cadastre. This measure thus captures variations in information-gathering on the location and distribution of valuable economic assets within the state’s territorial boundaries, most importantly land. Compared with the State Capacity Scores, both these indicators are more limited in their coverage. The STANCE dataset is geographically biased towards Western Europe and the Americas and includes only 86 polities that have been characterized by relatively durable forms of sovereign statehood for the last 200 years, while the cadastre dataset includes 78 countries, selected because of their relatively durable experience of democratic rule.

To identify differences in extractive capacity we use various tax ratios (all adjusted for GDP) from the ICTD/UNU-WIDER Government Revenue Dataset, widely considered the most accurate and reliable cross-national data source on taxation. The first main measure of interest is direct taxes, which encompass income tax, land tax, personal property tax, and taxes on profits and capital gains. Direct taxes are indicative of high extractive capacity, given that they are paid straight to the state by (usually reluctant) individuals and organizations. We also use personal income taxes and the non-resource component of indirect taxes, which include taxes on goods and services (e.g. VAT) and taxes on international trade, yet exclude resource-driven taxes on exports. Despite being somewhat less demanding to collect than taxes on income, profits, and capital gains, indirect taxes still need an effective tax administration. To cross-check our findings we also rely on measures of total tax revenue and total resource revenue, expecting the latter to be negatively associated with extractive capacity. In Appendix A, we examine the correlations between legibility and taxation, and find strong associations for each of our indicators except total resource revenue. This finding gains further support from the visual inspection of scatter plots (Figures A1–A5) of legibility and the different measures of extractive capacity.

³ When compared with the legibility measure, the STANCE index captures only that a census took place; it does not assess its quality.

3.2 Cross-national statistical analysis

Our statistical model is inspired by Lee and Zhang's (2017) analysis of legibility and its effects on public goods provision outcomes. We use an OLS model with country-decades as the unit of analysis. Our main explanatory variable is legibility, and our dependent variables are the various measures of taxation discussed above. We control for GDP per capita and political regime type (measured by using the Polity index), since economically more developed countries are more effective at taxation while democracies likely enjoy greater tax legitimacy. The geographic variables used by Lee and Zhang (2017) are also included in our model. Higher population density is known to promote tax collection. On the other hand, ruggedness of the terrain hampers both tax collection and census administration. We also include decade-fixed effects to control for period-specific factors, and cluster standard errors by country.

Table 1 shows our main results. Columns 1, 3, 5, and 7 show the results with legibility alone to ensure that there are no 'suppression effects' and that the results do not derive from the inclusion of particular controls. Columns 2, 4, 6, and 8 display the main indicators of interest with all the relevant control variables included.

Table 1: Legibility and variants of taxation: cross-national results

Dependent variable	Total taxes		Direct taxes		Indirect taxes		Total resource revenue		Personal income taxes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Legibility	6.456*** (0.644)	3.054*** (0.963)	2.766*** (0.385)	1.098*** (0.399)	1.358*** (0.384)	2.082*** (0.744)	-0.858* (0.510)	-0.790 (0.662)	1.743*** (0.295)	0.334 (0.326)
GDP per capita		3.346** (1.318)		1.754*** (0.420)		-1.371 (0.975)		0.605 (0.713)		1.387*** (0.390)
Democracy		1.845** (0.801)		0.628** (0.314)		0.497 (0.553)		-1.540 (1.082)		0.761*** (0.243)
Population density		-1.634*** (0.502)		-0.468 (0.372)		-0.358 (0.336)		-1.408* (0.729)		-0.542* (0.305)
Terrain ruggedness		-0.367 (0.644)		-0.131 (0.272)		0.308 (0.402)		-0.504 (0.729)		-0.180 (0.274)
Constant	17.753*** (1.097)	18.389*** (1.022)	5.531*** (0.538)	5.802*** (0.519)	9.173*** (0.694)	9.088*** (0.632)	4.697*** (1.481)	4.207*** (1.288)	2.873*** (0.409)	3.055*** (0.344)
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	225	225	217	217	214	214	130	130	205	205
R ²	0.512	0.614	0.440	0.537	0.118	0.163	0.090	0.227	0.333	0.480
Adjusted R ²	0.504	0.600	0.429	0.519	0.101	0.130	0.061	0.176	0.319	0.458
Residual Std. Error	7.330 (df = 220)	6.579 (df = 216)	3.318 (df = 212)	3.044 (df = 208)	4.196 (df = 209)	4.127 (df = 205)	5.021 (df = 125)	4.704 (df = 121)	2.884 (df = 200)	2.573 (df = 196)

Note: * p, ** p, *** p<0.01.

Source: authors' construction.

Overall, the results point in the hypothesized direction. Legibility is positive and statistically significant for different measures of extractive capacity. It positively predicts higher levels of revenue from overall taxes, direct taxes, and indirect taxes. Moreover, and again in line with our theoretical expectations, this effect does not exist for resource revenue. Since the standard deviation of the legibility measure is 1.01, the coefficients can be interpreted as the change in the dependent variable when a standard deviation in legibility is increased. An increase in legibility by one standard deviation is associated with an increase of 3.1 points in tax revenue, 1.0 point in revenue from direct taxes, and 2.1 points in revenue from indirect taxes as a percentage of GDP.

The fit of our overall statistical model is best for total tax revenues and direct taxes, while its explanatory power is severely limited for indirect taxes. This supports our hypothesis that the ability of states to process reliable information about their subject populations is less relevant to the collection of indirect taxes. The model also does not appear to explain anything about total resource revenues—a finding that might at least in part be driven by the significant drop in cases for this particular taxation measure. In Appendix B, we provide the results of ancillary analyses with different measures of information capacity (the STANCE index and the cadastre index) and using different estimation strategies. Our results are generally robust to these alternative approaches.

In all, our statistical analysis has shown that states with greater capabilities to gather and analyse information also tend to be more effective at taxing their subject population, even when we control for the socio-economic, political, and geographic characteristics of countries. Given that the development of extractive capacity is usually path-dependent (Ertman 1997), we proceed by focusing on the historical moment when the tax state was initially constructed to further examine the relationship between information capacity and fiscal capacity. This critical juncture is what the two historical case studies of Chile and Argentina will explore in greater detail.

4 Information capacity and tax state construction in 19th-century Chile

Scholars of the Latin American state coincide in tracing the emergence of the effective tax state in Chile to the period between 1840 and 1880 (Kurtz 2013; Saylor 2014; Soifer 2015).⁴ During this period, Chile experienced several significant commodity booms (Saylor 2014) and, more broadly, a process of rapid economic modernization (Ortega Martínez 2005). The country also won two wars against its northern neighbours, and some scholars attribute Chilean state development, and the formation of its tax capacity in particular, to these victories (Schenoni 2021). We also seek to explain the development of extractive capacity in Chile, but we do so by exploring intra-state dynamics, namely the ways in which the growth of information capacity underpinned tax state development.

