POLICY RESEARCH WORKING PAPER 10261

Using ORBIS to Build a Global Database of Firms with State Participation

Andrea Dall'Olio
Tanja Goodwin
Martha Martinez Licetti
Jan Orlowski
Fausto Patiño Peña
Francis Ratsimbazafy
Dennis Sanchez-Navarro

WORLD BANK GROUP

Abstract

This paper develops a novel methodology to construct a harmonized cross-country database of the state's footprint in markets: the Businesses of the State database. The methodology of the database is built on three criteria—(i) a harmonized definition of state-owned enterprises, (ii) identification of direct and indirect state ownership linkages at the national and subnational levels across the corporate sector, and (iii) classification of economic activities depending on their efficiency rationale—which conceptualize a framework to trace state presence in the corporate sector across economic activities. The database is constructed leveraging different firm-level data sources including the ORBIS Global Database, as the primary data source, which is then complemented with supplementary data sources (EMIS Intelligence, Factiva, Worldscope, Pitchbook, among others) to mitigate ORBIS's data limitations across countries and regions. The Businesses of the State database identifies an unprecedented number of firms with state participation across countries and economic activities, as well as providing novel insights on financial performance, economic performance, and governance of state-owned enterprises. A deep-dive analysis of 36 countries within the Businesses of the State database shows that 69 percent of state-owned enterprises operate in competitive activities (low efficiency-rationale for state participation), 16% are in partially contestable industries (moderate efficiency rationale), and 15 percent are natural monopolies (strong efficiency rationale). Furthermore, this analysis suggests that performance-based productivity of state-owned enterprises (revenue per worker) is negatively correlated with government control variables, such as government shareholding percentage and direct versus indirect government ownership.

Using ORBIS to Build a Global Database of Firms with State Participation

Andrea Dall'Olio, Tanja Goodwin, Martha Martinez Licetti, Jan Orlowski, Fausto Patiño Peña, Francis Ratsimbazafy, Dennis Sanchez-Navarro

Acknowledgments

1. Introduction

State-owned enterprises (SOEs) play a large and increasing role in the global economy. According to the International Monetary Fund (IMF), [1] global assets of SOEs more than tripled between 2000 and 2018, from US$13 trillion to US$45 trillion, about half of global GDP. SOEs are also large employers in many countries. For instance, despite several decades of privatization, SOEs still account for half of all public employment in EBRD countries (EBDR, 2021). An analysis of the role of firms with state participation across countries requires firm-level data. Aggregate data on SOEs can provide a fair overview of the relative size of the state footprint in terms of GDP or total employment. However, such aggregate figures provide overall estimates that are not precise enough for meaningful cross-country comparisons, due to differences in state ownership definitions and aggregation considerations. Hence, researchers analyzing government public provision of goods and services through SOEs, and the effect of SOEs on productivity, growth, and competitiveness, require access to firm-level data using a consistent definition of SOEs to allow for better cross-country comparisons. Firm-level data are also essential to analyze SOEs' economic implications, design SOE reform plans, and assess the adjustment costs and mitigation measures considering the interrelations of SOEs across multiple sectors in the economy. Firm-level data also allows for more targeted programs, for example regarding divestiture measures, management arrangements, corporate governance, and regulatory reforms. However, reform strategies are often constrained or delayed by the lack of readily available information.

This paper describes a novel methodology to develop a harmonized cross-country firm-level database of state-owned enterprises, the Businesses of the State (BOS) database, which maps the state footprint within the corporate sector and across economic activities.[2] The database is based on a uniform definition, a global data source, and a standardized protocol to include supplementary information from complementary sources beyond ORBIS (e.g., business registries, central depositories, central oversight bodies, MoF) (See Annex 3). This database is more comprehensive when compared to former or parallel efforts since it covers over 80 countries doubling the size of countries compared to the OECD coverage or quadruples the sample size of the IMF databases, focuses on developing countries, expands on the firm-level indicators beyond financial performance, covers all relevant economic sectors including natural monopolies and financial sector that are excluded in the IMF or EU databases, and does not trim the sample by any threshold on size (See Annex 1). The global BOS database provides unique country and firm-level information for research and policy interventions on areas such as the presence of SOEs by economic sector, the economic performance of SOEs, government ownership structures, and governance.

The main contributions of the global BOS database go beyond existing efforts expanding the coverage. First, the global BOS database builds on a harmonized definition that can be applied systematically across countries and separate from country-specific definitions that hinder comparability. Second, it not only covers direct linkages between a company and a government, but also identifies and spells out all indirect ownership linkages allowing to trace state participation throughout multiple levels (e.g., subsidiaries of subsidiaries of companies with state participation). Third, it allows to trace state participation in domestic markets but also across borders.[3] Fourth, the database captures the ability of the state to influence markets not only through majority owned companies, but also by lowering the

---

[1] https://blogs.imf.org/2020/05/07/state-owned-enterprises-in-the-time-of-covid-19/ (Gaspar, 2020).

[2] The database not only captures the conventional and traditional SOEs that are often linked to majority control of the state, but also expands the spectrum of entities in which the state can have a significant ability to influence through other businesses of the state.

[3] The ownership trees allow to identify the foreign based SOEs linked to a specific government (e.g., Angolan SOEs operating across borders). However, given the limitations regarding the data validation of the financial information of foreign based SOEs as well as data collection on the governance indicators, only few indicators of these companies are included in the global BOS database such as the company identification, company name, and countries of operation.

threshold of state ownership to 10%. This feature provides flexibility to assess and filter different subsets of companies (e.g., majority, blocking minority vs. minority) and run robustness checks across firms, but also to better understand how the nature and variation in ownership links can be important to understand firms' performance and the types of sectors where those are involved. Fifth, it covers both national and subnational governments, allowing to analyze how municipal governments can also influence markets, which is particularly relevant for decentralized economies. Sixth, it covers all relevant economic sectors with potential implications in the markets beyond traditional infra and network sectors offering a full overview of the state footprint even in commercial sectors without restricting by number of firms or any threshold in size. Last, the database compiles, complements, and validates information beyond ORBIS using official sources and government counterparts, minimizing the omission and measurement errors found in ORBIS.

Since control cannot be known ex ante, this paper proposes a definition of SOEs which allows to identify SOEs, based on the quantitative measurement of the state participation in firms, at different levels of ownership structure. The proposed SOE definition builds on the International Monetary Fund's (IMF) Government Finance Statistics Manual 2014 definition of SOEs. The concept of government control is at the heart of the definition of SOEs. As documented by the OECD, statistics show a tendency of governments to partially divest from SOEs up to a point where firms are no longer considered as SOEs according to national definitions, while still holding non-trivial, controlling stakes (OECD, 2009). Since it is impossible to identify state control without verifying each individual company's ownership structure, including the presence of golden (veto power) shares, we propose to use 10% as a proxy for state control. The use of a comparatively low ownership threshold allows for sensitivity and comparative analyses of state footprint for firms at multiple ownership thresholds levels, including the 25% and 50% commonly used thresholds and as well as cases of minority state control below 25%.

The proposed methodology reconstructs the full ownership structure linking each firm to the state as its ultimate shareholder, including foreign registered firms. For the most part and in contrast to the methodology presented here, existing SOE databases do not capture firms indirectly owned by the state (i.e., subsidiaries of state-owned enterprises), and only a limited number of them capture SOEs at the subnational level. The Global BOS database collects information about firms directly and indirectly linked to the state, mapping out all the subsidiaries of a government-controlled firm operating inside and outside each country with the help of a novel algorithm.

The database adopts a disaggregated sector taxonomy developed by Dall'Olio et. al. (2022) to classify SOEs based on industries' inherent technological features and market failures. There is a large heterogeneity of sectors in which SOEs are present and efficiency rationale for SOE participation in these sectors varies. The taxonomy provides a tool to assess the economic rationale for SOE presence and also serves to triage SOE interventions.[4] The three categories of classification of sectors are: natural monopolies, partially contestable sectors, and competitive sectors. The methodology to construct the sector taxonomy in Dall'Olio et. al. (2022) is complementary to this paper.

The Global BOS database leverages ORBIS, the proprietary firm-level data from Bureau van Dijk, as a primary source to construct country-level SOE data sets.[5] As pointed out by (Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, & Yesiltas, 2015), ORBIS is a good starting point to construct a firm-level data set because it provides administrative data for over 130 million entities worldwide covering all sectors

---

[4] For the alternatives of SOE reform and also different approaches based on the type of sector, please also refer to the CPSD SOE Knowledge note (Sanchez-Navarro, Goodwin, & Kikeri, SOE CPSD Knowledge note, 2021) and Private sector toolkit (WB, forthcoming).
[5] Also known as AMADEUS product for the ECA region.

in the economy. For each firm, ORBIS also provides a large set of indicators including information on the firms' ownership structure as well as their income statements and balance sheet data.

One of the key obstacles of identifying SOEs in ORBIS is the lack of a specific variable for this identification. The closest proxy variable provided by ORBIS is denoted as the Global Ultimate Owner (GUO), defined as the individual or entity at the top of the corporate ownership structure (majority owner). In theory, one could obtain a list of SOEs by selecting firms whose GUOs are government public authorities. However, the information provided by the GUO is limited to entities of which the state or public authority owns 25% or more (GUO25), or 50% or more (GUO50), and state ownership in SOEs could be lower and still provide the ability of the governments to influence over firms (e.g., golden shares). Moreover, the GUO variable does not include every shareholder at 25% or 50% or more level of ownership; it includes only the largest shareholder at the 25% or 50% threshold level, which may not be a state or a government. As a result, utilizing the GUO variable will yield an incomplete list of SOEs and firms with state participation – and therefore underestimate the businesses of the state. Furthermore, the GUO variable tends to misclassify some firms as private even if they are owned by the government.

The Global BOS database leverages a large set of alternative data sources to compensate for ORBIS's substantial coverage variation across countries and regions. In general, ORBIS representativeness in ECA countries is better compared to emerging economies in EAP and LAC (Gal, 2013) (Bajgar, Berlingieri, Calligaris, Criscuolo, & Timmis, 2020).[6] Furthermore, the ORBIS raw database suffers from several issues, including duplicates, omission, and measurement error. Several cleaning steps had to be implemented before the raw data in ORBIS could be processed and analyzed, as detailed in the following sections of the methodology. In addition, to complement and improve upon ORBIS's coverage of firm-level indicators, the Global BOS database leverages alternative data sources (EMIS, Factiva, Pitchbook, among others), World Bank proprietary data sets, country-level reports, and project information. For most countries, firm-level data collection has been also performed which allowed not only to produce a comprehensive data set of SOEs,[7] but also to complete the data initially available for each SOE in terms of employment, revenues, and other financial and firm level characteristics. World Bank country teams also reviewed and validated the country-level data to ensure completeness and accuracy of the information. The country-specific databases were also subject to technical exchanges with government counterparts.

The BOS database identifies an unprecedented number of firms with state participation across countries and economic activities, as well as providing novel insights on financial performance, economic performance, and governance of SOEs. A deep-dive analysis of 36 countries within the BOS database shows that 69% of SOEs operate in competitive activities (low efficiency-rationale for state participation), 16% are in partially contestable industries (moderate efficiency-rationale), and 15% are natural monopolies (strong efficiency-rationale). Furthermore, this analysis suggests that performance-based productivity of SOEs (revenues per worker) is negatively correlated with government control variables, such as government shareholding percentage and direct vs. indirect government ownership.

The remainder of this paper is organized as follows. Section 2 introduces the rationale for the analysis and describes the conceptual framework that defines and provides the structure of the Global BOS database. Section 3 presents the main data sources used. Section 4 describes the methodology to identify SOEs and create the Global BOS data set using ORBIS and additional data sources. Section 5 describes the structure

---

[6] Some empirical efforts to estimate models using firm-level data in ORBIS have documented the low number of observations in ORBIS in Australia, Chile, Iceland, Mexico, New Zealand, Israel, among others.

[7] A number of SOEs in the Global BOS database would not have been identified if some of these data sources had not been used.

and scope of the database. Section 6 presents a set of facts of SOE performance and sectoral participation, using the BOS database. Finally, section 7 concludes.

2. Conceptual Framework

This section introduces the three criteria that are the foundation of the conceptual framework of the Global BOS database:(i) a harmonized definition of SOEs, (ii) an identification of direct and indirect state ownership linkages at the national and subnational levels across the corporate sector, and (iii) a classification of economic activities depending on their efficiency rationale.[8] Together, these three criteria guarantee data comparability across countries and sectors.

A caveat of previous academic literature and policy-related workstreams is the lack of congruence with respect to the definition of SOEs across countries and economic activities. Often, the analysis of SOEs is connected to a local legal definition of companies that according to the national legal forms are defined as state-owned companies. However, this legal definition can vary substantially across countries.[9] This hinders the analyses of SOE performance within markets and distorts the aggregate economic implications of state participation. Academic attempts to identify SOEs in a systematic manner and assess their footprint usually define SOEs as firms which the state directly owns and in which governments have a majority share-holding percentage (Freund & Sidhu, 2017; Harrison et al, 2019; among others). Institutions that carry out economic and development policy work do not have a consensus on the definition of state-ownership. For example, the OECD, through the Product Market Regulation Indicator, defines SOEs according to the level of control of national or subnational governments, but also considers cases in which the government has special voting powers (i.e. nominating board of directors).[10] In 2020, the IMF Fiscal Monitor considered state-owned firms as those with state participation as low as 20 percent and also relied on other criteria to define SOEs, such as national legal forms (IMF, 2020).

To resolve this lack of harmonization, we propose an economic definition to define a firm as an SOE based on the government's control as well as their role in the market. That is, an entity is considered a state-owned enterprise for the purpose of the Global BOS database if it satisfies the following conditions:

I. It is controlled by government units or by other public corporations, proxied by a level of direct or indirect (i.e., through subsidiaries) participation of above 10%;[11]

II. It is recognized by law as a legal entity separate from its owners;[12]

III. It can generate profit or other financial gain for its owners;[13]

---

[8] This classification is described in further detail in Dall'Olio et. al. (2022).

[9] For instance, in Indonesia the state-owned enterprises are denoted as two different legal forms controlled by the governments: Badan Usaha Milik Negara (national) and locally owned (Badan Usaha Milik Daerah). For instance, SOEs are defined in Azerbaijani as public interest entities (PIEs), and in Mozambique as public enterprises and shareholding companies (World Bank, 2016).

[10] Even within different OECD workstreams, there is no harmonized definition. For example, an SOE survey conducted in 2015 denoted SOEs as corporate entities recognized by the national law in which only the central government exercised ownership and control (OECD, 2017), which differs from the criteria defined within the Product Market Regulation Indicator.

[11] The WB SOE policy tracker revealed that during the COVID-19 pandemic firms with as low as 10% could indeed receive significant support vis-à-vis their full privately-owned competitors. For instance, Lufthansa with 14% participation of the government received one of the largest programs in the aviation industry with a loan for over € 9 billion.

[12] In Poland, for instance, some municipal enterprises are registered as "municipal budget entities" (*samorządowy zakład budżetowy*). These entities are legally separate from the local governments that control them, but they are not separate legal persons. In this case, the entities are not considered as SOEs following the definitions here.

[13] It refers to the ability of the company to generate revenues and profits itself, but it does not limit the analysis to those actually reporting profits.

IV.   It is set up for the purposes of engaging in market production (i.e., it provides goods and/or services in exchange of a price).[14]

The SOE definition builds on the International Monetary Fund's (IMF) Government Finance Statistics Manual 2014 definition of SOEs, by adding an objective and quantitative criterion to proxy government control.[15] Generally, to assess the degree of "control" by the government on a corporate entity would require a firm-by-firm analysis. While a participation of 50 percent or more is sufficient to grant the state control over a corporate entity, this is not a necessary requirement: control can be achieved through a much lower equity participation and is not even limited to equity. For example, in a number of countries, governments have golden rights with a minority participation with the power to outvote other shareholders and directly influence the decisions of a firm.

To capitalize on the availability of shareholding information, and acknowledging that control is beyond majority ownership, we are setting the threshold for state participation at 10 percent to proxy government control. A 10% threshold for state shareholding is proposed to capture companies "controlled" by the state. The 10% is proposed to allow to flexibly manage the risks of overestimating the number of SOE (in those cases in which the government remains a minority shareholder vis-à-vis large private sector ones) with those of underestimating it (in cases of public companies for which control can be exercised with a lower level of ownership). At the same time, a 10% threshold is low enough to facilitate the application of different robustness checks with respect to different levels of state participation (i.e. majority-owned, more than 50 percent; blocking minority, between 25 and 50 percent; and minority participation, between 10 and 25 percent), while at the same time is not as low as to a large number of companies in which the government might not have any level of control.[16] This expands the scope of previous work on SOE analysis to trace state presence in the corporate sector, which has often been limited to majority-owned enterprises.