This section is organized accordingly: it begins by detailing the development of extractive capacity in Chile after 1840, drawing on primary and secondary sources. We then turn to showing that the same period saw the development of information capacity. Here we rely largely on the excellent work of Andrés Estefane (2012, 2016, 2017) on the evolution of state statistics in Chile. Third, we show that the development of information capacity underpinned the changes in extractive capacity in these decades in Chile. We focus here on the two mechanisms theorized above (the economies of scope from information capacity and the role of elite compliance with the development of information capacity) as facilitating conditions for increased extraction.

4.1 Fiscal capacity

Between 1840 and 1880, the Chilean tax state sharply expanded its capacity. Internal taxation increased almost 40-fold between 1840 and 1875—growing far faster than population, GDP per capita, or overall government revenue, and marking the emergence of an important complement to the importance of trade taxes. The state also developed the ability to collect a wider range of

⁴ By focusing on this earlier period, we bracket a robust debate in the case literature about the effects of a subsequent commodity boom on the fiscal state.

taxes, which reflected greater capacity. Whereas in 1840, the only internal taxes were monopoly rents on tobacco, playing cards, and salt, over the next two decades the state introduced a transfer tax on the sale of properties, licence fees for a variety of professions, and a series of increasingly standardized and systematized taxes on property value based on the assessment of values of more than 43,000 properties that generated significant revenue for state coffers.⁵

Though direct taxes never represented a large share of taxes (which continued to be generated mainly from customs duties), they did represent a significant amount of revenue, and increased especially in the 1870s, more than doubling between 1874 and 1880. An inheritance tax on wealth was introduced in 1878 and an income tax in 1879 (Sater 1979: 93ff). Though customs revenues continued to dominate state income, these other taxes represented a significant expansion in the state's ability to extract taxes directly from its citizens. Thus, by 1880, the Chilean state taxed wealth, income, inheritance, and the sale of properties; this constellation of revenue sources in the form of direct taxation reflected a massive expansion of extractive capacity; nothing less than the emergence of the modern tax state.

4.2 Information capacity

The state's informational capacity grew sharply during the same era. The Chilean state's limited capacity to collect information prior to 1840 is reflected in its census and cadastre initiatives of the 1830s, when the census reached only about half of the country's administrative units. Its director acknowledged its limitations by announcing that its population figures should be adjusted upward by 10 per cent due to errors and gaps in the enumeration process (Estefane 2017: 50). The cadastre, which relied on local commissions to provide reports on the income of rural properties, did collect information from every province in the country, though it managed to register only about 12,000 properties (Correa 2014: 8).

The country's Statistics Office, provisionally established in 1843 and formally opened in 1847, represented a major step toward centralized and systematic information collection. At times, its ambitious programme of 'national investigation' on a variety of subjects ranging from demography to topography and economic and political conditions (Estefane 2016: 42) exceeded the ability of its agents. Over time, bureaucratic experimentation, increased support from Congress and other government agencies, and increased compliance from society generated more information collection and systematization. This is reflected, for example, in improvements in the census. The 1854 census still relied on parish records and was characterized by obvious irregularities in data quality, but information was received from every administrative unit in the country within six months.

Other state agencies also developed the capacity to collect and organize statistical information. For example, judicial statistics began to be collected in the 1830s, and data on imports and exports in the port of Valparaíso were published from 1835; by 1843 data were collected from all official border crossings (Estefane 2017). Not surprisingly, given the fiscal dependence on trade taxes, commercial statistics 'had reached a higher level of development than other sectors' by the 1840s (Estefane 2017: 98). Medical statistics began to be compiled at a national level in 1845, supplementing the parish records that had previously been the state's source of vital statistics. Standard decimal weights and measures were acquired from France and distributed nationwide in the 1850s (Memoria del Ministerio de Interior 1859: 13). Most notably, the *Anuario Estadístico* or statistical yearbook began its coverage in 1848, though the first volume was not published until

⁵ See Correa (2014) for a detailed history of property taxation in Chile.

1860. Overall, then, by the 1850s we can already see significant gains in information capacity compared with two decades earlier.

The next two decades saw even greater information capacity forged. Census administrators worked steadily to ensure coverage of a mobile rural labour population, to overcome popular resistance, and to accurately count the indigenous populations of the southern frontier; and by the 1870s the systematization and efficiency of the census had greatly increased. State information capacity also improved via the secularization of vital statistics. Throughout the period under consideration, the state worked closely with Catholic Church officials to implement the census. This was possible because the clerical hierarchy saw the census as a sporadic activity that did not compete with its central place in the life of its parishioners. The state's efforts to collect vital statistics on a more regular basis generated greater tensions with a Church that sought to maintain its monopoly over the institutions of birth, death, and marriage. This was reflected in the fact that, although a first bill for a Civil Registry was introduced to Congress in 1868, it was not approved until 1884 (Estefane 2017: 169–70).

In 1869 the post of statistics inspector was created (Estefane 2012). Statistics inspectors conducted visits throughout the country, working to systematize local procedures of data gathering and processing. This initiative resulted in more timely and consistent data collection at the local level after the early 1870s (Estefane 2017: 177ff). Regular visits (*Anuario Estadístico* 1871–2: xiii) provided opportunities for training, for introducing standard practices, for evaluating the quality of local employees, and for overcoming in other ways the effects of autonomy and discretionality that were inherent in postings with multi-faceted job descriptions to the four corners of Chile's extended territory. Since statistics inspectors were responsible for data collection across a range of substantive areas, there was increased professionalization of statistics, which resulted in improvements in information capacity across many kinds of data the state sought to collect—an example of the economies of scope mechanism we suggested above.

4.3 Links and dependencies

Not only did these two developments occur coterminously, but there is strong evidence that information capacity was key to tax state development. A review of official state documents shows that limitations in the state's knowledge about its territory and population were a constant worry to Chilean tax officials in the early decades after independence. Tax governance confronted political, economic, and administrative problems generated by the absence of a definite territorial demarcation, a direct consequence of the lack of up-to-date information on geographical and statistical matters. In 1834, before the development of the state's information apparatus, the Minister of Finance attributed the parlous state of government revenues in part to the reliance on 'arbitrary and inefficient tax collection' necessitated by the lack of a well institutionalized infrastructure of data collection procedures in his agency (*Memoria del Ministerio de Hacienda* [MHAC] 1834: 10). Officials repeatedly cited the lack of statistics about regional administration and commerce as an obstacle to budgeting and fiscal administration, and the absence of reliable population statistics and administrative mapping as an obstacle to the allocation of Congressional seats and the implementation of voting procedures (MHAC 1834: 44–45).

Information capacity served directly to increase the state's extractive capacity. The implementation of the first republican census, conducted in the midst of the independence conflict, was explicitly justified by bureaucrats in the finance ministry to the politicians who had to authorize its adoption as a means to compile the information needed for tax collection (Estefane 2017: 38), and state officials repeatedly echoed this argument over the next decades. Similarly, the initial cadastre of the 1830s allowed the state to introduce direct taxation for the first time and was seen as a first

step toward the implementation of a systematic wealth tax (MHAC 1835: 18), though it included only rural property.

A new cadastral map completed in 1874⁶ included urban property for the first time (Correa 2014: 128), which allowed a tax on urban property value to be imposed in 1878 (Sater 1976: 326). By 1880, this tax, along with inheritance and income taxes, was the sixth largest in revenue, rising to fourth in 1882 and third in 1883 (Sater 1976: 328). Though all these taxes were eliminated during the subsequent nitrate boom, the fact that the state could so quickly generate so much revenue from new direct taxes was clearly a consequence of the information capacity it had developed in the preceding decades. Thus, information capacity provided the state with the data it needed to plan modernizing shifts in its tax structure and to implement those effectively.

The development of information collection, and therefore taxation, was also facilitated by elite compliance. Estefane documents the growing realization by landowners in the late 1860s that greater information about economic activity would provide more stable and predictable commodity prices and thus facilitate their economic decision-making. This view of the state's information collection came to outweigh distrust and the fear that it would be used to increase taxation. The result was that the Sociedad Nacional de Agricultores (national association of large landowners) came to take an active part in efforts to collect data on agricultural production (Estefane 2017: 171). The association, in its capacity as one of the country's most powerful interest groups, pushed state officials at the provincial level to collect more detailed and comprehensive figures, and national officials to expand the nationwide presence of the agrarian bureaucracy. Thus, elite compliance with information collection further contributed to tax state development.

5 Information capacity and tax state construction in 19th-century Argentina

A central feature of the development of the tax state in Argentina was the extensive nature of provincial autonomy during the decades after independence, especially in the so-called Rosista period (1828–52). After this period, two separate political units were created (the Estado de Buenos Aires and the Confederación), which joined forces in 1861 and consolidated Argentina as a single political unit with a federal organization. Like Chile from the 1840s onwards, Argentina after 1860 was characterized by political stability and world market integration.⁷ Our analysis confirms that the increasing ability of the Argentine state to reach its citizens and collect high-quality information crucially contributed to the creation of fiscal capacity. We show that the idea of economies of scope helps explain how improvements in information capacity allowed the Argentine state to establish efficient tax systems (in the sense of maximizing revenue) and the progressive consolidation of more 'information-intensive' taxes. We also show that elite compliance often affected the quantity and quality of information available to tax officials.

Closely following the organization of the Chile case study, the first and the second parts of this section show, respectively, the evolution of fiscal and information capacities in Argentina during the 19th century and up to the First World War. The last part concentrates on the interaction between these capacities and the relevance of the two aforementioned mechanisms.

⁶ Data from this have recently been digitized by Naím Bro Khomasi as part of a larger project on the political effects of land inequality in Chile.

⁷ Mazzuca (2021) argues that this period marked the onset of trade-based state formation in Latin America.

5.1 Fiscal capacity

The Argentine tax state expanded significantly from the 1860s onwards, in terms of both size and tax composition. During the first half of the 19th century tax yields stagnated (Garavaglia 2014; Gelman and Santilli 2006), even though tax levels in per capita terms were among the highest in Latin America, particularly in the Province of Buenos Aires (Garavaglia 2010). Moreover, trade taxes were the prevailing form of taxation, accounting for 83 per cent, 93 per cent, and 86 per cent of total revenues in the Province of Buenos Aires during the periods of liberal reform (1821–28), Rosas's domination (1829–52), and the 1850s, respectively, and for 95 per cent of total receipts of the Argentinean Confederation during the 1850s (Garavaglia 2014). Once the country was unified, total tax revenues jumped from 2 per cent of GDP in 1864 to 5 per cent during the early 1870s, increasing again during the last decade of the 19th century to 7 per cent of GDP and then maintaining this level until WWI (Ferrerres 2005).

Taxation diversified importantly from the 1890s onwards. Tax authorities introduced a variety of taxes on alcohol, beer, matches, sugar, tobacco, and wine in 1890 and 1891. Then, indirect internal taxes started to account for at least 40 per cent of total tax revenues and around 2 per cent of GDP. Given the rapid growth of the Argentine economy during this period, this implies a 12-fold expansion of indirect internal taxes in nominal terms between 1890 and 1913.

5.2 Information capacity

During the second half of the 19th century, information capacity also expanded significantly in Argentina. In the Province of Buenos Aires, the first systematic efforts to publish statistical information had begun with the Razones Estadísticas decree (1821) and the establishment of the Registro Estadístico de la Provincia de Buenos Aires (1822). But these initiatives remained piecemeal and the Rosista period was marked by a general decay in information capacity (Otero 2007).

Things changed during the early 1850s. From 1854 onwards, the regularly published *Registro Estadístico* offered information on issues ranging from trade to population demographics. The *Registro* was part of a broader set of state institutions in the Province of Buenos Aires that sought to collect and analyse relevant information. Among the most prominent was the Departamento Topográfico, which in 1864 published 'the first cadastral map covering most of the legally owned territory of Buenos Aires province' (Gautreau and Garavaglia 2012: 15). Interest in the creation of information capacity was not exclusive to the Province of Buenos Aires. Once Rosas was removed, the Argentine Confederation also created a statistical office, which executed a census in 8 of the state's 13 provinces (Otero 2007).

These institutional antecedents in both the Province of Buenos Aires and the Confederation set the stage for the rapid expansion of information capacity after unification. The National Statistics Office (created in 1864) regularly published trade and migration statistics and successfully implemented the first modern national population census in 1869. This trend continued during the subsequent decades. In 1894, the newly created Dirección General de Estadísticas was in charge of the collection and presentation of statistical information on demography, trade, economic activity, education, public finances, and transport. State officials also started to systematically collect specific sectoral information, as evident in the pursuit of manufacturing censuses in various provinces, the Agrarian Census of 1888, and censuses on national assets. Similarly, the second modern census (1895) provided information on population, manufacturing production, agrarian production, and trade. Sectoral censuses executed in 1908 were followed by the third modern census (1914) (Novick 2002; Otero 2007; Rayes 2016).

An international comparison shows that Argentina, like Chile, was among the first countries to carry out modern censuses.⁸ Furthermore, thanks to the data provided by Somoza and Lattes (1967), we can identify a sizable improvement in legibility between 1869 and 1895,⁹ which, at least to a certain extent, indicates a greater ability of census enumerators to reach the population and compile accurate information.

5.4 Links and dependencies

The development of information capacity can lay the groundwork for further improvements in other areas of state intervention. These economies of scope are clear when we look at the relationship between information capacity and extractive capacity. For instance, the limited revenues generated by the introduction of modern direct taxes in Argentina in the 1820s and 1830s can be explained by two problems from an administrative perspective: taxable assets were self-reported by taxpayers, which allowed underreporting (especially by wealthy people), and there was a lack of administrative capacity to enforce payments. In this context, and not surprisingly, the first collection efforts were outsourced to ‘tax-farmers’. One of the most important challenges of these agents was related to the need to create cadastres that, given the inflationary context, had to be constantly updated. When responsibility for the collection of direct taxes reverted to public administration (1836), this information challenge had to be tackled by the state. This became especially pressing in 1839 when, given the blockade of the Buenos Aires port, the state needed alternative resources to compensate for the fall in trade taxes. Consequently, an economic census of Buenos Aires was run and the number of potential taxpayers increased.¹⁰

The organization of the Departamento Topográfico offers further evidence on these economies of scope. Initially, the accumulation of cadastral information depended on landowners’ willingness to provide information and the effective collection of information by private agents (*agrimensores públicos*). In this individualized, map-based system, public officers only validated cadastral maps *a posteriori* (Gautreau and Garavaglia 2012). In order to cope with these problems and ensure the accumulation of compatible information, the Province adopted a standardization strategy throughout the 1840s and 1850s that reduced the discretion of the *agrimensores públicos*.