In line with the existing literature and practice, the Global BOS database only includes entities that are legally independent. That is, an entity is not considered an SOE if it is a branch of another company that operates as a single legal entity or if it is a part of a public authority without any juridical status. The rationale for the definition is that a branch of another company without autonomous legal status would not have an independent balance sheet, hence would be "consolidated" by the company which it belongs to while a branch of a public authority would be consolidated within the public sector. It is worth to note, that under this definition, we exclude those non-legally independent entities such as parastatals or dependencies under Ministries that which could operate and potentially influence markets (e.g., fiber backbone networks managed by government ministries). The database also distinguishes between corporatized or non-corporatized SOE. An SOE is corporatized if it operates under the company law as any other private sector company, with the only difference being that the government is a direct or indirect

---

[14] It excludes non-profit institutions and government units (see paras 2.30-31 of the GFSM).

[15] As established by the IMF (2020), although there are different national legal definitions of SOEs, there are three common elements to identify an SOE: it is controlled by the government, it is a legally separate entity from its owners, and it engages in market production and commercial activities.

[16] The 10% threshold is proposed as a proxy of control. Since control cannot be measured ex ante and can be linked to other ways of influence (e.g., golden rights, ability to influence the nomination of the board members), the state participation is used as an indicative variable of control. The threshold of 10% participation was determined after several robustness checks such that it balanced the trade-off between omitting companies with possible ability to influence while excluding companies in which the participation obeys to portfolio or diversification strategies. When testing different levels of participation under 10%, the results indicated that most of the companies were related to mutual funds and international investors (e.g., Blackrock). On the opposite side, when exploring thresholds over 25%, the risk of omitting companies with government participation was higher. After several exercises, the 10% threshold was robust enough to minimize the inclusion of companies with participation under the approach of portfolio investments, while allowing to capture large companies below blocking minority (e.g., Lufthansa). This comprehensive approach allows to do robustness and sensitivity analysis such that if the economic assessment of the policy agenda requires only the subset of companies with minority participation (10%-24%), blocking minority (25% or more), the BOS database offers a systematic variable to do so.

shareholder: the most common legal forms of corporatized SOEs are standard corporate forms such as Limited Liability, Joint Stock Company, among others. An SOE is non-corporatized if it is classified as a state company according to law, and this legal status is not a standard corporate form (i.e., no shareholder structure). A methodology was introduced to perform an analysis of the legal forms to be included in the final database (see Annex 2).

Finally, our database includes those firms that can generate profits and are engaged in market activities. The capacity to generate financial gains and profits implies that SOEs can participate in the market like any private actor that improves efficiency or adjusts prices to compete. This does not mean that only firms reporting profits are included in the Global BOS database, but rather that firms must have the capacity to generate revenues through the provision of goods and services. In this manner, a company is considered as an SOE if it provides a good or service which should be traded at economically significant prices. Based on this criterion, we exclude firms that offer goods or services which are often provided for free, such as firms in health or education services.

The second criteria of the BOS database is a novel algorithm tracing state participation in the corporate sector beyond directly owned firms. The ORBIS ownership information (known as the "links vintage files") only provides the connection for pairs of companies. In other words, the raw ORBIS files do not directly connect firms in such a way that for company C, we can identify company A as its indirect owner and connect all relationships as in A (y%) -> B (x%) -> C.[17] To address this issue and the lack of an SOE identifying variable in the ORBIS database, we developed an algorithm that retrieves the ownership structure of firms linked to the state using available information on shareholders and direct shareholding percentages.[18] Similar to a genealogic tree, the algorithm reconstructs the linkages at different ownership layers of all firms with state participation and their subsidiaries (and subsidiaries of those subsidiaries, etc.) applying at each stage the 10 percent threshold defined above.[19] The use of the algorithm allows for the identification of indirect shareholding relationships. The construction of the full ownership trees raises a set of methodological issues, including how to deal with loops or circular ownership links, what the maximum depth level for the tree should be, etc. The answer to these questions determines when the algorithm will stop the search for subsidiaries. These issues are addressed in section 4.2.

The output of the algorithm lays the foundation for understanding governments' corporate structures through ownership relationships at the national and municipal levels. This algorithm delivers the full extent of the linkages between a company and the state, which is key to understanding the real extent of fiscal risks and liabilities posed by SOEs and their subsidiaries, designing corporate governance forms, formulating transparency and accountability requirements, unveiling the risks of crowding out private investment, and addressing plausible distortions to competition (for example in the case of SOEs that are vertically integrated in upstream/downstream sectors). The algorithm captures all interrelated companies even if those are operating across borders. Figure 1 shows part of Pakistan's SOE 'ownership tree' constructed with the algorithm. The tree depicts a set of firms that are owned by the Federal Government of Pakistan (node labeled as "Pakistan") and a set of firms linked to local governments (nodes labeled as "Government of Khyber Pakhtunkhwa" and "Punjab Provincial Government").[20]

---

[17] There are cases in the ORBIS 'links' file, where the direct shareholding percentage is not reported, but a total shareholding is provided. For the latter, however, ORBIS does not report the full ownership path for obtaining the total share and therefore it cannot be used for the purpose of our methodology.

[18] A more detailed description of how the algorithm works is included in section 4.

[19] This involved the processing of more than 1.9 billion observations of ownership links and several terabytes of processing capacity to retrieve the trees.

[20] The algorithm allows to identify the unique code of the companies operating across borders and the country of location. For instance, the ownership tree reveals that SONANGOL, a large oil and conglomerate group has more than 200 subsidiaries operating in countries such as Brazil,

Figure 1. Ownership Tree: an example from Pakistan
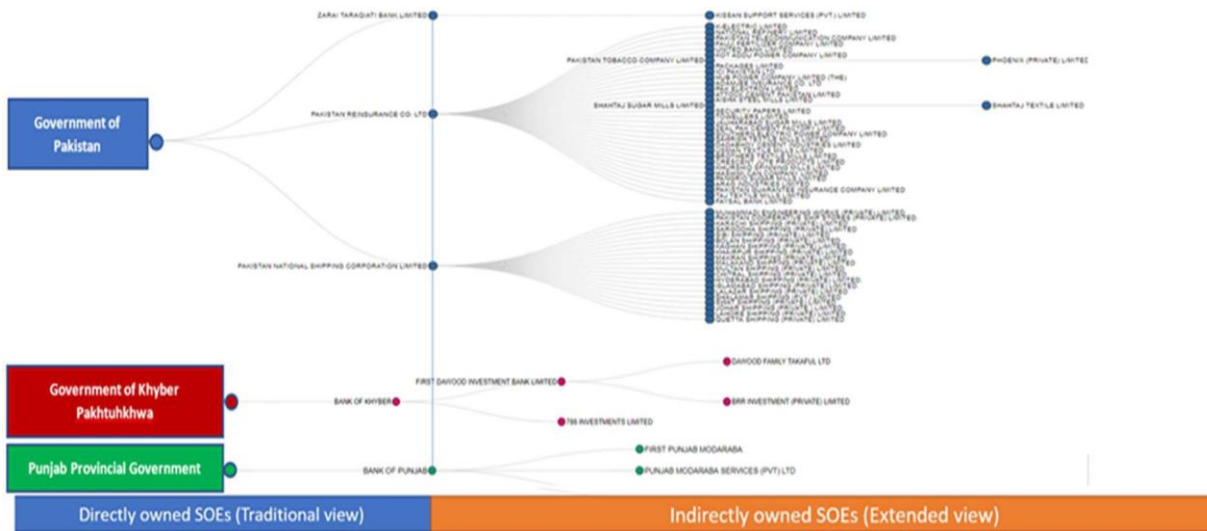
Note: The tree is read from left to right as follows: The left units refer to the parent entities that correspond to public authorities or government agencies of a specific country. The next layer (first layer) captures the **directly owned firms**, who can also be shareholders (of 10% or more) in the subsequent set of firms (2nd layer), and so forth. In the case of Pakistan tree, we observe at least 3 layers of ownership linked to the government.

Source: Authors' elaboration based on the Global BOS database.

The third criterion of the Global BOS database is a disaggregate classification of sectors according to their inherent technological features and market failures. This classification is developed by Dall'Olio et. al. (2022) to provide a guide for SOE interventions across disaggregate economic activities, which are close proxies for product/service markets. In particular, this taxonomy leverages the NACE Revision 2 Industry Classification to categorize 563 disaggregate economic activities (4-digit industries) into three categories: natural monopolies, partially contestable, and competitive sectors.[21] The categorization of a sector into one of the three categories is based on the market failures and the inherent technological features that characterizes the sector, i.e., an efficiency-based rationale. The three main categories of the taxonomy are the following, as explained in Dall'Olio et. al. (2022):

- **Natural Monopoly Sectors**: the economic literature identifies sectors in which it is not economically viable for more than one operator to provide the good/service. Typical examples are network industries (i.e., electricity transmission) characterized by sub-additivity in the cost structure which generates economies of scale. In other words, when provision by a single market player is the most efficient alternative, allocative efficiency cannot be achieved through profit maximization. This is the reason why the government might want to control the market power of the monopolist either through regulation or direct provision through SOEs.

- **Partially Contestable Sectors**: several sectors are characterized by some forms of market failures which could potentially be corrected through government ownership. Based on a

---

the US, and Spain. However, the global BOS database does not include the full set of variables such as financial module, ownership module for those companies given the limitations to validate that information. All companies in the global BOS database are operating domestically and fully validated under the methodological approach that is described in this paper. The cross-cutting indicators are also prepared only for domestic SOEs to be compared against domestic employment and GDP.

[21] The Statistical Classification of Economic Activities in the European Community, referred to as the NACE classification, is the industry standard classification system used in the European Union to classify economic activities.

comprehensive literature review, we identify three typologies of market failures which could potentially require corrective actions through state ownership: i) market power generated by structural barriers to competition, ii) under provision in the presence of positive externalities or uncertainty, and iii) risks connected to large/irreversible negative externalities.

- **Competitive Sectors**: these are sectors in which it is economically viable for multiple firms to compete to provide a good or service. Inherent market features, such as cost structure or demand characteristics, make entry into these sectors largely unproblematic. Furthermore, firms in commercial sectors are typically engaged in the provision of goods or services the consumption of which is either rivalrous or excludable. Given the competitive nature of these markets and private sector firms' ability to achieve economic efficiency without encountering significant market distortions, there is no strong economic rationale for SOE participation in them.

The BOS database excludes firms operating in sectors considered to provide public goods, which in some cases cannot be priced. In total, the data set excludes all 4-digit disaggregated sectors (classes) falling in the NACE Rev. 2 divisions of Public Administration & Defense, Education, Human Health & Social Work, and Activities of Extraterritorial Organizations; 2 disaggregated sectors in Finance & Insurance (pension funding and central banking), 8 disaggregated sectors within Arts, Entertainment & Recreation, and 6 within Other Services; a total of 52 out of 615. The exclusion is consistent with the harmonized definition of SOEs as entities that provide goods or services which can be traded at economically significant prices. Please refer to Dall'Olio et. al. (2022) for a more detailed discussion of the sector taxonomy.

3. Data

3.1 Bureau van Dijk (BvD) – ORBIS

Bureau van Dijk (BvD) provides two ways to access ORBIS data: the BvD-ORBIS online interface (Orbis.bvdinfo.com) and the BvD historic vintage files. The interface is the most straightforward way to access the data for each firm. By paying a subscription fee, a user can look up any company and if it exists, the interface will list all the information relevant to the search. For example, detailed information such as the list of shareholders and subsidiaries, sector of activity, balance sheets, and income statements among others can be viewed or downloaded online through ORBIS.

The platform is not designed for large volumes of data extraction because it puts a cap of 10,000 fields per export request. As a result, for data download requirements that exceed the cap limit, it is necessary to perform multiple downloads. Given the scope of the Global BOS database, the online download is also not a scalable option because the definition of SOEs requires both direct and indirect shareholding information and the interface lists only direct shareholders. To get the list of shareholders of the shareholders of a firm, the only option is to check each one of the first level shareholders and recover information manually.

Due to the challenges mentioned above, the best option for the construction of a large database such as the Global BOS database is to request access to the BvD historic vintage files, a copy of the entire raw

disks[22] that ORBIS can make available for downloads through File Transfer Protocol (FTP). ORBIS provides the option to access their historical vintage files, which are structured in different modules and years.[23] A set of technological specifications are required to process the files and build a single database. Raw vintage files can either be downloaded on the premises through a server such as Microsoft SQL server or Postgres, or hosted in a cloud-based data warehouse. Each of these methods has its costs and benefits. The local server option requires SQL coding and database management and administration skills to create, read, update, and delete tables. Because we are dealing with data in the order of 200Gb when compressed, the server must be large enough and needs to be expandable for additional updates. The data for the Global BOS database was obtained through a local Microsoft SQL server.

The main tables in the ORBIS raw "vintage disks" include:

- **Entities:** this file contains the list of all entities found in ORBIS. Each entity is identified with a unique ID called BvD ID number. The table also contains information on the entity type. The entity type categories include bank, financial company, insurance company, hedge fund, marine vessels, private equity firms, public authorities, managers, individuals, among others. This file also provides information on the country of operation.

- **Links:** this file provides the ownership information. More precisely, the linkages files (one per year) define the relationship between a pair of entities by providing the shareholding percentage.[24] Each pair of shareholder and subsidiary has its own row and there is one Links table for each year from 2007 to 2020. The Links table for 2019 has about 1.9 billion ownership links.[25] Additional cleaning to remove duplicates as well as to identify the most up-to-date information for each pair of entities was implemented.

- **Financials:** the financial module contains both the consolidated and unconsolidated balance sheets, income statements, and employment data. The information is presented in a long format where a row corresponds to the data for one entity for one year. Examples of variables in this table include the value of assets, debt, revenues, profits, stock turnover, and an estimation of some financial data for the upcoming year, among others. However, working with the financial module from the vintage files presented several limitations as documented by (Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, & Yesiltas, 2015). To overcome this and to work with the most up-to-date financial information, the global database used the financial module through the interface access.

Additional tables containing sector, industry classification, and contact information are also transferred with the main tables. The Links and Entities tables form the Ownership module of the ORBIS data, while the financial module consists of the financial data and the information in the additional tables. For the construction of the Global BOS database, we combine information from the vintage files as well as the online interface. For the identification of SOEs and the creation of ownership structures, we use ORBIS's

---

[22] The following files are available in the raw disks: Entities, Links per year, Industry classifications, Identifiers, Contact info, Additional company information, Legal information, Controlling shareholders, Basic shareholders, Headquarters, Overviews, Cash flow, BvD ID changes, Banks' financials, Insurances' financials, Industry's financials, and Key financials.

[23] ORBIS is not a single database but rather a collection of separate modules and partitions of data that required further processing to be able to construct the BOS database. For instance, the ownership links are provided for each year in a separate module from the annual financial indicators and from the entity type (atemporal file).

[24] The information includes the ownership participation of any company A in a company B, on the date of the link.

[25] This is the total number of connections identified in the form entity A has participation (x%) in company B previous to the cleaning developed by (Cusolito, 2020), which helped to remove historical ownership links that were no longer valid.

vintage files. Additionally, we retrieve the variables in Table 1 from the ORBIS interface to construct the four different modules of the Global BOS database.

Table 1 . Variables retrieved from ORBIS interface

Variable	Variable Name in Orbis	Description
Company ID	BvD ID number	Unique identifier for each entity provided by BvD for each company.
Name of the company	Company name Latin alphabet	Full name of the company (as registered)
Consolidation Code	Consolidation code	**Consolidated**, when the parent company reports financial statements with the results for the whole corporate group (parent and subsidiaries). **Unconsolidated,** when the subsidiaries and the parent company report in individual manner. The latter are the focus of the database, but we need to keep track of the consolidation code to avoid any double counting issues.
Year of incorporation	Date of incorporation	Year of start operations of the firm (incorporation date)
Legal Form	National legal form	National legal form. It varies by country.
Main Economic Activity (1-digit) and Description	NACE Rev. 2 Main Section (1 digit and description)	Sector classification of the main economic activity using NACE rev 2 classification - 1-digit and specific description.
Main Economic Activity (2-digit)	NACE Rev. 2, Core code (2 digits and description)	Sector classification of main economic activity using NACE rev 2 classification – 2-digit and specific description.
Main Economic Activity (4-digit)	NACE Rev. 2, core code (4 digits and description)	Sector classification of main economic activity using NACE rev 2 classification – 4-digit and specific description.
Country	Country	Country full name
Country ISO code	Country ISO code	ISO 2-digits code for the country
Employment	Number of employees	Total number of workers in the company (temporal and permanent)
Operating Revenues (Turnover)	Operating revenue (Turnover)	Operating revenues
Net profits/losses	Profit/(loss) after tax	After tax profit/losses

Source: Authors' elaboration

3.2 ORBIS data limitations and additional data sources

As (Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, & Yesiltas, 2015) pointed out, financial information in ORBIS for most European countries is superior to that of other countries because firms are required by law to file their financial information in Europe. In contrast, detailed financial data on SOEs are limited in the Middle East, North Africa, and Central Asia region due to weak oversight of SOEs and/or to a lack of direct access to financial statements (IMF, 2021). When information is not available in ORBIS, the Global BOS database is complemented with data from alternative sources, including official data from the Ministry of Finance, Treasury, business registries, tax authorities, local stock exchange (e.g., listed SOEs), open data sources, with a priority given to official government information in cases where a discrepancy is found (see Table A.2 in the Annex).