At the same time, the *demand* for cadastres by landowners increased. This was due to their higher interest in a better definition of property rights in the context of the progressive integration of the Buenos Aires economy to the world market (Gautreau and Garavaglia 2012). Therefore, thanks to the higher bargaining power of the state over *agrimensores públicos* and to elite compliance, information capacity improved during the mid-19th century. This, in turn, had an important indirect effect on fiscal capacity: cadastres contributed to a better definition of property rights, which in turn could foster exports and, therefore, trade taxes.

⁸ The 1869 census can be defined as a modern one since it was executed by a single government unit; it presented universal and uniform questions; data were collected simultaneously throughout the country; and the questions did not present extra-statistical considerations (Otero 2007). Before 1869, only Western European countries, Nordic countries, and the United States deployed modern censuses. From a Latin American perspective, only Uruguay (1852), Chile (1854), and Costa Rica (1864) ran a modern census before Argentina.

⁹ Following Lee and Zhang’s (2017) operationalization strategy, the Myers Score for Argentina is 23.71 in 1869 and 17.78 in 1895.

¹⁰ This tax potential was not fulfilled during the 1840s because of political instability (Gelman and Santilli 2006; Santilli 2010).

Similarly, the improvement in information capacity that came from the unification of customs and the generation of continuous trade series beginning in 1861 was central in the revenue-maximization strategy that characterized tariff policy during this period (Hora 2009; Tarsitano 2010). Indeed, thanks to information capacity, public authorities could discuss the tax-elasticities of the different types of imports (Memoria del Ministerio de Hacienda 1864), as well as acquire information on market prices in foreign ports that helped them to set the official price to be imposed on imports (Rayes et al. 2020).

The striking rise in indirect internal taxes during the 1890s described above is explained not only by the expansion of a new tax base or the fiscal urgencies derived from the Baring crisis, but also by the greater interest of the Argentinean state in the production of sectoral information after the onset of industrialization during the 1880s (Rocchi 2005). Once more, this was not only a top-down process. For instance, in the preparation of the national census of 1914, public authorities asked for the collaboration of the Sociedad Rural and the Unión Industrial Argentina (Otero 2007: 83), the main lobbies of agrarian and industrial production in the country. In fact, until 1914 the Unión claimed that population censuses had a free-trade bias (Otero 2007: 107).

We also find clear links between the expansion of sectoral information and fiscal capacity. For instance, the Memoria del Ministerio de Hacienda (1897: L–LVI) discusses the effect of tax-rate increases on alcohol production by using monthly statistics on alcohol production, consumption, and stocks from 1895 to 1897. Public officials stated that, whereas information on these variables was available since 1891, they were focusing on the post-1895 period because the establishment of the Oficina de Control y Estadística in 1895 assured the quality and robustness of the data (MHAC 1897: LIV). Overall, this suggests that the effective consolidation of revenue sources beyond trade taxes depended on the availability of information, the quality of which was, at least in part, determined by elite compliance.

6 Conclusion

In this paper we have combined cross-national statistical analysis and in-depth historical case studies of Argentina and Chile to explore the relationship between two crucial aspects of state capacity: information capacity and fiscal capacity. Our empirical findings indicate that information capacity contributes to tax state development. States require accurate information about their subject populations in order to effectively mobilize revenues.

These findings also connect with and contribute to a number of broader debates in the study of states and development. For students of *state capacity* our analysis cautions against treating fiscal capacity as synonymous with overall state strength, given the informational foundations of taxation we have identified for the period under study. This paper similarly questions the well established claim that state capacity is multidimensional. While our findings reinforce the view that state capacity has multiple and interrelated dimensions, they also urge scholars to move beyond this truism and further theorize and empirically explore how core dimensions of state capacity relate to and affect each other.

For research on the rise of *fiscal states* our study brings information capacity, a currently underappreciated factor, to the analytical forefront. Existing scholarship draws on a variety of theoretical approaches to explain why there are persistent differences in the ability of states to tax, including geography (e.g. Herbst 2000; Nunn and Puga 2012; Stasavage 2011), ethnic diversity and other forms of social inequality (e.g. Easterly and Levine 1997; Engerman and Sokoloff 2002), and international war and external threats more generally (e.g. Centeno 2002; Hui 2005; Tilly 1990).

Another body of work emphasizes the crucial role of institutions, in particular colonial institutions, and their long-run consequences for the fiscal state (e.g. Frankema and Booth 2019; Mahoney 2010). While these structural and historical-institutional explanations are certainly useful for explaining enduring differences in fiscal capacity, a growing scholarship shows that they do not pay sufficient attention to intra-state determinants. This is precisely what a focus on information capacity and its relationship to the construction of fiscal states can bring to the table.

We close the paper by pointing to avenues for future research to expand this line of work. Empirically, our cross-national statistical analysis would have benefited from greater geographical and temporal coverage of datasets on information capacity. Thus, one possible future contribution to the study of information capacity and its relationship to fiscal capacity would be an extension of the legibility dataset compiled by Lee and Zhang (2017) to include micro-level historical census data from before 1960, which would help trace the prevalence of age heaping further back in time. Similarly, a cross-national cadastre dataset that builds on D'Arcy and Nistotskaya (2017, 2018) by including multiple observations on the frequency and quality of land cadastres since the early 19th century would further test the claims advanced in this paper and extend their applicability to a wider set of countries.

Another possible extension of this study would be the development of detailed case studies based on multiple primary sources and relevant secondary literature, not just on land cadastres, in other world regions. While our focus on 19th-century Argentina and Chile has illustrated the importance of taking a historical perspective on the construction of information capacity and the fiscal state, systematic cross-regional comparisons would allow further testing of our argument.

This said, the two Latin American states—in many ways typical cases of tax state development, when seen from a global perspective—already illustrate the usefulness of the framework developed in this paper. Thus, we hope to have laid the foundation for a research programme that takes information capacity and its relationship to fiscal capacity seriously when studying tax states and development.

References

- Adcock, R., and D. Collier. (2001). 'Measurement validity: a shared standard for qualitative and quantitative research'. *American Political Science Review*, 95(03): 529–46. <https://doi.org/10.1017/S0003055401003100>
- Anuario Estadístico (various dates). Santiago: Government of Chile.
- Brambor, T., A. Goenaga, J. Lindvall, and J. Teorell (2020). 'The lay [*sic*] of the land: information capacity and the modern state'. *Comparative Political Studies*, 53(2): 175–213. <https://doi.org/10.1177/0010414019843432>
- Bräutigam, D., D. Fjeldstad, and M. Moore (2008). *Taxation and state-building in developing countries: capacity and consent*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511490897>
- Centeno, M.A. (2002). *Blood and debt: war and the nation-state in Latin America*. University Park, PA: Pennsylvania State University Press.
- Centeno, M.A., A. Kohli, and D. Yashar (eds) (2017). *States in the developing world*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9781316665657>
- Correa, L.A. (2014). 'Catastro de propiedad en Chile: orígenes y evolución'. Unpublished PhD dissertation. Barcelona: Universitat de Barcelona.

- D’Arcy, M., and M. Nistotskaya (2017). ‘State first, then democracy: using cadastral records to explain governmental performance in public goods provision’. *Governance*, 30: 193–209. <https://doi.org/10.1111/gove.12206>
- D’Arcy, M., and M. Nistotskaya (2018). ‘The early modern origins of contemporary European tax outcomes’. *European Journal of Political Research*, 57: 47–67. <https://doi.org/10.1111/1475-6765.12214>
- Downing, B. (1992). *The military revolution and political change: origins of democracy and autocracy in early modern Europe*. Princeton, NJ: Princeton University Press.
- Easterly, W., and R. Levine (1997). ‘Africa’s growth tragedy: policies and ethnic divisions’. *The Quarterly Journal of Economics*, 112(4): 1203–50. <https://doi.org/10.1162/003355300555466>
- Engerman, S.L., and K.L. Sokoloff (2002). ‘Factor endowments, inequality, and paths of development among new world economics’. *Economia*, 3(1): 41–109. <https://doi.org/10.1353/eco.2002.0013>
- Ertman, T. (1997). *Birth of the Leviathan: building states and regimes in medieval and early modern Europe*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511529016>
- Estefane, A. (2012). ‘Burócratas ambulantes. Movilidad y producción de conocimiento estadístico en Chile, 1860-1873’. *Revista Enfoques*, X(17): 123–46.
- Estefane, A. (2016). ‘La institucionalización del sistema estadístico chileno: debates y problemas prácticos 1843-1851’. *Estudios Sociales del Estado*, 2(4): 35–73. <https://doi.org/10.35305/ese.v2i4.90>
- Estefane, A. (2017). ‘Elusive numbers: state knowledge and bureaucratic organization in Chile 1750–1911’. Unpublished PhD Dissertation. Stony Brook University, New York.
- Evans, P. (1995). *Embedded autonomy: states and industrial transformation*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9781400821723>
- Ferreres, O. (2005). *Dos siglos de economía argentina, 1810-2004: historia argentina en cifras*. Buenos Aires: Fundación Norte Sur.
- Frankema, E., and A. Booth (eds) (2019). *Fiscal capacity and the colonial state in Asia and Africa, c.1850–1960*. New York: Cambridge University Press. <https://doi.org/10.1017/9781108665001>
- Garavaglia, J.C. (2010). ‘Algunos aspectos preliminares acerca de la “transición fiscal” en América Latina, 1800-1850’. *Illes i Imperis*, 13: 159–92.
- Garavaglia, J.C. (2014). ‘Dispute for the nation: incomes and fiscal policy in the construction of the Argentine state, 1850–1865’. *Investigaciones de Historia Económica*, 10(1): 34–45. <https://doi.org/10.1016/j.ihe.2013.01.002>
- Gautreau, P., and J.C. Garavaglia (2012). ‘The weak-state cadastre: administrative strategies to build territorial knowledge in post-colonial Argentina (1824 to 1864)’. *Cartographica*, 47(1): 29–49. <https://doi.org/10.3138/cart0.47.1.29>
- Gelman, J., and D. Santilli (2006). ‘Entre la eficiencia y la equidad. Los desafíos de la reforma fiscal en Buenos Aires. Primera mitad del siglo XIX’. *Revista de Historia Económica - Journal of Iberian and Latin American Economic History*, 24(3): 491–520. <https://doi.org/10.1017/S0212610900000616>
- Hanson, J.K., and R. Sigman (2021). ‘Leviathan’s latent dimensions: measuring state capacity for comparative political research’. *The Journal of Politics*, 83(4): 1495–510. <https://doi.org/10.1086/715066>
- Hendrix, C. (2010). ‘Measuring state capacity: theoretical and empirical implications for the study of civil conflict’. *Journal of Peace Research*, 47(3): 273–85. <https://doi.org/10.1177/0022343310361838>
- Herbst, J. (2000). *States and power in Africa: Comparative lessons in authority and control*. Princeton, NJ: Princeton University Press.
- Hora, R. (2009). ‘La política económica del proteccionismo en la Argentina, 1870-1914’. *Investigaciones y Ensayos*, 58: 237–83.

- Hui, V. (2005). *War and state formation in ancient China and early modern Europe*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511614545>
- Kurtz, M. (2013). *Latin American state building in comparative perspective: social foundations of institutional order*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9781139019668>
- Lee, M. (2020). *Crippling Leviathan: how foreign subversion weakens the state*. Ithaca, NY: Cornell University Press. <https://doi.org/10.7591/cornell/9781501748363.001.0001>
- Lee, M., and N. Zhang. (2017). ‘Legibility and the informational foundations of state capacity’. *The Journal of Politics*, 79(1): 118–32. <https://doi.org/10.1086/688053>
- Leftwich, A. (2008). ‘Developmental states, effective states and poverty reduction: the primacy of politics’. UNRISD Project on Poverty Reduction and Policy Regimes. New York: United Nations Research Institute for Social Development.
- Levi, M. (1988). *Of rule and revenue*. Berkeley, CA: University of California Press.
- Mahoney, J. (2010). *Colonialism and postcolonial development: Spanish America in comparative perspective*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750328>
- Memoria del Ministerio de Hacienda (various dates). Buenos Aires: Government of Argentina.
- Memoria del Ministerio de Hacienda (various dates). Santiago: Government of Chile.
- Memoria del Ministerio de Interior (various dates). Santiago: Government of Chile.
- Novick, S. (2002). *Legislación referida a censos y estadísticas en la Argentina: 1854-1991*. Buenos Aires: IIGG/UBA.
- Nunn, N., and D. Puga (2012). ‘Ruggedness: the blessing of bad geography in Africa’. *The Review of Economics and Statistics*, 94(1): 20–36. https://doi.org/10.1162/REST_a_00161
- Olson, M. (1993). ‘Dictatorship, democracy, and development’. *The American Political Science Review*, 87(3): 567–76. <https://doi.org/10.2307/2938736>
- Ortega Martínez, L. (2005). *Chile en ruta al capitalismo: cambio, euforia y depresión 1850-1880*. Santiago: LOM Ediciones.
- Otero, H. (2007). *Estadística y nación. Una historia conceptual del pensamiento censal de la Argentina moderna, 1869-1914*. Buenos Aires: Prometeo.
- Queralt, D. (2019). ‘War, international finance, and fiscal capacity in the long run’. *International Organization*, 73(4): 713–53. <https://doi.org/10.1017/S0020818319000250>
- Rayes, A. (2016). ‘La construcción de las estadísticas oficiales argentinas de exportación, c. 1880-1930.’ *Estudios Sociales Del Estado*, 2(4): 96–120. <https://doi.org/10.35305/ese.v2i4.92>
- Rayes, A., R. Castro, and F. Ibarra (2020). ‘Números oscuros. La valoración de las importaciones argentinas, c. 1870-1913’. *Revista Uruguaya de Historia Económica*, 17: 25–48. doi.org/10.47003/RUHE/10.17.02
- Rocchi, F. (2005). *Chimneys in the desert: industrialization in Argentina during the export boom years, 1870–1930*. Stanford, CA: Stanford University Press.
- Rueschemeyer, D. (2005). ‘Building states – inherently a long-term process. An argument from theory’. In M. Lange and D. Rueschemeyer (eds), *States and development: historical antecedents of stagnation and advance*. New York: Palgrave Macmillan, pp. 143–64. https://doi.org/10.1057/9781403982681_7
- Sandbrook, R., M. Edelman, P. Heller, and J. Teichman. (2007). *Social democracy in the global periphery: origins, challenges, prospects*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511491139>
- Santilli, D.V. (2010). ‘El papel de la tributación en la formación del Estado. La contribución directa en el siglo XIX en Buenos Aires’. *América Latina en la Historia Económica*, 17(1): 31–63. <https://doi.org/10.18232/alhe.v17i1.428>