Key alternative data sources for the preparation of the Global BOS database include EMIS intelligence, which covers more than 5.5 million listed and unlisted firms across more than 125 emerging markets. Through EMIS intelligence, it is possible to identify a company's main sector of operation, as well as its financial statements at the unconsolidated level, audit status, and year of incorporation, among other

variables. EMIS collects information from official registries and government sources (e.g., Dion Global Solutions Limited in India).[26] The financial and ownership information in the global BOS database is also cross-checked against other publicly available sources that collect information in real time for listed and large firms around the globe such as Factiva, World Scope, and Global 2000 (Forbes).

The Global BOS database also benefited substantially from existing knowledge within the World Bank and other research and development institutions. To facilitate the analysis and to improve coverage of key financial, ownership, and employment variables, we constructed country-level data sets (registries). Through each registry, we were able to identify additional firms and ownership structures that were not present in ORBIS. The registries draw on data on SOES collected by the World Bank in coordination with government counterparts in the context of 150 operational projects and 20 analytical support projects from 2015 to 2019. Some examples include the information obtained for the preparation of the Integrated State-Owned Enterprise framework (iSOEF) reports for countries such as Angola, Niger, and Chad as well as country-level ASAs like the Pakistan Advisory Support to Public Expenditure Management. In addition, the Global BOS database incorporates publicly available information from other multilateral institutions or international organizations. Reports and databases constructed by other institutions such as the OECD (2017), the IMF (2021), the EBRD (2020), and the Inter-American Development Bank (2019) were also used.

Finally, the construction of the BOS database relied on field work carried out by World Bank country teams in consultation with country and sectoral experts. Country teams provided important reports and databases collected under ongoing client dialogue (i.e., PER in Mozambique), reviewed the information, and provided important insights to ensure the accuracy of the information. Teams provided expert knowledge on:

- Corporate legal structures/forms

- Business registry and statistical institute databases

- Government participation in key enabling sectors

The information from World Bank projects, external sources beyond ORBIS, and those provided by country teams greatly improves upon the coverage of any existing BOS database. (The rest of the paper will refer to information used to complement ORBIS data as the *'SOE Supplementary Database.'*) With the supplementary database, we were able to (i) identify entities in ORBIS that have links to the state even though the ORBIS ownership files do not identify that link, (ii) substantially expand coverage of financial variables, and (iii) add variables included in the Governance module of the BOS Global Database, such as legal entity type, administrative ministry line, government participation, among others. As shown in Figure 3 using only ORBIS leads to an underestimation of the state footprint across most countries. This is due to at least two issues: (i) *measurement error* (misclassification of firms with state participation as private firms such as the case of the state airline in Angola), and (ii) *omission error* given that some firms are not captured in ORBIS. In particular, the multi-phased methodological approach and supplementary databases allowed to compare the results from the GUO variable in ORBIS against the country counterfactual to evidence that there is a substantial number of firms incorrectly classified as privately owned (Figure 2).

---

[26] They also collect information from global providers such as MarketLine, Euromonitor International Ltd., Oxford Economics, SBI Securities, etc.

Figure 2. Number of SOEs in SOE that are misclassified as private in ORBIS in selected countries
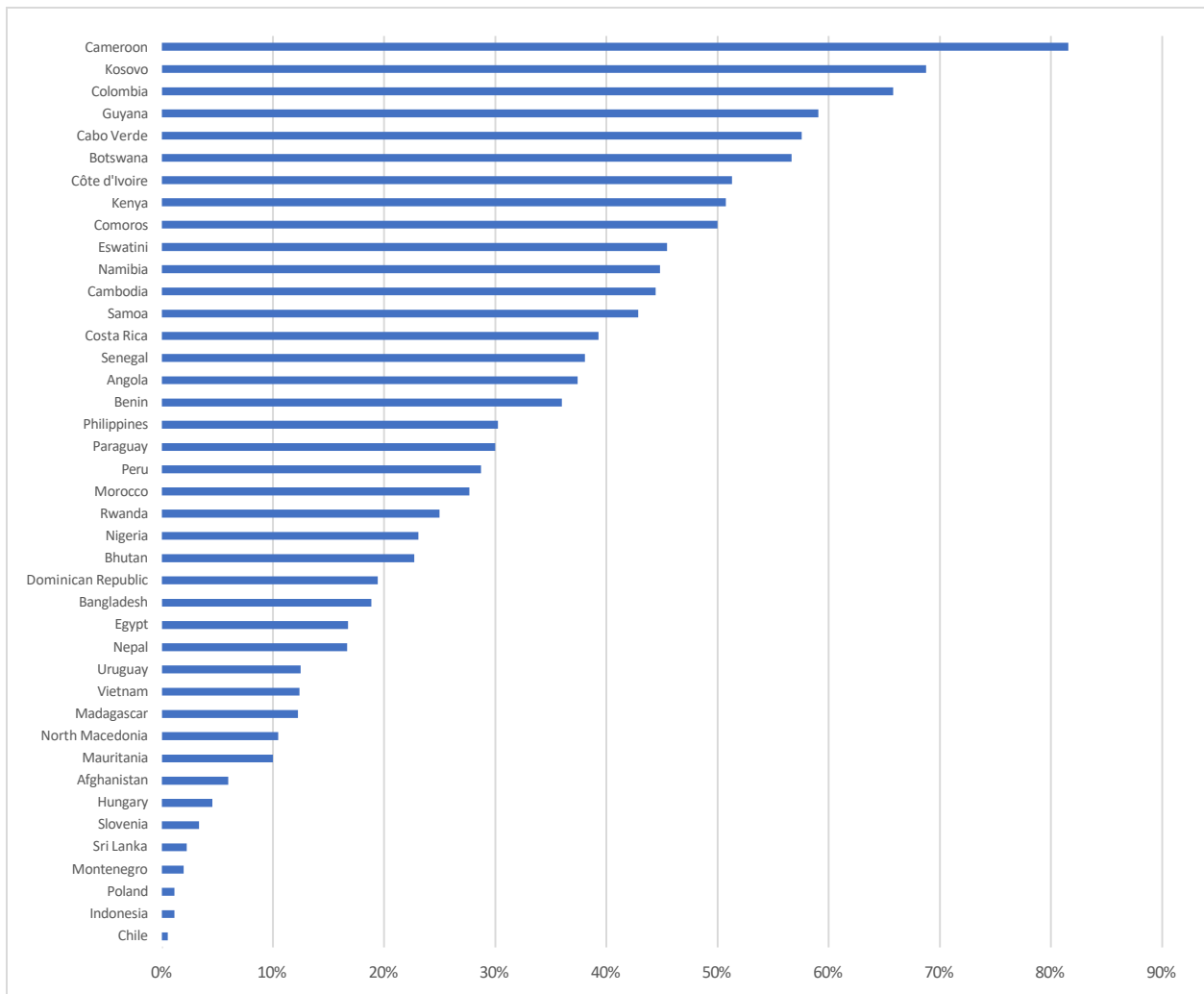
Source: Authors' elaboration based on ORBIS data and BOS database

We find that omission error varies significantly across regions and countries. For instance, in Bosnia and Pakistan, at least 40% of the final SOE registry corresponded to firms that were classified as private firms in ORBIS. Supplementary data sources to ORBIS also allowed to reduce the omission error from firms that were not covered in ORBIS even in ECA countries where ORBIS tends to have a better coverage according to the literature (Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, & Yesiltas, 2015).

Figure 3. Number of state-owned firms identified by country and source of information
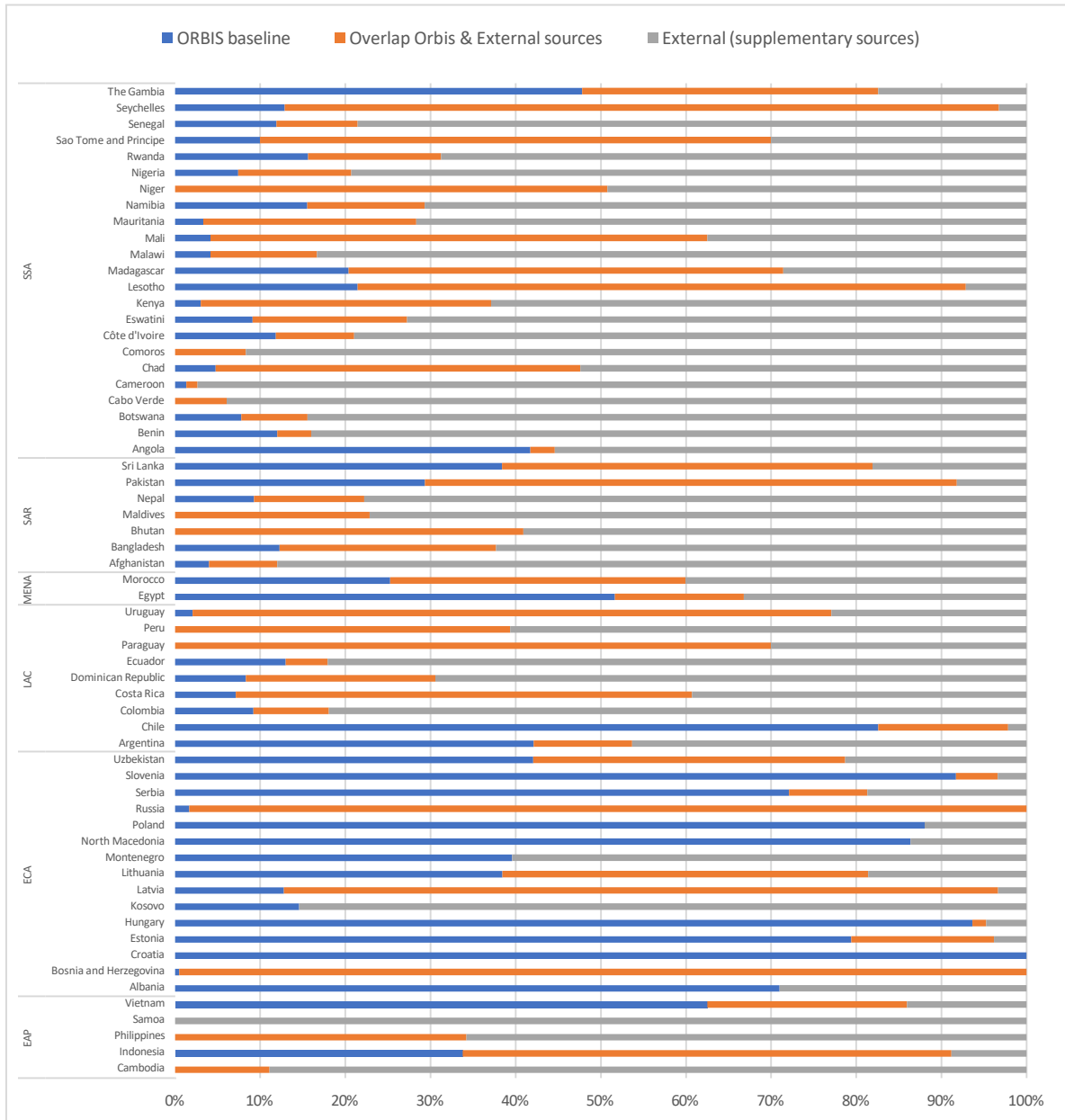
Note: Baseline information refers to the ORBIS first baseline using the algorithm developed without complementing any databases, overlap firms mean firms identified in ORBIS and external sources, and External Supplementary sources refer to the firms that were identified only through external sources.

Source: Authors' elaboration based on ORBIS and Global BOS database.

On average, supplementary databases increased the number of firms identified with state participation by 47%, although with variation across regions. ORBIS is a limited source to provide a full landscape of the state footprint in the markets. Using supplementary databases (See Annex 3) beyond ORBIS as well as implementing a protocol to validate the information with government counterparts was key to overcome these limitations. For example, in the SSA region more than 60% of the firms with state participation were identified through complementary and official sources (Figure 3). In Moldova, Montenegro, and Serbia,

the SOE landscape was expanded by more than 18% by including information from official sources (e.g., business registries, tax authorities). This evidence confirms the importance and value added of a multi-phased approach leveraging complementary databases and the country-specific validation, which minimizes the omission error and offers a more comprehensive landscape of SOEs.[27]

Our methodological approach also improves ORBIS' original coverage of financial and employment information to get a more accurate overview of the real footprint of the state in the markets. Across the current countries included in the Global BOS database, the use of complementary sources enlarged not only the total number of firms with state participation by more than 4 times, but it also improved the coverage of revenues and employment by more than 4.4 and 3.7 times, respectively (Figure 4).[28] Hence, these examples showcase the importance of combining the different sources of information to not only identify the full set of firms with state participation, but also to obtain the highest coverage with respect to financials and economic performance variables beyond ORBIS. Some country examples are presented in Annex 3. Nonetheless, it is important to note that some data gaps remain in the global BOS database despite the multiple efforts and sources implemented to complement the information including dialogue with some government counterparts. This lack of available information signals opportunities to improve transparency and underscores the importance for improving monitoring and accountability mechanisms.[29]

Figure 4. Value added of the approach and complementarities through the multi-phased approach
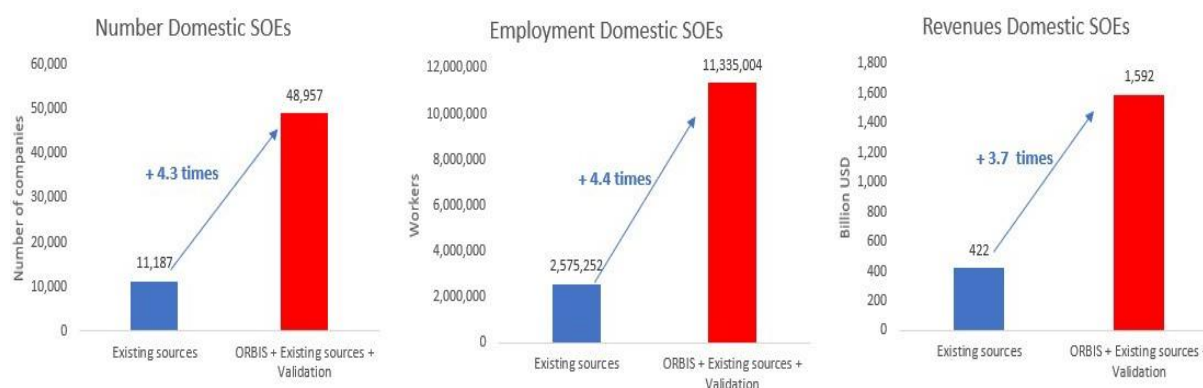
Source: Authors' elaboration based on global BOS database.

4. Methodology

This section identifies all the necessary steps to construct the Global BOS database. To introduce the novelty in our methodology, it starts by explaining why we could not simply use the GUO information provided by ORBIS to identify the SOEs. Then, it provides the algorithm that we follow from start to finish to extract, clean, and combine the ownership and financial information that make up the database.

---

[27] Some specific country examples of the complementarity across sources are added in the annex.

[28] For example, the baseline scenario for Pakistan suggested only 383,000 employees reported as the total number of workers in SOEs. By including the information from supplementary databases, as well as carrying out the country validation, employment increased by 36% to 523,000 workers (Figure A.1 in the Annex).

[29] Some countries with coverage under 50 percent of the total companies identified with state participation include Bolivia in LAC; Tunisia, Morocco, Jordan and the Arab Republic of Egypt in MENA; Maldives and Sri Lanka in the SAR region; Benin, Eswatini and Rwanda in the SSA region; Turkey in ECA; and Indonesia in the EAP region.

As well presented in (Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, & Yesiltas, 2015), ORBIS is a proprietary and entity-level global database constructed by the Bureau van Dijk (BvD) with the information collected from national business registries and statistical institutes (see Table A.1 in the annex for the full list of sources used by ORBIS). The simplest way to identify SOEs is through the Global Ultimate Owner (GUO25 or GUO50) variable available in ORBIS's "Links" table. The GUO25 and GUO50 variables indicate the unique majority shareholder of an entity at a 25% or more participation level and at a 50% or more participation level, respectively. Each entity in the ORBIS database has at most one GUO25 and/or GUO50.[30] An entity may not have a GUO if no shareholder owns shares at or above the threshold level.

However, proxying SOEs as the firms whose Global Ultimate Owner (GUO) is a public authority is an unreliable source of information. The GUO variable allows to identify firms whose unique majority shareholder is a state or a government. However, this method of identification of SOEs is limited for the following reasons:

- **GUO25 only captures the highest shareholder at or above the 25 threshold level of participation**. GOU25 variables do not capture all the shareholders that satisfy the threshold requirements. For example, if a state or a government owns 25% of entity A, and a private firm also owns A at 26%, assuming all other entities have less than 25% of A's equity, only firm B will be listed as firm A's GUO25. As a result, firm A will not be included in the list of SOEs when the GUO25 is used as an SOE identifying variable even though it has 25% participation of the state.[31] In other words, the variable GUO25 misses all the firms owned by a state or a government at 25% or more when other shareholders have higher shares.

- **GOU25 and GOU50 suffer from the lack of completeness of ORBIS data, which results in a large share of firms being missed.** The GOU variable suffers from omission and measurement errors resulting from the completeness and accuracy of ORBIS data. In the former, it refers to companies that are operating in the markets but are not covered by ORBIS, whereas the measurement error is linked to firms that are incorrectly classified as private.