- Sater, W. (1976). 'Economic nationalism and tax reform in late nineteenth-century Chile'. *The Americas*, 33(2): 311–35. <https://doi.org/10.2307/980789>
- Sater, W. (1979). 'Chile and the world depression of the 1870s'. *Journal of Latin American Studies*, 11(1): 67–99. <https://doi.org/10.1017/S0022216X00022306>
- Savoia, A., and K. Sen (2015). 'Measurement, evolution, determinants, and consequences of state capacity: a review of recent research'. *Journal of Economic Surveys*, 29: 441–58. <https://doi.org/10.1111/joes.12065>
- Saylor, R. (2013). 'Concepts, measures, and measuring well: an alternative outlook'. *Sociological Methods & Research*, 42(3): 354–91. <https://doi.org/10.1177/0049124113500476>
- Saylor, R. (2014). *State building in boom times: commodities and coalitions in Latin America and Africa*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199364954.001.0001>
- Schenoni, L. (2021). 'Bringing war back in: victory and state formation in Latin America'. *American Journal of Political Science*, 65(2): 405–21. <https://doi.org/10.1111/ajps.12552>
- Schumpeter, J. (1954). 'The crisis of the tax state'. International Economic Papers 4. Translations prepared for the International Economic Association. London: Macmillan.
- Scott, J. (1998). *Seeing like a state: how certain schemes to improve the human condition have failed*. New Haven, CT: Yale University Press.
- Slater, D. (2010). *Ordering power: contentious politics and authoritarian Leviathans in Southeast Asia*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511760891>
- Soifer, H.D. (2008). 'State infrastructural power: approaches to conceptualisation and measurement'. *Studies in Comparative International Development*, 43: 231–51. <https://doi.org/10.1007/s12116-008-9028-6>
- Soifer, H.D. (2013). 'State power and the economic origins of democracy'. *Studies in Comparative International Development*, 48: 1–22. <https://doi.org/10.1007/s12116-012-9122-7>
- Soifer, H.D. (2015). *State building in Latin America*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9781316257289>
- Soifer, H.D., and M. vom Hau. (2008). 'Unpacking the strength of the state: the utility of state infrastructural power'. *Studies in Comparative International Development*, 43: 219–30. <https://doi.org/10.1007/s12116-008-9030-z>
- Somoza, J., and A. Lattes (1967). 'Muestras de los dos primeros censos nacionales de población, 1869 y 1895'. Buenos Aires: Centro de Investigaciones Sociales/Instituto Torcuato di Tella.
- Stasavage, D. (2011). *States of credit: size, power, and the development of European polities*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9781400838875>
- Tarsitano, A. (2010). 'Período 1866-1880'. In V. Oscar and H. García (eds), *Historia de la Tributación Argentina*. Buenos Aires: Asociación Argentina de Estudios Fiscales, pp. 181–215.
- Thies, C. (2005). 'War, rivalry, and state building in Latin America'. *American Journal of Political Science*, 49: 451–65. <https://doi.org/10.1111/j.1540-5907.2005.00134.x>
- Tilly, C. (1985). 'War making and state making as organized crime'. In P. Evans, D. Rueschemeyer, and T. Skocpol (eds), *Bringing the state back*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511628283.008>
- Tilly, C. (1990). *Coercion, capital, and European states, AD 990–1990*. Cambridge, MA: B. Blackwell.
- vom Hau, M. (2008). 'State infrastructural power and nationalism: comparative lessons from Mexico and Argentina'. *Studies in Comparative International Development*, 43(December): 334–54. <https://doi.org/10.1007/s12116-008-9024-x>
- vom Hau, M. (2012). 'State capacity and inclusive development: new challenges and directions'. Working Paper 2. University of Manchester: Effective States and Inclusive Development Research Centre (ESID). <https://doi.org/10.2139/ssrn.2141771>

Appendix A: Initial tests of association

As preliminary evidence for the relationship we propose, we explore correlations between the average legibility score and different measures of taxation for the 1980–2010 period, with countries/country-decades as units of analysis. If we are correct that information capacity is associated with extractive capacity, then the correlations should be positive and statistically significant. The results shown in Table A1 provide support for this expectation. The correlations are relatively strong and in the predicted direction.

Table A1: Correlations between legibility and different forms of taxation

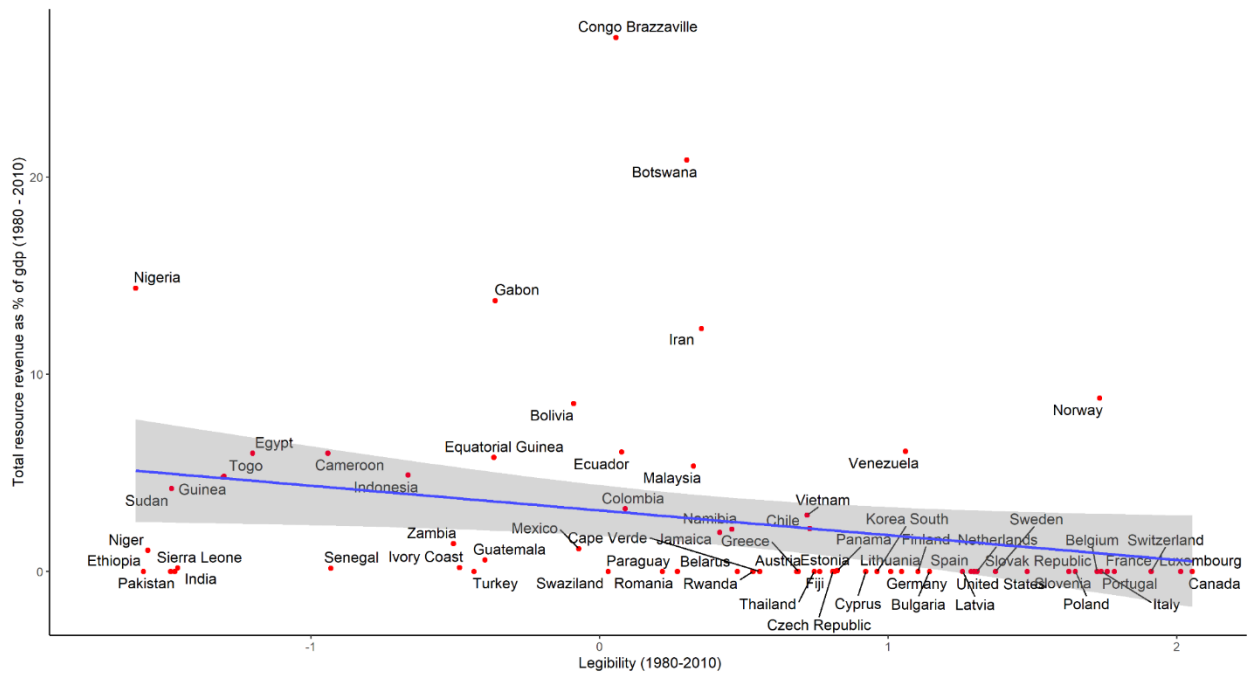
	Total taxes	Direct taxes	Personal income taxes	Indirect taxes	Total resource revenue
Legibility (1980–2010)	0.769	0.711	0.555	0.469	-0.258

Source: authors' construction.