To fully capture the comprehensive list of directly and indirectly owned SOEs (i.e., SOE subsidiaries) we developed and implemented a new methodology which uses the information collected by ORBIS, complements it with other sources and elaborates it to fully capture the ownership of the state in the corporate sector. The steps for the identification of SOEs detailed in this methodology yield a more accurate count of SOEs globally. A detailed description of the functions of the algorithm is presented in section 5.

Even adopting a 25% threshold for state participation, our analysis shows that the GUO variable tends to significantly underestimate the real presence of the state as a shareholder in most of the countries analyzed. When comparing the number of SOEs directly owned by a government at 25% or above with those identified through the GUO25 variable in ORBIS for the same set of 60 countries, we find that on average, the Global BOS database quadruples the number of firms with state participation per country (see

---

[30] Depending on the threshold (25 or 50) used in the GUO variable in ORBIS, it can lead to different "ultimate" owner entities. On the contrary, our methodology captures the full ownership structure of ownership using a lower threshold that allows to identify all paths with state participation to identify the ultimate shareholder accordingly.

[31] This example considers both direct and indirect total shareholding by either a state entity or private firm.

Figure 5), quadruples the number of workers (from 2.5 million up to 11 million) and offers a threefold increase in the operating revenues (from USD 422 million to USD 1.5 billion). For some cases, such as Uganda, Afghanistan, and Maldives, BOS database identifies over 50 times more companies with state participation.[32] The assessment of the omission and measurement error in ORBIS and implications for the creation of the BOS database are described in section 3.2.

Figure 5. Number of SOEs identified in the Global BOS database compared to baseline identified through the GUO25 variable (GUO25 = 1)
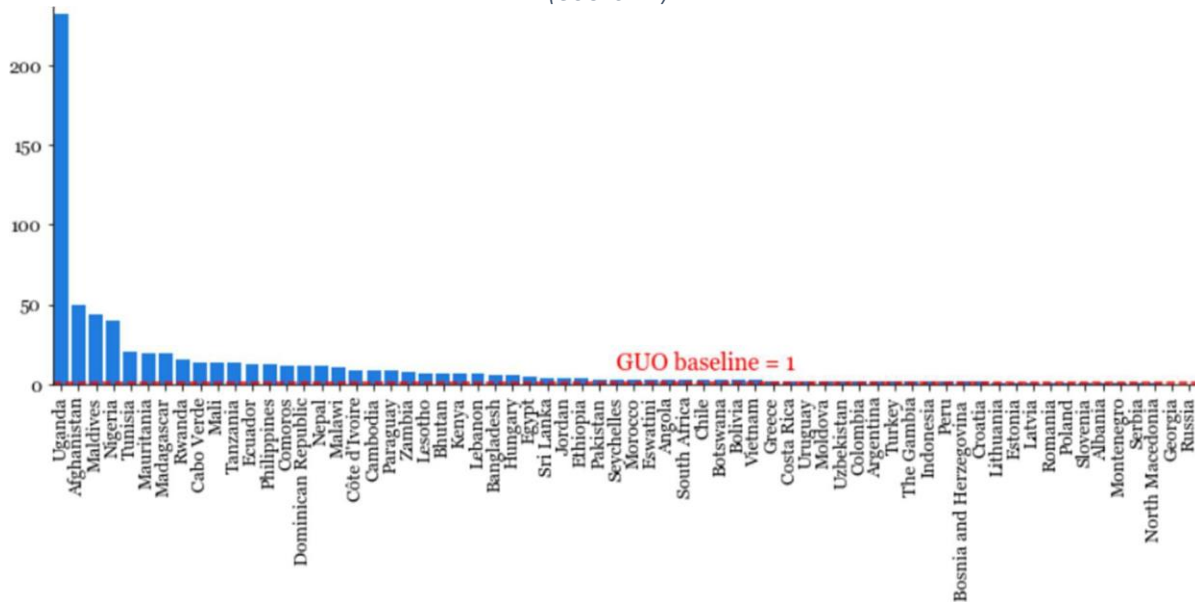
Note: Countries whose number of final SOEs appear to be below GUO25 benchmark respond to measurement error such that ORBIS classifies according to GUO25 some entities as corporations whereas those are public authorities (e.g., local municipalities in Russia). The issues with the measurement errors found in ORBIS are discussed in section 3.2

Source: Authors' elaboration based on ORBIS and Global BOS database.

4.1 The Global Businesses of the State (BOS) database methodology

The development of the BOS database required a multi-phased approach to ensure a comprehensive coverage and to ensure the accuracy of the information. A sequence of methodological steps was implemented to build a multi-modal and comprehensive database of the firms with state participation across the globe. The steps required for the database are: (i) the identification of the firms with state participation based on the harmonized definition and the implementation of a novel algorithm, (ii) integrate financial and governance indicators using, (iii) integrate the ownership information, (iv) the quality control including an extensive cleaning and cross-validation of the information as well as country team validation (Figure 6).

---

[32] This only compares companies with 25% or more participation and within the same sectors as defined by our methodology to ensure consistency.

*Figure 6. Overview of the methodological steps to develop the BOS database*

1. **Build the registry of firms with state participation**

2. **Integrate financial information and governance indicators**

3. **Integrate the ownership information identified through the creation of the ownership trees (degree of relationships and state participation) as variables.**

4. **Quality control including additional cleaning and country team validation**

Source: Authors' elaboration

4.1.1 Identify State-Owned Enterprises

The first step in the methodology is to apply the novel algorithm described in the conceptual framework to the 1.9 billion ownership links (or pair of shareholding relationships) in ORBIS to retrieve SOE ownership structures. The entire tree of SOE ownership was created using the cleaned Links and Entities tables from the ORBIS *Ownership Module* as of 2019. The algorithm identifies all firms directly or indirectly owned by a public entity or government of 10 percent participation or more. The steps to construct the ownership trees are the following (Decision Tree in Figure 7).

Figure 7. Decision tree implemented in the algorithm (Top-to-bottom approach)
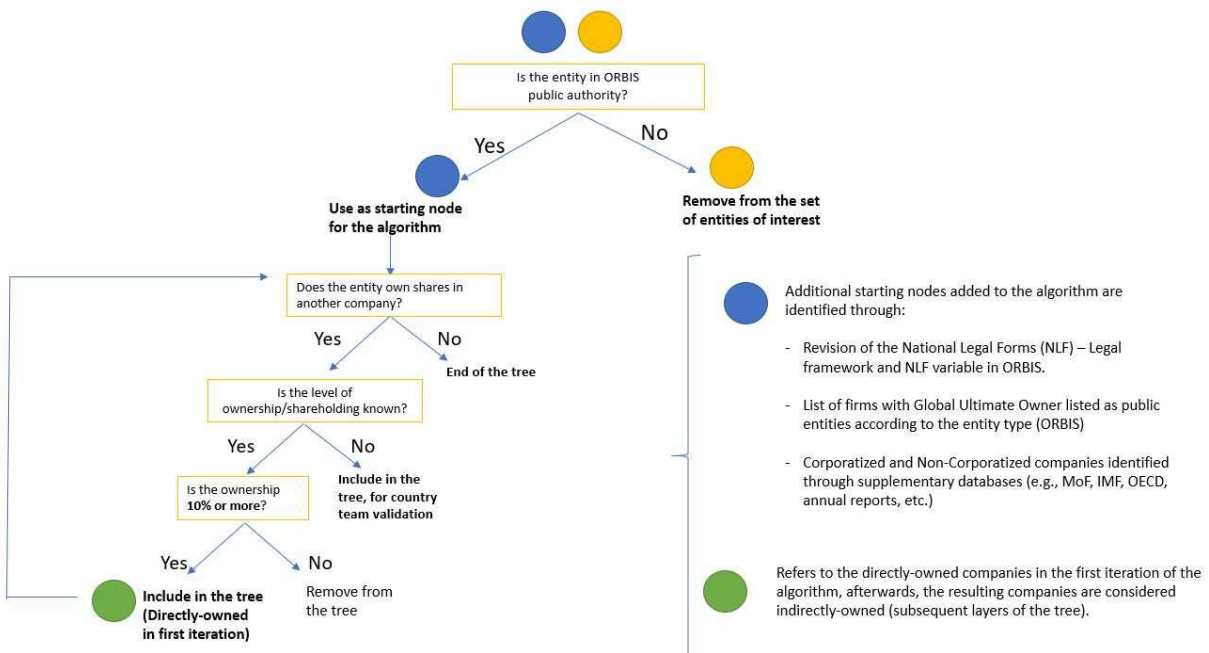
The main idea of the algorithm is to produce for each country the full ownership tree (i.e. direct and indirect subsidiaries) of: (i) "Public Authorities"; (ii) entities identified as SOEs in the National Legal Form; (iii) entities identified as SOEs from the Supplemental database; and (iv) entities identified as SOEs in the GUOs. In each step, the algorithm ensures that firms are classified as either corporatized or non-corporatized based on the national legal form, and that branches and inactive entities are removed. The detailed algorithm runs as follows:

1. Use the clean Links file[33] to identify the set of entities that are labeled as "Public Authority" under the variable "Entity Type" and their respective BvD ID (company ID). ORBIS classifies governments and government institutions, such as ministries or municipalities, under this category.

2. Use the information on BvD IDs, shareholders, and direct percentage to construct the ownership tree for each entity labeled as "Public Authority".[34] These ownership trees are constructed using the iterative and recursive process described below:

   - For each public authority, identify its direct subsidiaries (entities and their BvD IDs that report the public authority's BvD ID as a direct shareholder).

   - Using the direct percentage, keep all direct subsidiaries for which the public authority's shareholding direct percentage is greater than or equal to 10%. If the shareholding percentage is unknown, and the subsidiaries are not branches or inactive firms, keep the direct subsidiary so as not eliminate potential SOEs.

   - Next, for these direct subsidiaries linked by a known shareholding percentage, identify the set of their subsidiaries. To do this, identify the entities (and their BvD IDs) that report a direct subsidiary's BvD ID as a shareholder, and did not already appear in the first (previous) level of subsidiaries.

   - Using the shareholding percentage again, keep all subsidiaries for which the direct subsidiary's shareholding percentage is greater than or equal to 10%. If the shareholding percentage is unknown, keep the subsidiary.

   - Continue this iterative process until there are no further subsidiaries.

3. Combine all the ownership trees of public authorities, in a given country, into one government ownership structure and identify the unique entities (and their BvD IDs) within this ownership structure.

4. Use the ORBIS interface to double check and eliminate all inactive entities or branches.[35] We define the remaining active set of entities as Corporatized SOEs.

---

[33] (Cusolito, 2020) developed the script to remove outdated information in the links file and to obtain the file denoted as "clean" links to run our algorithm.

[34] We omit from Sovereign Wealth Funds (SWF) and institutional investors (e.g., IFC) as either appearing as public authorities (starting nodes) or potential subsidiaries.

[35] ORBIS provides a status indicator to differentiate active from inactive entities. Similarly, ORBIS also indicates in the entity's national legal form whether the entity is a branch or not.

5. Identify the legal forms of non-corporatized SOEs in the country. To do this, we use the variable "National Legal Form" in the ORBIS interface and identify all entities (and their BvD IDs) that report the legal forms of non-corporatized SOEs for the variable "National Legal Form" and with the support of the country teams we verify whether those are fully aligned to our definition.[36]

6. Use the ORBIS interface to eliminate inactive entities or branches identified in step 5. We refer to the remaining active set of SOEs as Non-corporatized SOEs.

7. Compare the sets of Corporatized SOEs and Non-corporatized SOEs, to see if there is any overlap.

8. For each entity in the Non-corporatized SOE set and not in the Corporatized SOE set, construct the ownership tree, only if this tree exists.[37] Any extra entities (and their BvD IDs) from these additional trees are added to the set of Corporatized SOEs.

9. Search for the SOEs of the SOE Supplementary Database within the Corporatized and Non-corporatized sets of SOEs. We identify those entities that are in the SOE Supplementary Database and not in the Corporatized or Non-corporatized sets of SOEs and classify them into either the Corporatized set or the Non-corporatized set based on the legal form information within the SOE Supplementary Database.

10. Search for each entity identified in step 9 within the ORBIS interface to determine if this entity is present in ORBIS with its ownership information and whether it might have missing or incorrect ownership links that may have resulted in its omission from the Corporatized or Non-corporatized sets.

11. For those entities (and their BvD IDs) found in the ORBIS interface in step 10), we proceed to construct the ownership tree in the same way as in step 2, only if this tree exits. Any extra entities (and their BvD IDs) from these additional trees are added into the set of Corporatized SOEs.

12. Use the ORBIS interface to identify all the firms, for which the GUO is a "Public Authority" in the variable "Type of Entity". We cross check this set of firms within the Corporatized and Non-corporatized sets of SOEs. For those extra firms that are missing in these sets, we classify them into either the Corporatized set or the Non-corporatized set based on the legal form reported in the variable "National Legal Form" in ORBIS.

13. For the extra firms identified in step 12), construct the ownership tree in the same way as in step 2 only if this tree exists. Any extra subsidiaries (and their BvD IDs) from these additional trees are added into the set of Corporatized SOEs.

14. Use the information of sector from variable "NACE Rev. 2, core code (4 digits)" in ORBIS for all the entities identified in the ORBIS Database and the information of sector from the SOE Supplementary Database for only those entities identified through the SOE Supplementary Database to eliminate all entities that fall within one of the excluded sectors described in Section 2.

15. The remaining entities in the Corporatized and Non-corporatized sets make up the registry of SOEs in the country.

---

[36] The review of the legal framework in each country was conducted with the support of the WB country local teams.
[37] It means that there are firms reporting the No-corporatized SOE as a shareholder.

The algorithm above traces every link of state participation once the starting node is correctly determined as a public authority. However, we identified cases where ORBIS incorrectly classified a government-related entity as a corporation, and it was therefore not included in the algorithm (see decision tree in Figure 7). To correct this, we analyzed the national legal definition for each country to identify additional starting nodes that needed to be included in the exercise.[38] These additional starting nodes were included in the algorithm to map additional subsidiaries and firms that were not identified in the first run. Another tool that we employed to minimize the measurement error consisted of using the information from the GUO variable described in section 2 to identify potential additional entities as starting nodes. Inclusion of GUO variable also addresses cases where a firm is linked to a single public authority through multiple ownership "paths" and each individual path does not exceed the threshold, but in aggregate the public authority holds shares above the threshold.

The iterative use of the algorithm to identify SOEs in ORBIS, combined with additional information from the supplementary database yields an unprecedented coverage of SOEs. The application of the algorithm to ORBIS may fail in cases where the entity at the starting node is wrongly classified or not even included at all in ORBIS, or where some ownership link is not recorded. We used the supplementary database to complement the entities identified through the algorithm and then applied the algorithm to the new set of entities identified. This iterative process helps retrieve a universe of SOEs larger than previously identified in many countries.[39]

In sum, the SOE registry for a specific country is built through five complementary steps to minimize the measurement error. The full list of SOEs in a country (SOE registry) is built through a combination of (i) ownership trees run through the public authorities identified in the ORBIS 'entities file', (ii) ownership trees built on additional entities identified as government-owned firms through the review of the national legal form, (iii) ownership trees built on additional nodes identified through the global ultimate owner information, (iv) list of firms that satisfy the SOE definition in supplementary and official databases (e.g., MoF), and (v) the latter list of companies' ownership trees (if the firms are in ORBIS).

4.1.2   Incorporating Financial, Economic Performance, and Governance Variables

Once the registry of SOEs is completed, the second stage consists of complementing the data with financial and corporate governance information. For this purpose, we used the ORBIS interface and the SOE Supplementary Databases. It is important to note that ORBIS presents information on consolidated and unconsolidated statements of firms. As explained by the database manual from Bureau van Dijk (ORBIS, 2011), consolidated accounts are composed of financial information for the mother company and all its subsidiaries. Unconsolidated accounts correspond to financial information of just the specific company, excluding the financials of its subsidiaries. (Cusolito, 2021) shows that some countries report only consolidated accounts, others report only unconsolidated accounts, and others report both in ORBIS.

Given that the BOS database provides the full ownership structures, unveils all subsidiaries, and presents the information where each company is the unit of observation, we focus on collecting unconsolidated financial accounts of domestic SOEs to avoid any double counting issues. Since the BOS database includes all separate legal entities including those that can be subsidiaries of other companies in the database, we

---

[38] Close coordination with the regional and country-level experts across GPs and across EFI were key to identify these set of firms based on the national legal forms.
[39] Some cases, the algorithm could not find the specific company because it referred to non-corporatized forms, where firms did not report specific shareholding information. To minimize this measurement error, external and complementary databases, as mentioned in section 3, were essential to cross-check the findings of the ownership trees and complement as needed. New firms identified through external sources were also included as starting nodes in case those could provide further subsidiaries in ORBIS.

focus on collecting unconsolidated financial information to avoid double counting issues that may overestimate the footprint of the state. When only consolidated information is available, we document that in the consolidation variable and indicate in the respective subsidiaries to avoid any overestimations.

The steps to construct the BOS Global Database financial and governance modules are the following:

1) For each country, identify the set of unique company identifiers in the SOE registry (using their BvD ID numbers) and export the list of firms of interest to the ORBIS interface.

2) Retrieve the key unconsolidated variables defined in section 2 using the ORBIS interface as of 2019 for the firms of interest, to be consistent with the ownership links created as of 2019.

3) For firms identified through supplementary sources, allocate a unique identifier using a similar structure as ORBIS and use complementary databases (as EMIS, Pitchbook, and those described in section 3) to retrieve the variables of interest (e.g., sector, year of incorporation) and financial unconsolidated information as much as possible for employment, revenues, net after tax profits/losses.[40]

4) In case that a specific company is found both in ORBIS and in the SOE Supplementary Database, but where there are information discrepancies, use the official data collected through the SOE Supplementary Database.

5) Although the main focus of financial data is to obtain unconsolidated financial statements for all companies, when we can only find consolidated information despite the use of supplementary databases, the financial information is added to the database and indicated as C1/C2 (consolidated) in the consolidation code variable and mark as n.a. (not available) for the subsidiaries to indicate that those are covered under the financials provided by the parent company.

6) To differentiate between corporatized and non-corporatized SOEs, create a variable (Corporate type) which specifies whether a company is corporatized or not based on the analysis of the national legal forms and/or the company's incorporation act.

4.1.3   Merging the ownership module with Financial, Economic Performance, and Governance Variables

The final stage consists of merging the ownership information with financial, economic, and governance variables through the unique company identifier. For this purpose, we developed another algorithm to transpose all ownership links in the form of a matrix that combines different layers of state participation into a single row. In that way, the ownership links can be merged with the SOE registry using the unique firm identifier (BvD ID identifier). Figure 8 shows an example of the 'ownership tree flipping' algorithm's input with three levels of ownership. In our example, entity 1, which is a public authority, owns entity 2 at 74.8%, which in turn owns 10.26% of entity 3.

---

[40] The unique identifier of each company follows the same structure XX00000j, where XX corresponds to the ISO 2-digits code to denote the country of operation of the company, and j is a unique value provided by ORBIS to identify the firm.

Figure 8. Example of SOE ownership structure before the application of the "ownership tree flipping algorithm"

Firm_1 ⌄	Share_1 ⌄	Firm_2 ⌄	Share_ ⌄	Firm_3 ⌄
PH*J00S4480	74.80	PH0000023009	10.26	PHAS94009983

To construct the full ownership tree, we use the following "ownership tree flipping algorithm":

1) For each country, compile all ownership sub-trees. These files come in the format of a csv file for each public authority or government-related entity in a country.

2) Complement the ownership tree with supplementary information, particularly for those links reported as existing connections by ORBIS that did not have specific percentage of participation (when available).

3) Run the "ownership tree flipping algorithm" to transpose the final matrix containing the ownership connections by public entity into a table showing ownership connections by company, removing any potential duplicates, and maintaining multiple links where they exist *(flipping the tree)*. Companies owned by multiple state entities are flagged in the data module sheet of the excel file, where only the largest shareholder is displayed. The ownership module shows all the linkages in separate rows without adding up any state participation across entities (Figure 9). Further, a company can have state participation through one single ministry, but the ministry's ownership may be expressed through different sets of intermediate firms (ownership "paths"). Such cases are also selected, cleaned, and harmonized by the algorithm.

Once the full ownership tree is developed, the final step consists of merging the ownership information with the financial information collected previously. To avoid duplicates of financial and economic variables, each entity will have only one row in the final database. This means that to produce the final Global BOS database, when an entity has multiple linkages to the government, we select only the record with the highest direct participation and merge that row with its corresponding financial data. The other records of the same entity are kept in another file separate from the Global BOS database.

Figure 9. Example of SOE ownership structure after the application of the "ownership tree flipping algorithm"

Firm Identifier ⌄	Company Name ⌄	Participation by direct shareholder ⌄	First level shareholder ⌄	Participation by second level shareholder ⌄	Second level shareholder ⌄	⌄
PH0000023009	United Coconut Planters Bank - UCPB	74.8	PHILIPPINES			
PHAS94009983	UCPB-CIIF FINANCE AND DEVELOPMENT CORPORATION	10.26	United Coconut Planters Bank - UCPB	74.8	PHILIPPINES	

4.2   Ensuring the accuracy of the information: Quality control

In addition to standard cleaning and quality control procedures, we also implemented a set of cleaning rules to ensure the alignment of the data to the harmonized definition of SOEs. After these steps were implemented, the data was shared with the country teams and sector experts as an extra validation step before sharing the information with the government counterparts for final validation.

4.2.1   Cleaning Rules: Standard cleaning and quality control

The ownership and financial modules in ORBIS went through various cleaning steps before we were able to identify SOEs and extract their financial and economic data. Building on the lessons and literature using ORBIS information (Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, & Yesiltas, 2015), the following cleaning steps were applied:

- *Remove the duplicates from the ownership Links data.* Using the subsidiary id, shareholder id, and type of relation to identify each unique observation, we retain the most recent ownership information if duplicates exist.

- *Reconstruct the Year variable.* The variable Year takes the value of year for the closing date if the closing date falls on or after June 1st. Otherwise, we assign the previous year to that variable. The Year variable indicates the fiscal year for which the financial data were reported.

- *Remove inactive links.* For each year, we remove all obsolete links that were no longer valid by using ORBIS's variable "Active archived" that indicates whether a link is active or obsolete. This step is essential to ensure that the information on ownership is correctly specified every year.

- *Convert local nominal currency.* The balance sheets and income statement data are stored in local and nominal currency in the ORBIS files. For our purposes, we converted all nominal values to real US Dollars with 2005 as the base year.

The cleaned version of the Ownership module is the starting point for the creation of the global BOS database. After the ownership module is cleaned, the following cleaning rules were further applied:

- *Drop entities that are no longer active in the markets*: Using the "Active" status variable in the ORBIS interface, we delete all firms flagged as inactive because they correspond to firms that were dissolved, liquidated or under bankruptcy.

- *Complement data gaps when information as of 2019 was not available:* When there was no information available in terms of sector of operation, revenues, employment, or profits as of 2019, we employed data reported as of 2018 or 2017 as the best proxy. The financial information was deflated to report prices as of 2019 using the WBG GDP deflators. All financial indicators are reported in thousand USD.[41]

- *Review financial data referred to as unconsolidated information to avoid double counting (or overestimation of the state footprint):* Although ORBIS provided a variable to indicate whether a company reports as a consolidated unit (as parent with their subsidiaries) or unconsolidated (as individual), we did several cross-checks to ensure firms with information were reporting in unconsolidated terms.[42]

- *Review ownership information:* Cross-check with existing information to confirm the level of state participation as of 2019. When official sources or complementary information suggested firms do

---

[41] When data was provided in local currency, the exchange rate as of December of 2019 by the Central Bank was employed to convert the information to USD.
[42] When the company only reported consolidated information, it was used for the aggregate dashboard, but to avoid double counting issues, the information of the subsidiaries is not included. Those firms are flagged as n.a., to indicate their information is captured by the parent company in the database.

not have state participation or is below 10%, those firms and their subsidiaries are excluded from the. This extra step helped to update/remove linkages that were no longer valid as of 2019.

4.2.2   Cleaning Rules: Ensuring alignment with conceptual framework and SOE harmonized definition

- *Drop sectors not included in the database:* As per the global SOE definition described earlier, we exclude public administration, defense, social security (including pension funds), education, human health, social work activities, and activities of extraterritorial organizations sectors, among other activities.[43]

- *Drop entities that are not aligned with the SOE definition:* ORBIS includes ministries, public authorities, regulators, and Central Banks as entities. We carefully reviewed that those are not counted as SOEs.[44]

- *Keep domestic firms only:* Although the database allows to trace both domestic and foreign-based firms linked to a specific country-government, for the purpose of the subsequent data validation and for the creation of the country-level dashboards (e.g., revenues as % GDP), only domestic firms are kept in the database. Domestic firms are defined as those with the same country ISO code as the government.[45]

- *Drop entities that refer to branches as opposed to companies*: Using the ORBIS interface branch variable, we delete all branches that are not SOEs because they are not separate legal entities.

4.2.3   Country Team Validation: Data Validation and Integration

Even after combining the information from ORBIS and the SOE Supplementary databases, important data gaps remain. To address this issue, we engaged with country teams and country experts to help review the information collected and to retrieve as much of the missing information as possible. The steps carried out through the Country Team Validation are the following:

1) *Validate the list of SOEs identified by the methodology.* If there are firms that are not SOEs, either because they are incorrectly identified or have gone through a privatization or liquidation process, they are eliminated. Also, the country team provides information on any missing SOEs and any relevant variables.

2) *Validate that the information from the different modules of the database is correct.* Country teams help to complete the data that are missing across the different modules by suggesting alternative sources of information or reach out to relevant government counterparts to collect missing information (e.g., IGAPE in Angola, FONAFE in Peru). For example, in some instances, ORBIS reports more than one value of the shareholding percentage. In this case, the country team needs to determine which value is correct.

---

[43] Creative arts and entertainment activities related to libraries, museums, botanical gardens, and reserve activities (Code 90-91 in the NACE sector classification) are also excluded from the database.

[44] In the case that the regulator also has a dual role as a market player such as a port authority that also operates the port, then it is included as an SOE. It required a careful review of those entities and related activities by the SOE global team and country-level teams.

[45] For example, SONANGOL and its domestic subsidiaries are captured in the database as long as the unique identifiers of those firms (first 2 digits referring to the country of operation) coincide with the Government of Angola (AO). The foreign subsidiaries are identified and traced but only used for some overall indicators (e.g., presence across countries and type of sectors).

After the country team validation, including review by sector experts from other areas such as transportation and infrastructure when applicable, the database is shared with the relevant government counterparts for a final review and non-objection.

5. Scope of the Database

The Global BOS database is a cross-section database as of 2019 made up of four modules (i) economic characteristics, (ii) ownership relationships, (iii) economic and financial performance, and (iv) governance. The four modules of the database (Figure 10) provide a landscape to fully measure the footprint of the state using the most complete and updated information from different data sources described in the following section. The full description of the variables in the BOS database is in Annex 6.

Figure 10. *Modules and key variables in the global BOS database*
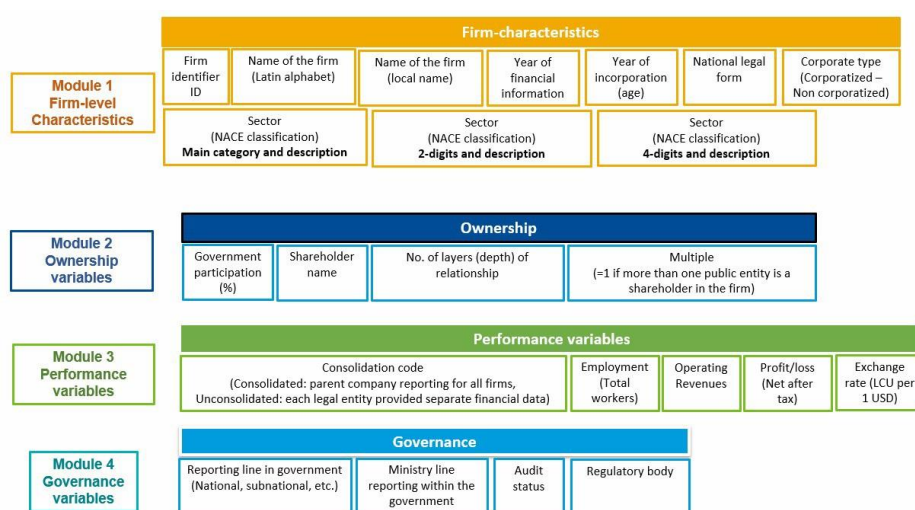
Source: Authors' elaboration

The first module provides information related to firm-level characteristics such as country of operation, economic activity, sector of operation, year of start of operations (proxy for the age of the firm), among others. Previous studies suggest that in many countries SOEs are present in a wide variety of sectors, but they are mostly predominant in natural monopolies such as transportation or utility sectors (IMF, 2021) or network industries (e.g., telecom) (OECD, 2017). However, for the first time, the Global BOS database adopts the NACE classification of economic sectors.[46] Other information included such as the country of operation, legal form, and age of the SOEs, enables cross-region and cross-country comparisons on SOE prevalence and differentiation by firm-level characteristics in different sectors of the economy.

The second module refers to government ownership and provides information on state participation (percentages) and who exercises the ownership functions on behalf of the government (which public authority the SOE is linked to). This module spells out all the ownership relationships (i.e., ownership trees) across different layers. For each SOE, the Global BOS database identifies the firm's shareholders and their shares at different levels. By design, the last shareholder is always a public authority or the government. That is, for a specific company, the database can provide as many layers of information as

---

[46] We are aware that some companies may operate in more than one activity or participate in activities that are bundled (i.e. electricity transmission and electricity distribution). To address this issue, the database collects information on the primary sector of operation of companies, as this sector is the most important in terms of companies' operations.

needed to trace its relationship back to the government. This module also provides a novel indicator denoted as multiple links to denote whether more than one public entity (e.g., different ministries) act as shareholders in the company on behalf of the government and therefore it can provide more than one ownership path that connects the firm to the government.[47] For indirectly owned firms, the number of layers can go from two (e.g., Bangladesh, Cabo Verde) to as deep as thirteen levels (e.g., the Russian Federation). Private firms with no ultimate links to public authorities or governments are not included in the data set.

The third module collects economic and financial performance variables focused on employment, operating revenues (turnover), and net profit/losses (after tax). Researchers use these variables to analyze SOE performance, profitability, and contribution to the economy. However, in many countries, data on SOE income statements and balance sheets are hard to obtain. The global BOS database provides information on total employment (permanent and temporal workers), operating revenues and net profit/loss after tax. The financial information is provided in unconsolidated terms (i.e., parent company reporting as single entity and subsidiaries reporting their respective operations) to avoid potential double counting issues.

The fourth module contains variables related to corporate governance, including the reporting line ministry, level of government, audit status, among others. This final module identifies which firms are owned by the central or subnational (e.g., municipal governments), the ministry line and oversight entity of the SOE, the sectoral regulator, and the audit status.

6. Application: New Facts on SOEs Based on the BOS Database

In this section, we present descriptive statistics of SOEs for a subsample of countries in the database. In particular, we characterize SOE patterns for employment, revenues, and state participation by region, type of contestability, type of ownership, age, and size. In the BOS database, employment is measured as the number of workers that are employed by the SOE. Operating revenues is measured as the income that is generated by the SOE from business operations. As explained in section 2.2.2, the BOS database provides information on the shareholding percentages between the government and their direct subsidiaries and between directly owned SOEs and their indirect subsidiaries (and between subsidiaries of those subsidiaries). Using this information, we classify firms into three groups of state participation: majority owned SOEs (shareholding percentages higher than 50%), SEOs with blocking minority participation (shareholding percentages between 25% and 50%) and SOEs with minority participation (shareholding percentages between 10% and 25%). A firm is considered an SOE with majority government participation if the shareholding percentage is above 50% for all ownership layers that link it to the government. If at any ownership layer, the shareholding percentage between two firms is less than 50%, then all the firms beyond that layer are not considered as SOEs that have a majority government participation.[48] In the same manner, a firm is an SOE with blocking minority participation if the shareholding percentage is above 25% for all ownership layers that link it to the government, but less than 50% for at least one of the ownership layers that link it to the government. Last, a firm is an SOE with minority state participation if the shareholding percentage is above 10% for all ownership layers that link

---

[47] For instance, in Angola, the Banco de Poupanca e Credito is owned 75% directly by the Government of Angola, and 15% is owned indirectly by the national social security. The ownership module spells out these different relationships to understand the different paths that connect the firm with the government as a shareholder.

[48] For example, consider a group of SOEs (firms A, B, and C) that are linked to a government public authority (PA) in the following way: PA (X%) ->A (Y%) -> B (Z%) -> C. For companies A, B, and C to be considered all majority owned, then X%, Y%, and Z% should be all greater than 50%. Suppose that X% was greater than 50%, while Y% was less than 50%, then company A would still be considered an SOE with majority participation while company B would not. In addition, company C, regardless of the value of Z%, would not be considered an SOE with majority participation, given that its direct owner does not satisfy the criteria of a shareholding percentage above Y%.

it to the government, but less than 25% for at least one of the ownership layers that link it to the government.

Currently, the BOS database includes 53,675 state-owned enterprises, as defined by the definition in section 2.2.1, across 80 countries. However, the coverage is varied across countries, even after implementing the different data validation efforts explained in section 4.4.3. Given this, to guarantee robustness in our results, we focus on 36 countries for which the coverage on employment and revenues is at least 70%. Tables A.5.1 to A.5.10 (in the Annex) present aggregate statistics and descriptive statistics of SOEs by region, type of contestability, type of ownership, age, and size. In our sample of 36 countries, we have a total of 40,596 SOEs with employment and revenues totaling at around 9 million workers and 1.2 trillion USD, respectively. Of these SOEs, 76% have a majority state participation, 14% are SOEs with blocking minority participation, and 10% are SOEs with minority state participation.