Further evidence comes from visual inspection. The scatter plots shown in Figures A1 to A5 cross average legibility scores (1980–2010) on the X-axis with measures of income from direct taxes (1980–2010), income from indirect taxes (1980–2010), income from personal income taxes (1980–2010), total tax revenue (1980–2010), and total resource revenue (1980–2010) on the Y-axis, again with countries as the unit of analysis. Especially for direct taxes, the scatter plots reveal a positive association with legibility. We also note that the greater variation of direct tax income found in countries with relatively high legibility scores (above 0) suggests that legibility appears to be necessary, but not sufficient, for effective direct taxation to occur. The scatter plot for total resource revenue reinforces the expectation that legibility is positively related to taxation. Moreover, and again as predicted, the relationship between legibility and total resource revenue is, if anything, a negative association, since resource revenues require little information capacity compared with taxation. Closer inspection also shows that the relationship between legibility and indirect taxes is rather weak, possibly because the taxation of trade and services is comparatively less 'information-intense' than that of direct taxes.

In sum, these initial tests of association based on the most comprehensive quantitative data available establish that information capacity and extractive capacity are positively related to each other and provide tentative support for our hypothesis. For the 1980–2010 period at least, information capacity appears to be necessary for effective extraction.

Figure A5: Scatterplot of legibility and resource revenue



Source: authors' construction.

Appendix B: Robustness checks

A number of robustness checks indicate that our finding of a positive association between information capacity and the extractive capacity of states holds even when other measures of information capacity and/or different modelling strategies are used. We first show results for two alternative statistical models. The model used in Table B1 includes country-fixed effects, which allow us to control for time-invariant characteristics of particular countries. In this model, legibility is again positive and statistically significant for overall and direct taxes, while the coefficient also points in the right direction for indirect taxes. Table B2 shows results from a random effects model. Here, we find evidence that legibility positively predicts higher levels of revenue from overall taxes, direct taxes, and indirect taxes.

The remaining tables show that when we employ alternative measures of information capacity, the results are broadly supportive of our findings for legibility. Tables B3–B5 show that replacing legibility with the STANCE information capacity index (Brambor et al. 2020) as the main explanatory variable in the model leads to less conclusive results, but this is arguably because the STANCE index is more attuned to measuring the ‘performative’ aspects of census-taking and related state activities than the actual quality of the census. Table B6 presents the results for the cadastre index as yet another alternative measure of information capacity. Given that the cadastre index is cross-sectional, we had to adjust our modelling strategy and aggregate across the 1980–2010 period. The results are less powerful when compared with our findings for legibility, but the coefficients still point into the predicted direction.

Table B1: Legibility and variants of taxation: results of two-way linear fixed effects regression

Dependent variable	Total taxes		Direct taxes		Indirect taxes		Total resource revenue		Personal income taxes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Legibility	2.791*** (0.752)	1.840** (0.837)	1.593*** (0.341)	0.932** (0.390)	0.821** (0.398)	0.512 (0.473)	-0.290 (0.698)	-0.192 (0.780)	1.645*** (0.366)	1.122*** (0.412)
GDP per capita		6.590*** (1.645)		1.241** (0.609)		0.050 (0.737)		2.224* (1.305)		1.655** (0.647)
Democracy		0.427 (0.579)		0.375 (0.264)		-0.442 (0.324)		0.758 (0.508)		0.787*** (0.270)
Population density		-5.589** (2.801)		0.425 (1.179)		2.175 (1.454)		-5.581** (2.178)		-1.885 (1.266)
Observations	225	225	217	217	214	214	130	130	225	225
R ²	0.100	0.210	0.156	0.237	0.035	0.065	0.003	0.118	0.141	0.241
Adjusted R ²	-0.626	-0.462	-0.544	-0.433	-0.771	-0.763	-1.075	-0.928	-0.564	-0.416
F-statistic	13.763*** (df = 1; 124)	8.060*** (df = 4; 121)	21.875** (* (df = 1; 118)	8.944*** (df = 4; 115)	4.253** (df = 1; 116)	1.952 (df = 4; 113)	0.173 (df = 1; 62)	1.978 (df = 4; 59)	20.199** (* (df = 1; 123)	9.550*** (df = 4; 120)

Note: p-values: * < 0.05; ** < 0.01; *** < 0.001

Source: authors' construction.

Table B2: Legibility and variants of taxation: results of random effects regression

Dependent variable	Total taxes		Direct taxes		Indirect taxes		Total resource revenue		Personal income taxes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Legibility	5.669***	2.238***	2.418***	1.075***	1.341***	1.235***	-0.85*	-0.778	2.097***	0.700**
GDP per capita		5.843*** (0.765)		1.858*** (0.334)		0.265 (0.417)		0.021 (0.740)		1.799*** (0.372)
Democracy		0.388 (0.449)		0.459** (0.212)		-0.102 (0.265)		-0.110 (0.435)		0.575** (0.225)
Population density		-1.889*** (0.631)		-0.427 (0.297)		-0.073 (0.389)		-1.833*** (0.603)		-0.997*** (0.327)
Constant	18.774*** (0.780)	18.605*** (0.659)	5.657*** (0.337)	5.613*** (0.303)	9.379*** (0.405)	9.397*** (0.409)	2.98*** (0.633)	2.821*** (0.624)	5.855*** (0.368)	5.864*** (0.333)
Observations	225	225	217	217	214	214	130	130	225	225
R ²	0.428	0.584	0.362	0.479	0.216	0.219	0.032	0.103	0.290	0.414
Adjusted R ²	0.425	0.576	0.359	0.469	0.212	0.204	0.025	0.074	0.287	0.403
F-statistic	164.547***	307.450***	121.800***	194.496***	57.862***	58.028***	4.247**	14.353***	90.107***	154.663***

Note: p-values: * < 0.05; ** < 0.01; *** < 0.001

Source: authors' construction.