Europe & Central Asia (ECA) is the region with the best coverage, so that 15 countries of the countries in our reduced sample are in this region (Table A.5.1). The median employment of SOEs for countries in this region is 28 workers. In comparison to the other regions the median employment is much lower as the median SOE in each of the remaining regions has more than 100 workers (Table A.5.2). The median Latin America (LAC) SOE is more than ten times as large as the median SOE in ECA. These employment patterns are also maintained for operating revenues. That is, the median firm in ECA has the lowest level of revenue while median LAC SOE has the highest level of revenue. Although most SOEs are majority owned across all regions, firms in LAC are associated with higher government control as more than 90% of SOEs are majority owned. Majority owned SOEs in East Asia & Pacific (EAP) account for only 61% of all SOEs; the share of SOEs with blocking minority participation is the highest for this region at 25%.

In our sample of 36 countries, SOEs are present predominantly in competitive sectors that are viable for private participation. We find that 69% of SOEs operate in competitive activities (e.g., manufacturing of textiles), 16% are in partially contestable industries (e.g., air transportation services) and 15% are Natural Monopolies (e.g., water and sewerage) (Table A.5.3). For the full description of the sectors and their classification, please refer to Dall'Olio, et al. (2022). This suggests that in many countries, SOEs do not operate based on an efficiency-based rationale, but there may be other considerations for the participation of the state in the economy. In terms of employment and revenues, SOEs in competitive industries are smaller than firms in natural monopoly and partially contestable sectors (Table A.5.4). The median employment of an SOE in competitive, natural monopoly, and partially contestable activities is 28, 35, and 45, respectively. The revenues of the median SOE in natural monopoly sectors is 1.5 times that of the median SOE in competitive industries. For the median SOE in partially contestable sectors, operating revenues are 2.4 higher than the median SOE in competitive markets. As expected, governments have higher control of SOEs in natural monopoly sectors, as 90% of natural monopoly SOEs are majority owned. A striking finding for our sample is that SOEs in competitive industries are associated with higher control than SOEs in partially contestable sectors, although the efficiency rationale for state participation is much weaker in these sectors.

Two-thirds of the SOEs in our sample of 36 countries are directly owned by governments, while the remaining SOEs are indirect subsidiaries (Table A.5.5). This shows the importance of considering indirectly owned companies in measuring the state footprint within economic activity. Comparing directly owned SOEs to indirect subsidiaries, we find that there are almost no differences in the median level of employment (Table A.5.6). However, the median directly owned SOE is smaller than the median indirect subsidiary in terms of operating revenues. SOEs that are directly owned by the government are also associated with higher government control in terms of shareholding participation. More than 91% of

directly owned SOEs have a majority participation of the government. On the contrary, less than half of indirectly owned SOEs (48%) have majority participation by the government.

Most of the SOEs in the sample of 36 countries are in the age between 6-25 years, which account for the largest share of revenues and employment when compared to other age groups. We study the SOEs in our sample for different age groups: 0-5 years, 6-15 years, 16-25 years, 26-50 years, and more than 50 years. The first finding informed by the sample is that most SOEs are aged between 6 and 25 years (Table A.5.7). The age groups "6-15 years" and "16-25 years" also account for the highest share of employment and revenues among the different age groups. The second finding is that there is a positive relationship between age and economic activity. That is, SOEs that are older have higher levels of employment and revenues. The last finding is that SOEs are associated with higher government control. For example, 83% of SOEs aged between 26 and 50 are majority owned, while 74% of SOEs aged between 0 and 5 years are majority owned.

Large SOEs (i.e., over 250 workers) in the sample account for 77% of the total employment and 86% of the revenues in the sample. Last, we assess patterns of SOEs in our sample associated with SOE size (Tables A.4.9 and A.4.10) We classify firms into four size groups based on employment: 1-19 employees, 20-100 employees, 100 to 250 employees, and more than 250 employees. We find that SOEs with more than 250 workers account for 77% of total employment and 86% of total revenues, despite only representing 28% of all SOEs in our sample. With respect to state participation, there is no clear pattern between SOE size and state participation. For example, SOEs with 1 to 19 employees have the lowest share of firms with majority state participation, while SOEs with 20 to 100 employees have the highest share of firms with majority state participation.

Some evidence suggests that the higher state participation, the lower the labor productivity among SOEs even controlling by sector, age, size, and other relevant variables. To understand the implications of the different firm characteristics described above on SOE performance, we regress a proxy of SOE performance on region, type of contestability, type of ownership, age, size, subnational SOE dummy, and government shareholding percentage (Table 2). The performance measure we use is operating revenues per worker (i.e., performance-based productivity). Region, type of contestability, type of ownership, age, and size are dummy variables categorized according to the different variable categories described above. In addition, the variable we include for state participation is the minimum shareholding percentage for all the ownership layers of an SOE that link it to the government.[49] Our regional dummy suggests that SOEs in LAC perform the best, when compared to other regions.

There are differences in performance based on the sector of operation. SOEs in natural monopolies perform worse than SOEs in competitive industries, which could be explained by the fact that governments' rationale for SOEs in natural monopoly sectors are not motivated only by efficiency considerations and are often linked to public service obligations (i.e., supply of water or postal services). There are no significant differences in performance-based productivity between SOEs in competitive sectors and partially contestable sectors. We find that subnational SOEs seem to underperform compared to those owned by the central government. Larger SOEs are associated with higher levels of performance- based productivity. Also, older SOEs are associated with higher productivity (until the age of 25), but the significance is not robust to the inclusion of clustered standard errors.

---

[49] For example, consider a group of SOEs (firms A, B, and C) that are linked to a government public authority (PA) in the following way: PA (50%) ->A (25%) -> B (30%) -> C. The value of shareholding percentage for firm A in the regression is 50%, while the value of shareholding percentage for firms B and C is 25%, as this would be the minimum shareholding percentage for all the ownership layers that links them to PA.

With respect to government control variables, we find that closer control of the government might be related to lower performance of the SOE in our sample. Directly owned SOEs seem to have a lower performance-based productivity. Likewise, the higher the share-holding percentage, the lower an SOEs' productivity. This suggests that government participation could be associated with a relatively lower performance of SOEs, supporting a large strand of literature that has documented that SOEs underperform in markets.[50]

Table 2. Performance of SOEs explained by SOE characteristics

VARIABLES	(1) Labor productivity (Operating revenues per worker)	(2) Labor productivity (Operating revenues per worker)
Directly owned	-0.371***	-0.371***
	-0.029	-0.089
State participation (continuous variable)	-0.008***	-0.008***
	0.0000	-0.002
Region: Europe & Central Asia	-0.088*	-0.088
	-0.048	-0.123
Region: Latin America & Caribbean	0.982***	0.982***
	-0.133	-0.311
Region: South Asia	-0.277*	-0.277
	-0.156	-0.205
Region: Sub-Saharan Africa	-0.021	-0.021
	-0.104	-0.187
Sector contestability: Natural Monopoly	-0.131***	-0.131
	-0.024	-0.151
Sector contestability: Partially Contestable	-0.031	-0.031
	-0.031	-0.261
Size :100-250	0.494***	0.494***
	-0.032	-0.121
Size :20-100	-0.046*	-0.046
	-0.025	-0.111
Size :250+	0.432***	0.432***
	-0.033	-0.145
Age :16-25 years	0.081**	0.081
	-0.035	-0.119
Age :26-50 years	0.03	0.03
	-0.036	-0.112
Age :50+ years	-0.298***	-0.298
	-0.047	-0.272
Age :6-15 years	-0.043	-0.043
	-0.035	-0.121

---

[50] Forthcoming analytical analysis is exploring these relationships in more detail including a comparison in performance between private owned companies and SOEs using ORBIS data.

Subnational (=1 if municipal)	-0.565***	-0.565***
	-0.027	-0.068
Constant	4.393***	4.393***
	-0.059	-0.239
Observations	**29,712**	**29,712**
R-squared	0.137	0.137
Sector FE	Yes	Yes
Cluster	No	Yes: Sector 2d

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

7. Conclusion

This paper describes a novel methodology for the construction of a global data set of state-owned enterprises, the Global Business of the State (BOS) database. The BOS database is more comprehensive than existing efforts to date since the BOS database (i) builds on a harmonized definition and expands the threshold beyond majority owned companies allowing for a more comprehensive and yet flexible approach, (ii) identifies and unveils both direct and indirect linkages of companies including presence at the national and subnational level as well as across borders, (iii) includes all type of sectors including natural monopoly sectors, which are relevant for assessing the state footprint in the markets, (iv) adjusts for the measurement error in ORBIS through the extensive and systematic use of supplementary databases and country team validation that allows to identify firms that are considered as private in ORBIS, (v) minimizes the omission error including companies that are operating in markets with state participation with the support of complementary sources and government counterparts, but are not captured in ORBIS; (vi) includes non-corporatized companies that are aligned to our proposed definition, but for which the shareholder information in ORBIS is incomplete, and (vii) includes countries with fewer than 30 firms and does not restrict the companies to any threshold in assets. As a result, the Global BOS database includes data on more than 48,000 firms across 80 countries, whereas the second largest database by the number of SOEs, the EBRD database (Borkovic & Tabak, 2020), covers 17,600 firms in 25 countries.

To construct the database, we use the ORBIS data set provided by Bureau van Dijk. The ORBIS database provides the ownership structures between shareholders and their direct subsidiaries and firms' financial data. However, low data coverage and outdated ownership structures found in the ORBIS database restrict researchers' ability to use it directly. Moreover, the lack of a uniform SOE definition restricts the ability to identify SOEs across countries to perform cross-country comparisons. Given the data caveats of ORBIS, we complement the information of this database with different data sources including WBG project Databases, external BOS databases, and country-specific firm-level data sets.

Our methodological approach allowed us to identify four times more companies with state participation across 80 countries. It also improved the coverage of financial information by 4.4 times in employment reaching more than 11 million workers, and 3.7 times more revenues surpassing USD 1.5 billion, to provide a more comprehensive measure of the real state footprint in the markets. The Global BOS database sheds light on the very diverse set of sectors in which the state participates as a market player, in many cases without a clear rationale for state's direct intervention in the form of an SOE. In many countries, SOEs are

still predominant in commercial activities that can be viable for the private sector[51] such as fiber and textile manufacturing (e.g., the Arab Republic of Egypt), manufacture of food and cardboard (e.g., Bolivia), meat production (e.g., Botswana), groundnuts production (e.g., The Gambia), and even financial intermediation firms and digital payments (e.g., Russia, Vietnam) that could be served by the private sector. State participation in these sectors could potentially crowd-out private sector participation and thus private sector competitiveness, and as such, warrants further attention from policy makers.[52] Nonetheless, it is important to note that the BOS database collects only the primary activity of the SOEs, even though some companies could be multi-product. Since the BOS database captures the different legal entities even if under a conglomerate group, the risk of misclassification of the main activity is limited, but for the policy agenda and assessment of options of reforms this can be complemented with other tools (e.g., MCPAT).

The Global BOS database shows that state ownership and relationships among firms with state participation can be very complex. For the first time, the Global BOS database provides an in-depth view of the ownership relationships between the government and the corporate sector. It allows to identify countries where there are large conglomerate groups (e.g., Egypt and Vietnam) with strong upstream and downstream relationships. It also provides new tools to assess the degree of vertical integration of some firms across markets, which can also be relevant to understanding potential implications in terms of competition and firm dynamics.

Last, the Global BOS database has the potential to inform key reform agendas in areas such as corporate governance, anticorruption, debt management, among others.[53] The rich information contained in the database has been used to inform country-level analysis around SOE reform, but its uses go beyond corporate governance. For instance, the SOE agenda is directly relevant to macro-fiscal management given SOEs potential impact on direct and contingent public sector liabilities. SOEs are also important employers in certain markets, with implications for the jobs and poverty agenda. Finally, SOEs have become more relevant in the context of the COVID-19 pandemic, as both critical elements in the production of emergency protective and medical equipment in the case of manufacturing firms, but also as recipients of state support, for example in the airline industry (Sanchez-Navarro, Martinez-Licetti, & Perrotet, 2020).

---

[51] (Sanchez-Navarro, Goodwin, & Kikeri, SOE CPSD Knowledge note, 2021). CPSD SOE Knowledge note (2021).

[52] Who's the BOSs : Shedding New Light on Businesses of the State, (World Bank, 2022).

[53] Ibid.

Annexes

Annex 1. Comparability with other existing SOE databases

Most available SOE databases are not comprehensive as they either focus on collecting sectoral data or have limited coverage and scope. Databases that focus on specific sectors are primarily centered on economic sectors commonly associated with SOE presence, such as infrastructure and finance. Examples of infrastructure databases include the World Bank Database of Infrastructure State-Owned Enterprises (Herrera Dappe, et al., 2022), covering 19 countries and 135 SOEs between 2000 and 2018, and the 2017 State-Owned Enterprises Public Projects (SPI) database (PPIAF/The World Bank, 2017) compiled by the World Bank's Public-Private Infrastructure Advisory Facility. Analyses of SOE presence in the financial sector include the work of La Porta, Lopez de Silanes and Shleifer (2002), who assembled data on government ownership of banks in 92 countries, the cross-country data set of state-owned banks compiled by Andrianova, Demetriades, & Shortland (2012) for 1997-2007, and the WB State Bank Privatization database that covers 70 countries between 1995-2017 (Can, Calice, Diaz, & Masseti, 2020).

Other existing databases compile national or regional SOE data, although with limited global country coverage and scope. Hence, a comprehensive, global cross-country database that identifies SOEs and their financial data has never been compiled. Existing efforts so far are limited to certain regions and in particular, coverage of developing countries has been lacking. For example, in 2012, 2015, and 2017, the OECD conducted an exercise to identify the presence of SOEs across 40 economies, which provided aggregated data for about 2,400 firms including information on the number of SOEs and their sectoral distribution based on a self-reporting survey tool of participating countries.[54] However, the OECD effort relied heavily on the local SOE definitions, which can vary substantially across jurisdiction. The IMF also collected firm-level data for about 10,000 SOEs leveraging the ORBIS database from *Bureau van Dijk* for the period 2014-2016 although this one excludes SOEs operating in natural monopolies or sectors where private firms are barely present (IMF, 2019), and omits countries with fewer than 30 SOEs or restricts the sample to the largest companies implementing some thresholds in assets.[55] The original exercise covered 20 countries in Central, Eastern and Southeastern Europe (IMF, 2019) and in 2021, it was expanded to the Middle East, North Africa, and Central Asia region (IMF, 2021). The IADB led an exercise to collect financial performance information of non-financial SOEs for 16 countries in the LAC region covering the period 2010-2016 (Musaccio & Pineda, 2019). Other efforts include an analysis of majority-owned Chinese SOEs (Freund & Sidhu, 2017) (Harrison, Meyer, Wang, Zhao , & Zhao, 2019); the IMF Fiscal Monitor, which included firms with 20 percent of state participation and above (IMF, 2020); and the EBRD cross-country study on SOEs with about 12,000 companies with at least 25% participation of the state (Borkovic & Tabak, 2020). Table A.1 summarizes existing comparable SOE databases.

A global SOE database cannot be simply compiled by combining the various existing databases due to the lack of a uniform definition of SOEs. In absence of a commonly accepted definition of SOEs,[56] regional and sectoral exercises cannot be merged to create a single repository. As shown in Table A.1, each one of the aforementioned databases employs a different definition of SOEs with varying degrees of state ownership thresholds, and some of them follow countries' official SOE definition, which varies significantly and complicates comparability even further. For example, the survey-based exercise led by the OECD quoted above collects information on the corporate entities recognized as SOEs by the respective national laws

---

[54] https://www.oecd.org/publications/the-size-and-sectoral-distribution-of-state-owned-enterprises-9789264280663-en.htm

[55] Given the low coverage for some countries such as Albania and Kosovo many were not included in the exercise. Annex 4 (IMF, 2019) indicates that the IMF database does not include neither SOEs in natural monopolies nor countries with less than 30 SOEs and limit the assessment to those with total assets over USD 100,000.

[56] (IMF, 2020). https://www.elibrary.imf.org/view/books/089/28929-9781513537511-en/ch03.xml

(OECD, 2017). The use of different definitions naturally poses a limit to cross-country analysis outside the limited set of countries covered by the individual databases.

Table A.1 . Summary of existing cross-country SOE databases

	World Bank Businesses of the State (BOS) database	World Bank (2021 Infrastructure) [57]	IMF[58]	ADB[59]	EBRD[60]	EU[61]	OECD[62]	IDB[63]
Countries	81 developing countries	19	20	9	25	28 EU	40	18 (LAC only)
# SOEs captured	57,000+ firms with state participation.	135 SOEs	10,000 SOEs	12,742 SOEs	17,600 SOEs	950 SOEs focusing on sectors: electricity, gas, and railways.	2,467 SOEs for 39 OECD countries, 55,341 for China.	1,019 SOEs
Period	2019	2000-2018	2014-16	2010-2018	2014-2016	2008-2013	2012, 2015, 2017	2010-2016
SOE definition	Unified definition across countries (10%+) including the full ownership tree (i.e., path with company name and specific participation that connect the companies with the state)	50%+ state participation	At least 25%	At least 50.01% participation (default of ORBIS GUO)	At least 25% participation	For analysis purposes: At least 20% ownership	*As defined by local authorities and respondents to survey including majority (50%+ participation) and minority participation (from 10 to -49%)*	*As defined by local authorities and respondents to survey*
Sectors	NACE 4 digit; all sectors including Infra financial sector as well as real sector (e.g, Agri, manufacturing) and services. It excludes health and Education, public administration.	Infrastructure assets in power and transportation sectors	Only 1-digit sectors (only total values by sector, not-firm level data)	All sectors	Non-financial SOEs to prevent distortion	Non-financial SOEs	Only 1 digit sectors (only total values by sector, not-firm level data published)	Non-financial SOEs
Focus of Analysis	State footprint, breakdown by type of sector, financial indicators, governance indicators, firm-level characteristics, and ownership structure	Financial performance, benchmarked against POEs	Financial performance against POEs, ownership structure and productivity.	Financial performance.	Financial performance of SOEs	Financial performance of SOEs	Number of firms, total workers, and sector (digits) of operation	Financial performance of SOEs;

[57] Infrastructure State-Owned Enterprises: A Tale of Inefficiency and Fiscal Dependence (Herrera Dappe, et al., 2022)

[58] State-Owned Enterprises in Middle East, North Africa, and Central Asia: Size, Costs, and Challenges (IMF, 2021)

[59] Reforms, Opportunities, and Challenges for State-Owned Enterprises (Ginting & Naqvi, 2020)

[60] Economic Performance of State-Owned Enterprises in Emerging Economies: A Cross-Country Study (EBRD, 2020)

[61] State-owned Enterprises in the EU: Lessons Learnt and Ways forward in a Post-Crisis Context (European Commission, 2016)

[62] OECD Size and sector distribution of state-owned enterprises. (OECD, 2017)

[63] Fixing State-Owned Enterprises: New Policy Solutions to Old Problems (Musaccio & Pineda, 2019)

Variables	Sector of operation, Age of company, Revenues, Employment, Profit/loss, State participation, Level of government (central, subnational), audit status, Ownership structure (participation and shareholders)	Accounting financial data, SOE ownership structure and income (netted fiscal transfers)	Revenues, employment, Return on Assets. Corruption (WGI), and productivity	Financial data, Ownership data	22 indicators on financial performance	Share of SOEs in total rail turnover vs OECD Public Ownership Index in Rail; Financial performance of SOEs	Number of firms, total of labor, and total of revenues	Financial performance and netted fiscal transfers.
Source	Orbis; government's sources, EMIS intelligence, Factiva, government sources (full list in annex)	Government sources and company websites; EMIS; stock exchanges	Orbis; S&P Capital IQ; Surveys to public authorities; external sources	Orbis, external databases	Orbis, government sources, external databases	Orbis, Amadeus databases; government sources; external databases	SOE Surveys to authorities (2012, 2015 and 2017)	SOE Surveys to authorities

Source: Authors' compilation based on listed sources.

Preliminary results underscore the power of the methodological approach developed in this paper to capture an unprecedented number of firms with state participation, in comparison to other existing databases. Figure A.1 shows a comparison between the number of firms captured by the World Bank's Global BOS database and the OECD database. We found that in almost all cases, the BOS database captures significantly more firms with state participation and in some cases, as many as 30 times more than comparable exercises, as in the case of Poland. Figure A.2 shows a similar finding when comparing the BOS database to the IADB database of SOEs.

Figure A.1 Comparison of Country-level SOE figures. World Bank's Global BOS database vs. OECD database
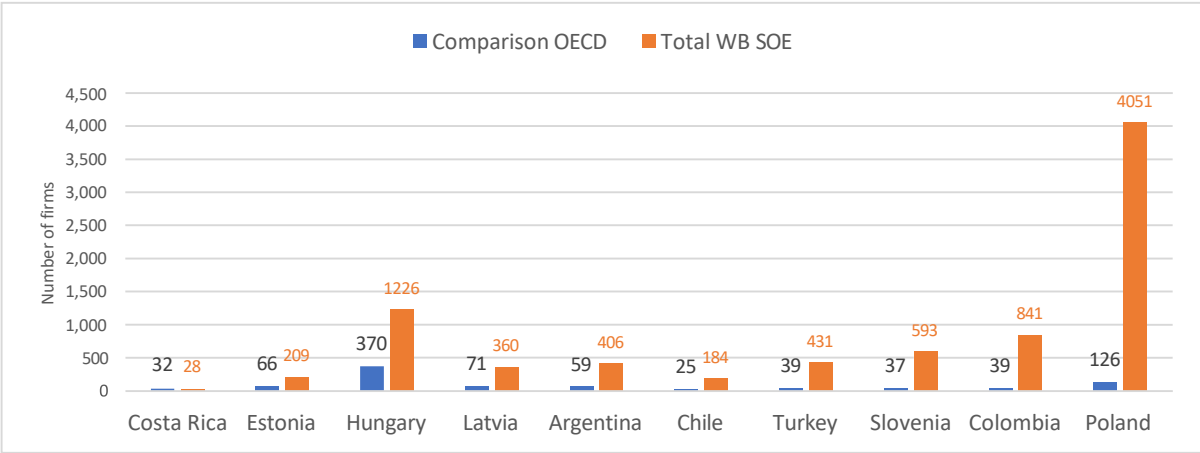
Source: Authors' compilation based on Global BOS database and OECD (2017) data

Figure A.2 Comparison of Country-level SOE figures. World Bank's Global BOS vs. IADB database

Comparison number of firms with state participation WB SOE vs. IADB database
(IADB =1)
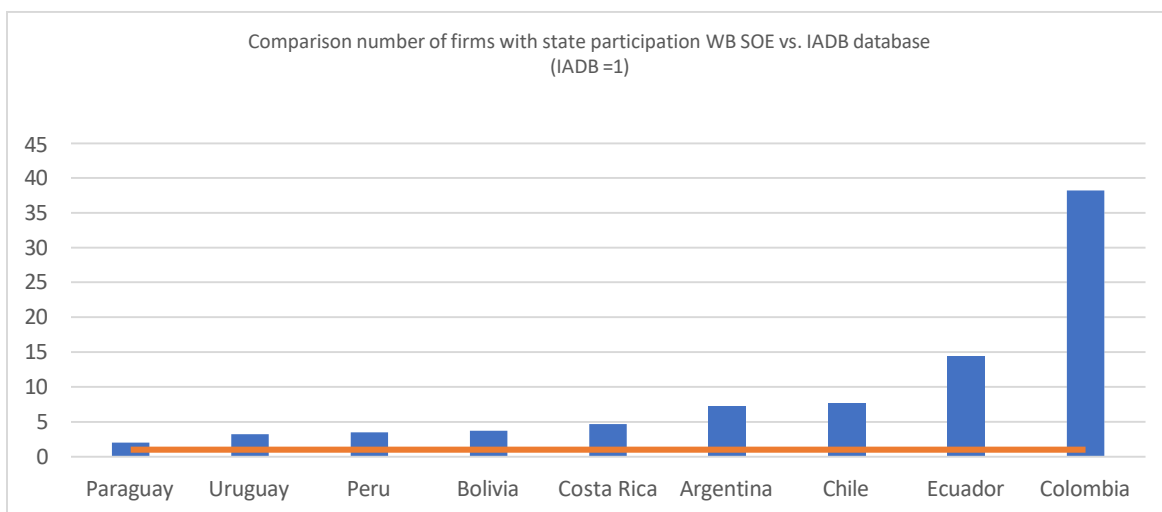
Source: Author's compilation based on Global BOS database and IADB (2019) data

Furthermore, the database fills an important gap in the coverage SOEs in developing countries. The number of SOEs captured through this methodology is significantly larger than the number of SOEs captured by comparable exercises led by other international organizations as seen in Figure A.3. The BOS database even covers more companies and sectors than the approach proposed by the IMF (2019) using ORBIS.

Figure A.3 Number of countries captured by cross-country BOS databases
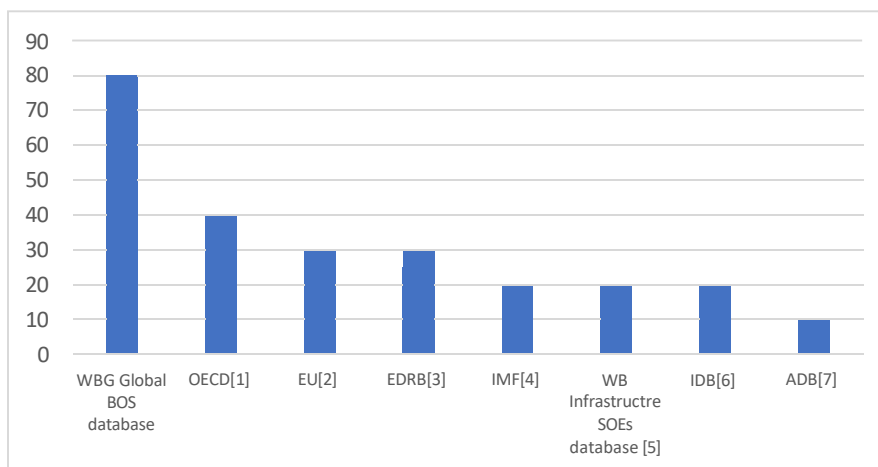
Source: Global BOS database and [1] OECD Size and sector distribution of state-owned enterprises (2017); [2] State-owned Enterprises in the EU: Lessons Learnt and Ways forward in a Post-Crisis Context (2016); [3] Economic Performance of State-Owned Enterprises in Emerging Economies: A Cross-Country Study (2020); [4] State-Owned Enterprises in Middle East, North Africa, and Central Asia: Size, Costs, and Challenges (2021); [5] Infrastructure State-Owned Enterprises: A Tale of Inefficiency and Fiscal Dependence (2021); [6] Fixing State-Owned Enterprises: New Policy Solutions to Old Problems (2019); [7] Reforms, Opportunities, and Challenges for State-Owned Enterprises (2020).

In addition, this approach allows for the first time to conduct comparable cross-country assessments. Figure A.4 shows a country-level set of indicators created for Slovenia using the Global BOS database. This approach offers the possibility to create standard SOE indicators such as the number of SOE enterprises operating in the country, revenues, and employment, as well as indicators that describe the breakdown between directly and indirectly owned SOEs, the distribution of SOEs by sector according to a sector

taxonomy based on the degree of contestability, the distribution of SOEs by government participation (including a category for minority participation from 10% to 25%) and the percentage of national vs subnational SOEs.

Figure A.4 Example of country-level output from Global BOS database: Slovenia dashboard
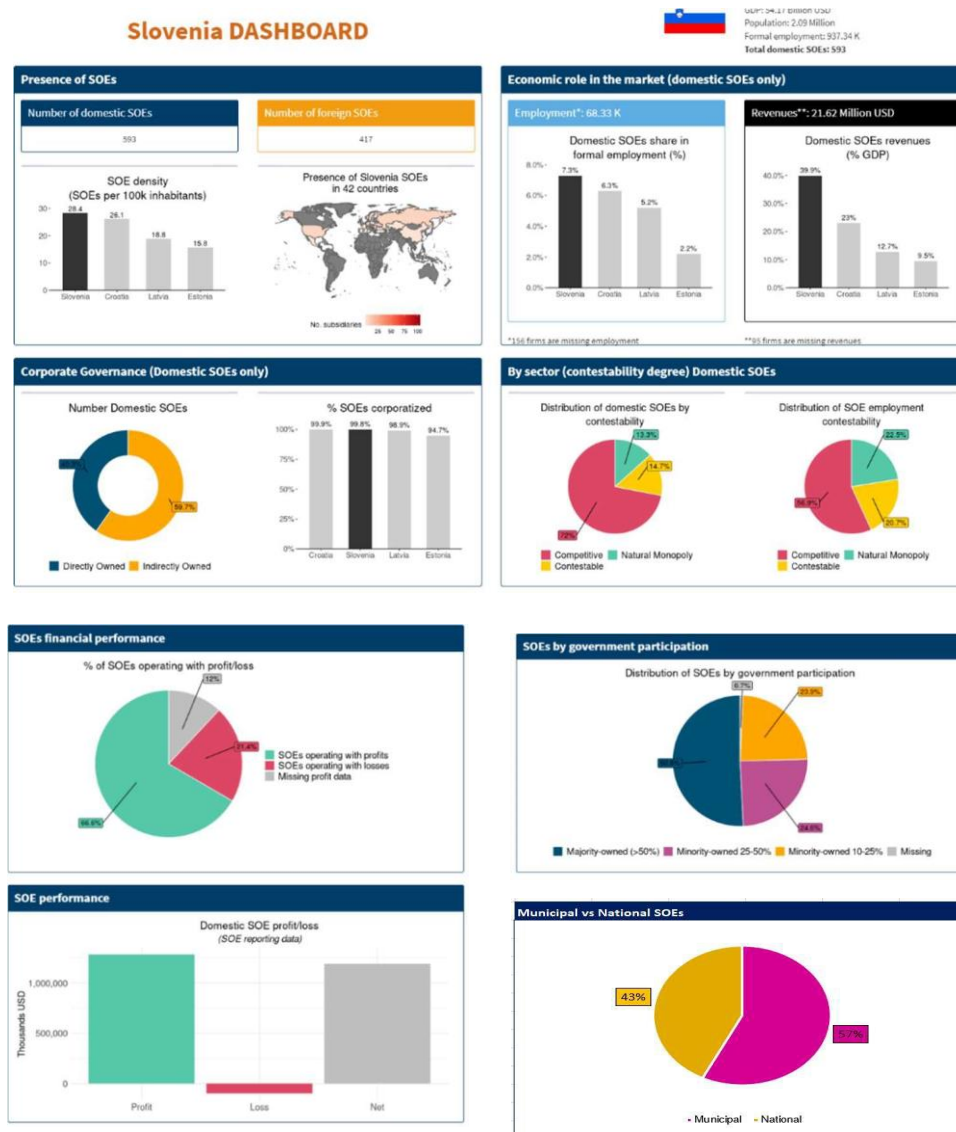
Source: Author's elaboration based on Global BOS database

Annex 2. Identification of non-corporatized SOEs

The BOS database identifies an SOE's corporatization type based on its national legal form. An SOE is corporatized if its corporate legal form corresponds to a standard legal form that any private sector firm can, with the only difference being that the government is a direct or indirect shareholder. Some common legal forms of corporatized SOEs are Limited Liability, Joint Stock Company, among others. An SOE is non-corporatized if it is classified as a state company according to law, and this legal status is not a standard corporate form (i.e. does not have shareholder structure).

Given that national corporate law and state-ownership laws vary, countries are characterized by different types of national legal forms. For instance, in Indonesia SOEs according to national state-ownership law are denoted with two different legal forms: Badan Usaha Milik Negara (controlled by the central government) and Badan Usaha Milik Daerah (controlled by the local government). Similarly, in Azerbaijani SOEs are defined as public interest entities (PIEs), in Mozambique as public enterprises and shareholding companies, and in Rwanda as corporate entities recognized by national law as an enterprise in which the state exercises ownership. In some cases, national legal forms of state ownership include state corporations as well as regulators or administrative agencies with some commercial activities (e.g., parastatals).

Given that many SOEs are corporatized and the large heterogeneity in corporate and state-ownership laws, national legal forms cannot be solely used as a variable to identify SOEs. However, the use of national legal forms is essential to ensure that non-corporatized SOEs are included in the BOS database. As most non-corporatized SOEs do not have a shareholder structure, these companies do not provide shareholding percentage information in ORBIS's links file. As a result, these companies will not be identified as SOEs using an approach that relies on linkages to public authorities in ORBIS, as these linkages are not present. For this reason, we incorporate into the identification of SOEs as well as the construction of ownership structures the information pertaining to national legal forms. In this manner the identification approach ensures that all companies -even under different legal forms- are included in the database when they satisfy the criteria of the SOE definition. For this, national legal forms were retrieved from an ORBIS variable with the same name and then cross-checked with local country teams with the legal expertise to corroborate that those were aligned to the definition proposed in section 2. Once those legal forms were verified, we identified the companies with legal forms associated to state ownership by law as non- corporatized SOEs. In addition, some of these companies can have subsidiaries. Hence, for these non- corporatized SOEs we explored whether those companies are shareholders in other subsequent companies using the algorithm in Figure 7.

Annex 3.