Table B3: STANCE information capacity index and variants of taxation: results of preferred model with time-fixed effects and country-clustered standard errors (see Table 1)

Dependent variable	Total taxes		Direct taxes		Indirect taxes		Total resource revenue		Personal income taxes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
STANCE information capacity	15.058 (11.074)	-3.380 (8.641)	3.790 (5.038)	-5.655 (3.965)	4.005 (3.518)	2.132 (4.322)	0.589 (4.032)	5.206 (4.187)	7.372 (4.857)	0.475 (3.863)
GDP per capita		7.584*** (1.775)		3.549*** (0.839)		0.027 (0.668)		-0.625 (0.912)		2.574*** (0.678)
Democracy		1.728 (1.878)		0.509 (0.673)		0.113 (0.936)		-2.356* (1.220)		0.625 (0.538)
Population density		0.153 (0.904)		-0.877 (0.823)		0.073 (0.405)		-1.225* (0.643)		-0.423 (0.654)
Terrain ruggedness		-2.329* (1.192)		-0.219 (0.527)		-0.994* (0.571)		0.705* (0.410)		-0.384 (0.607)
Constant	10.062 (8.101)	19.031*** (5.939)	3.584 (3.709)	7.973*** (2.619)	6.884*** (2.453)	8.587*** (2.935)	2.765 (2.595)	0.982 (2.146)	-1.114 (3.416)	2.171 (2.422)
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	110	110	111	111	111	111	70	70	105	105
R ²	0.187	0.601	0.094	0.597	0.085	0.134	0.049	0.325	0.141	0.523
Adjusted R ²	0.156	0.569	0.060	0.565	0.051	0.066	-0.009	0.237	0.107	0.483
Residual Std. Error	9.693 (df = 105)	6.924 (df = 101)	4.755 (df = 106)	3.235 (df = 102)	3.316 (df = 106)	3.289 (df = 102)	3.429 (df = 65)	2.981 (df = 61)	3.996 (df = 100)	3.041 (df = 96)

Note: p-values: * < 0.05; ** < 0.01; *** < 0.001

Source: authors' construction.

Table B4: STANCE information capacity index and variants of taxation: results of two-way linear fixed effects regression

Dependent variable	Total taxes		Direct taxes		Indirect taxes		Total resource revenue		Personal income taxes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
STANCE information capacity	10.445* (5.385)	2.857 (5.239)	3.947 (2.404)	0.526 (2.103)	1.828 (2.624)	1.246 (2.497)	2.989 (3.859)	2.874 (2.609)	6.593** (2.723)	3.908 (2.754)
GDP per capita		6.771** (2.969)		1.621** (0.787)		-0.025 (0.928)		6.727*** (1.706)		1.809* (1.031)
Democracy		-1.410 (0.953)		-0.139 (0.385)		-1.332*** (0.467)		-0.763 (0.572)		0.633 (0.504)
Population density		4.072 (5.036)		4.469** (1.718)		5.254** (2.021)		-20.88*** (3.033)		-0.540 (2.238)
Observations	110	110	111	111	111	111	70	70	114	114
R ²	0.056	0.288	0.039	0.373	0.007	0.231	0.017	0.630	0.080	0.198
Adjusted R ²	-0.609	-0.272	-0.601	-0.095	-0.680	-0.365	-0.938	0.201	-0.551	-0.416
F-statistic	3.763* (df = 1; 64)	6.170*** (df = 4; 61)	2.695 (df = 1; 66)	9.371*** (df = 4; 63)	0.485 (df = 1; 65)	4.643*** (df = 4; 62)	0.600 (df = 1; 35)	13.597** (df = 4; 32)	5.864** (df = 1; 67)	3.952*** (df = 4; 64)

Note: p-values: *<0.05; **<0.01; ***<0.001

Source: authors' construction.

Table B5: STANCE information capacity index and variants of taxation: results of random effects regression

Dependent variable	Total taxes		Direct taxes		Indirect taxes		Total resource revenue		Personal income taxes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
STANCE information capacity	17.07*** (4.946)	5.208 (4.530)	5.894*** (2.232)	0.506 (1.993)	4.454** (2.043)	2.927 (2.163)	-0.041 (2.703)	3.392 (2.895)	10.32*** (2.450)	5.20** (2.418)
GDP per capita		9.208*** (1.170)		3.571*** (0.455)		1.243*** (0.475)		-1.330* (0.751)		2.550*** (0.517)
Democracy		-1.117 (0.811)		-0.121 (0.371)		-1.032** (0.419)		-0.205 (0.601)		0.477 (0.456)
Population density		-0.115 (1.118)		-0.466 (0.530)		0.280 (0.520)		-1.573** (0.631)		-0.893 (0.544)
Constant	12.40*** (3.812)	14.764*** (3.145)	3.601** (1.714)	5.061*** (1.395)	7.111*** (1.499)	8.052*** (1.500)	1.805 (1.972)	0.634 (1.939)	0.317 (1.824)	1.883 (1.650)
Observations	110	110	111	111	111	111	70	70	114	114
R ²	0.299	0.594	0.170	0.517	0.270	0.309	0.006	0.139	0.260	0.474
Adjusted R ²	0.293	0.578	0.162	0.499	0.263	0.283	-0.008	0.086	0.253	0.455
F-statistic	43.868 ***	150.773 ***	21.244 ***	112.638 ***	39.515 ***	46.418 ***	0.424	10.432 **	38.140 ***	97.240 ***

Note: p-values: *<0.05; **<0.01; ***<0.001

Source: authors' construction.

Table B6: Cadastre index and variants of taxation: results of cross-sectional regression

Dependent variable	Total taxes		Direct taxes		Indirect taxes		Total resource revenue		Personal income taxes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cadaster	0.049*** (0.012)	0.012 (0.011)	0.015** (0.006)	0.001 (0.006)	0.004 (0.005)	0.004 (0.007)	-0.010* (0.005)	-0.007 (0.005)	0.016*** (0.005)	-0.0001 (0.006)
GDP per capita		7.267*** (1.629)		3.368*** (0.832)		-0.053 (0.930)		-1.040 (0.938)		2.906*** (0.816)
Democracy		1.229 (3.420)		0.036 (1.724)		0.360 (1.955)		-0.674 (2.165)		-0.376 (1.822)
Population density		-0.509 (1.113)		-0.563 (0.563)		-0.590 (0.635)		-1.483** (0.550)		-0.159 (0.555)
Terrain ruggedness		-0.825 (0.992)		0.111 (0.505)		0.049 (0.568)		-0.127 (0.507)		-0.618 (0.488)
Constant	19.217*** (2.086)	18.345*** (2.643)	6.144*** (1.033)	5.857*** (1.304)	10.438*** (0.896)	10.230*** (1.512)	3.008*** (0.999)	3.963** (1.734)	2.937*** (0.932)	3.314** (1.315)
Observations	59	59	55	55	57	57	42	42	56	56
R ²	0.223	0.583	0.111	0.503	0.012	0.031	0.083	0.299	0.138	0.416
Adjusted R ²	0.210	0.543	0.094	0.453	-0.006	-0.064	0.060	0.202	0.122	0.358
Residual Std. Error	9.646 (df = 57)	7.332 (df = 53)	4.615 (df = 53)	3.588 (df = 49)	4.064 (df = 55)	4.180 (df = 51)	3.656 (df = 40)	3.368 (df = 36)	4.155 (df = 54)	3.553 (df = 50)

Note: p-values: *<0.05; **<0.01; ***<0.001

Source: authors' construction.