Table A.2 Complementary data sources for the global BOS database beyond ORBIS

Database	Vendor/Source	Coverage	Variable coverage	
			Financial Information	Ownership information
EMIS Intelligence	EMIS	5.5 million listed and unlisted firms in 125 emerging markets	Yes	Yes (name of shareholders). SOEs can be more easily identified for countries with centralized SOE ownership.
Factiva/Factset	Dow Jones/Factset Research	Listed and unlisted companies	Only for listed companies	Yes (binary variable on SOE)
Worldscope	Thompson Reuters/Refinitiv	Listed companies	Yes	Yes (common shares outstanding with information on as "government owned company or majority owned by government".
Global 2000	Forbes	Largest 2000 firms worldwide measured by sales, profits, assets and market value	No	No
PMR	OECD/WBG	OECD & 40+ developing economies	No	Yes
PitchBook	PitchBook	3.3 million firms worldwide covering public listed and private (unlisted) companies	Yes	Yes (for public companies, including list of shareholders and percentages held)

The Banker Database	The Financial Times	Firms in the financial sector in more than 190 countries covering 5,000+ public, private, government and subsidiary banks accounting for more than 90% of total global banking assets	Yes	No
Fitch Connect	Fitch Solutions	Financial data on 36,000+ banks, 12,000+ insurers, 117 sovereigns, and 3,000+ corporates globally across developed and emerging markets.	Yes	No
	IJGlobal (part of the Euromoney Institutional Investor PLC group)	Firms in the global infrastructure finance sector, including 636 SOEs globally		
Uniworld online	Uniworld	Firms with headquarters in over 200 countries and 20,000 industries	Yes (revenue, employees)	No
Bloomberg	Bloomberg Finance L.P.	Data for more than 5 million legal entities, including public & private companies, sovereign entities, funds, governments, agencies, and municipalities	Yes	Yes
Banker's Almanac	LexisNexis Risk Solutions	Firms in the financial sector covering 200,000 financial institutions globally	Yes	Yes
CB Insights	CB Insights	Global database on private firms and investor activities.	Yes	Yes (investment and acquisition data only)
S&P Capital IQ	S&P Global Inc.	Data on more than 62,000 public firms and 18.2 million private firms globally.	Yes	Yes
Extractive Industries Transparency Initiative (EITI)	EITI	Firms in the extractive industries sector, covering 60 SOEs in 20 EITI countries	No (revenue data at the mining project level)	Yes (tagged as SOE)

Source: Authors' elaboration

Annex 4.

Figure A.5 Value added of the methodological approach beyond ORBIS – Country-level examples

I.   Identification of the full registry of SOEs
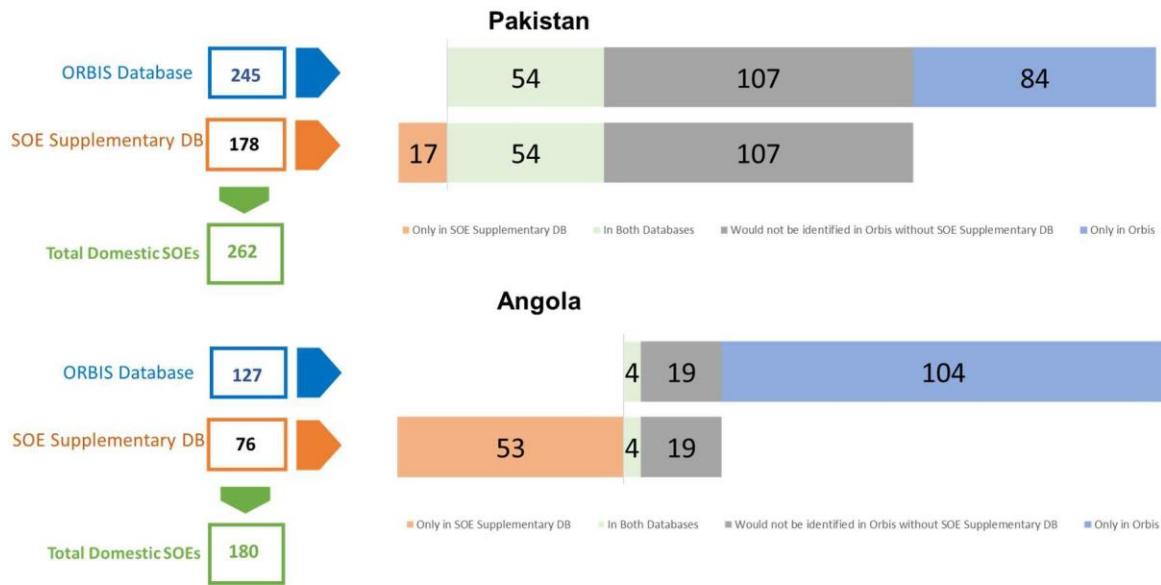
Source: Authors' elaboration based on Global BOS database.

Figure A.6 Value added improving coverage for Financials & Performance Variables
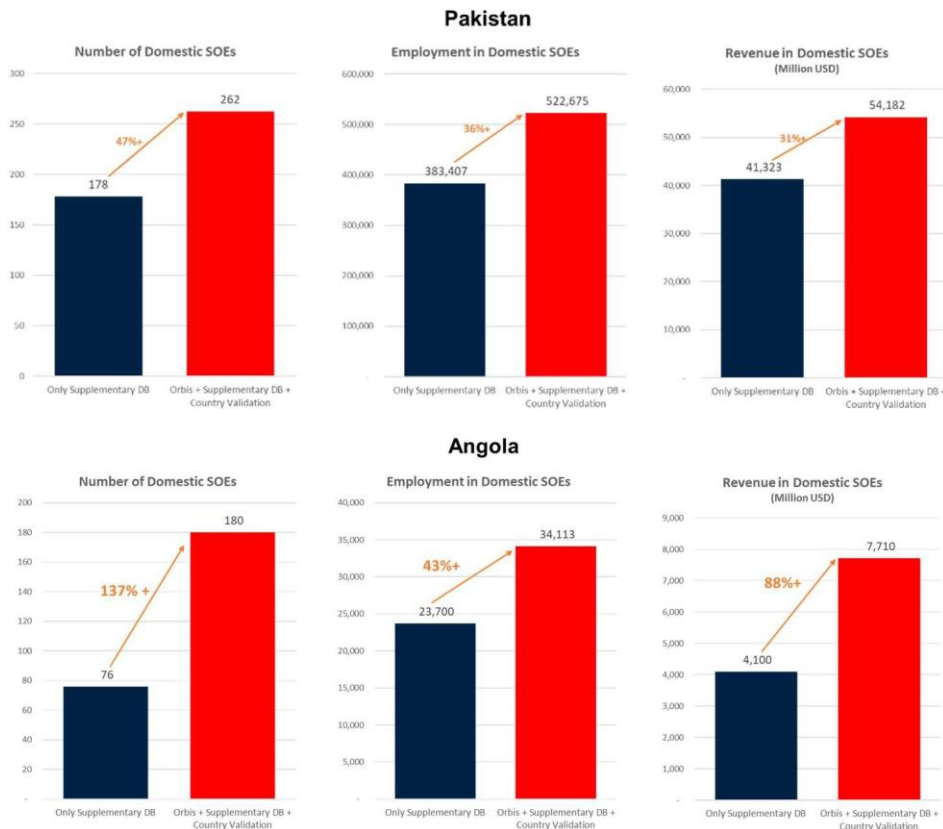
Source: Authors' elaboration based on Global BOS database.

Annex 5. Descriptive Statistics of the BOS database
Table A.5.1. Aggregate Statistics of Sample by Region

Region	No. countries (70%+ coverage)	Total firms	Total revenues (thousand USD)	Total employment	Firms directly owned	Firms with minority participation [10-24%]	Firms with blocking minority participation [25-50%]	Firms with majority participation [50%-100%]
East Asia & Pacific	2	2,160	144,800,000	901,251	847	308	530	1,311
Europe & Central Asia	15	37,228	925,500,000	6,755,836	25,732	3,191	4,669	26,779
Latin America & Caribbean	4	189	33,070,622	162,704	143	4	14	171
South Asia	5	504	70,867,819	959,967	303	64	54	384
Sub-Saharan Africa	10	515	21,091,076	233,462	347	45	59	400
Total	**36**	**40,596**	**1,195,329,517**	**9,013,220**	**27,372**	**3,612**	**5,326**	**29,045**

Table A.5.2. Descriptive Statistics of Sample by Region

Region	Employment				Operating Revenues (thousand USD)				Level of state participation			
	Min	Median	Std. Dev	Max	Min	Median	Std. Dev	Max	Min	Median	Std. Dev	Max
East Asia & Pacific	1	120	1,969	46,508	0	7,194	369,706	5,618,062	10	50	24	100
Europe & Central Asia	1	28	4,516	729,281	0	530	399,939	29,853,835	10	100	31	100
Latin America & Caribbean	3	301	1,778	13,037	157	17,546	533,514	4,585,247	12	100	23	100
South Asia	3	281	6,438	75,000	0	7,066	545,451	7,098,000	10	99	33	100
Sub-Saharan Africa	1	168	1,324	18,031	0	6,492	125,463	923,094	10	100	30	100

Table A.5.3. Aggregate Statistics of Sample by Contestability of Sector

Sector type	Number of firms	Operating revenues	Employment	Firms directly owned by the state	Firms with minority participation [10-24%]	Firms with blocking minority participation [25-50%]	Firms with majority participation [50%-100%]	Subnational firms
Competitive	27,212	517,500,000	3,856,672	18,068	2,662	3,563	18,796	13,467
Natural Monopoly	6,201	146,600,000	1,748,488	5,062	185	437	5,412	4,773
Partially Contestable	6,295	527,900,000	3,377,853	3,886	653	1,094	4,395	2,235

Table A.5.4. Descriptive Statistics of Sample by Contestability of Sector

Sector type	Employment				Operating revenues (thousand USD)				State participation			
	Min	Median	Std. Dev	Max	Min	Median	Std. Dev	Max	Min	Median	Std. Dev	Max
Competitive	1	28	1,832	238,997	0	534	285,833	22,558,914	10	100	32	100
Natural Monopoly	1	35	1,734	78,484	0	816	185,092	5,958,143	10	100	24	100
Partially Contestable	1	45	10,454	729,281	0	1,308	802,163	29,853,835	10	100	33	100

Table A.5.5. Aggregate Statistics of Sample by type of ownership (direct vs. indirect)

Ownership type	Number of firms	Operating revenues (thousand USD)	Employment	Firms with minority participation [10-24%]	Firms with blocking minority participation [25-50%]	Firms with majority participation [50%-100%]	Subnational firms
Indirectly owned	13,224	746,100,000	4,025,562	2,841	4,030	6,349	3,009
Directly owned	27,372	449,200,000	4,987,658	771	1,296	22,696	17,561

Table A.5.6. Descriptive Statistics of Sample by type of ownership (direct vs. indirect)

Sector type	Employment				Operating revenues (thousand USD)				State participation			
	Min	Median	Std. Dev	Max	Min	Median	Std. Dev	Max	Min	Median	Std. Dev	Max
Indirectly owned	1	32	2,690	238,997	0	1,688	489,219	25,176,507	10	49	31	100
Directly owned	1	30	5,062	729,281	0	477	345,079	29,853,835	10	100	23	100

Table A.5.7. Aggregate Statistics of Sample by Age

Age bins	Number of firms	Operating revenues (thousand USD)	Employment	Firms with minority participation [10-24%]	Firms with blocking minority participation [25-50%]	Firms with majority participation [50%-100%]	Subnational firms
0-5 years	6,080	40,052,657	393,619	613	933	4,331	3,622
6-15 years	13,283	287,600,000	1,495,136	1,249	1,847	8,809	7,127
16-25 years	10,174	351,600,000	2,842,481	920	1,268	7,457	5,186
26-50 years	6,638	196,200,000	1,930,631	445	624	5,279	3,253
50+ years	2,682	248,500,000	1,536,169	233	417	1,962	1,056

Table A.5.8. Descriptive Statistics of Sample by Age

Sector type	Employment				Operating revenues (thousand USD)				State participation			
	Min	Median	Std. Dev	Max	Min	Median	Std. Dev	Max	Min	Median	Std. Dev	Max
0-5 years	1	12	706	41,422	0	210	57,037	2,235,392	10	100	32	100
6-15 years	1	25	852	58,319	0	394	278,548	20,486,548	10	100	32	100
16-25 years	1	37	8,260	729,281	0	934	561,361	29,853,835	10	100	31	100
26-50 years	1	61	1,751	73,829	0	1,548	258,722	7,377,555	10	100	29	100
50+ years	1	110	2,757	78,484	0	3,857	764,532	23,652,460	10	100	32	100

Table A.5.9. Aggregate Statistics of Sample by Size (Employment)

Size (Number of workers)	Number of firms	Operating revenues (thousand USD)	Employment	Firms with minority participation [10-24%]	Firms with blocking minority participation [25-50%]	Firms with majority participation [50%-100%]	Subnational firms
1-19	12,641	57,354,500	91,088	1,556	2,038	8,133	6,990
20-100	12,632	72,575,061	565,742	762	1,084	10,435	8,793
100-250	3,799	142,500,000	601,396	304	593	2,852	1,708
250+	11,524	922,800,000	7,754,994	990	1,611	7,625	3,079

Table A.5.10. Descriptive Statistics of Sample by Size (Employment)

Sector type	Employment				Operating revenues (thousand USD)				State participation			
	Min	Median	Std. Dev	Max	Min	Median	Std. Dev	Max	Min	Median	Std. Dev	Max
1-19	1	5	6	19	0	85	192,966	20,486,548	10	100	33	100
20-100	20	38	21	100	0	741	69,162	5,050,109	10	100	28	100
100-250	101	151	42	250	0	5,450	275,275	9,981,498	10	100	31	100
250+	251	599	11,878	729,281	0	20,257	899,255	29,853,835	10	100	32	100

Annex 6. Dictionary of variables included in BOS database

Module in BOS database	Variable in BOS database	Description	Category
Firm-level characteristics	Firm identifier	Unique identifier in the form XX0000j, where first 2 digits correspond to ISO country code.	Text
	Country	Country full name	Text
	Company name	Full company name in latin alphabet	Text
	Local Company Name	Full company name in local language	Text
	Year of Information	Year of the financial information collected.	Numeric
	Year of start of operations	Year of incorporation of the firm (proxy of age)	Numeric
	Sector of operation (1 digit)	NACE rev. 2 classification - 1 digit (main category)	Numeric
	Sector of operation description (1d-description)	NACE rev. 2 classification - 1 digit (Description)	Text
	Industry of operation (2 digits)	NACE rev. 2 classification - 2 digits	Numeric
	Industry of operation description (2d-description)	NACE rev. 2 classification - 2 digits (Description)	Text
	Activity classification (4-digits)	NACE rev. 2 classification - 4 digits	Numeric
	Activity description (4-d - description)	NACE rev. 2 classification - 4 digits (Description)	Text
	National legal form	Legal form of the company based on national law	Text
	Corporate type (i.e., Corporatized, Non-corporatized)	Categorical variable indicating whether the firm is corporatized or non-corporatized	Categorical: Corporatized Non-corporatized
Performance module	Firm size (number of employees)	Total number of workers in the firm (permanent + temporal)	Numeric
	Consolidation code	Indicator of the level of consolidation of the financial information provided. The focus of the exercise is unconsolidated, but when that is not available, we provide the consolidated information and identify accordingly.	Categorical: C - consolidated U- unconsolidated.
	Revenues (Unconsolidated) in thousand US dollars	Operating revenues in thousand US dollars	Numeric
	Profit/loss (Unconsolidated) in thousand US dollars	Net profit/loss after tax in thousand US dollars	Numeric

	Exchange rate (LCU per 1 USD)	Exchange rate as official in December of 2019 (Local currency per 1 USD)	Numeric
Governance module	Reporting line	Level of government linked to the company as shareholder.	Categorical: 1- Central, 2- Municipal, 3- Both, 4-Other
	Reporting line within government	Name of the ministry line that acts as shareholder on behalf of the government	Text
	Audit status	Status of the latest financial statements provided by the company	Categorical: - Audited - Unaudited - Undefined
	Regulatory body (if different from reporting line)	Name of the government (sectoral) regulator that oversights the firm if different from the ministry line	Text
Ownership module	Participation by direct shareholder (%)	Shares held by the direct shareholder in the firm (in percentage)	Numeric
	First level shareholder	Name of the direct shareholder of the firm.	Text
	Participation by second level shareholder (%)	Shares held by the 2-level shareholder in the firm (in percentage)	Numeric
	Second level shareholder	Name of the second level shareholder of the firm.	Text
	Participation by third level shareholder (%)	Shares held by the 3-level shareholder in the firm (in percentage)	Numeric
	Third level shareholder	Name of the third level shareholder of the firm.	Text
	….	…	
	Participation by thirteenth level shareholder (%)	Shares held by the 13-level shareholder in the firm (in percentage)	Numeric
	Thirteenth level shareholder	Name of the 13-level shareholder of the firm.	Text
	Multiple ownership links	Categorical variable to indicate cases where more than one public authority is a shareholder in the firm.	Categorical: 1 - If more than one link is related to the government 0 - only one public entity is a shareholder is the firm

Bibliography

Andrianova, S., Demetriades, P., & Shortland, A. (2012). Government Ownership of Banks, Institutions and Economic Growth. *Economica*, 449-469.

Bajgar, M., Berlingieri, G., Calligaris, S., Criscuolo, C., & Timmis, J. (2020). Coverage and representativeness of Orbis data . *OECD Science, Technology and Industry Working Papers* .

Bank, W. (2022). *Who's the BOSs : Shedding New Light on Businesses of the State.* Washington, DC: World Bank.

Borkovic, S., & Tabak, P. (2020). *Economic performance of state-owned enterprises in emerging economies: A cross-country study.* EBRD.

Brown, T., & Potoski, M. (2003). Transaction Costs and Institutional Explanations for government service production decisions. *Journal of Public Administration Research and Theory Vol. 13 (4)*, 441-468.

Can, A., Calice, P., Diaz, F., & Masseti, O. (2020). Recent trends in Bank Privatization. *Policy Research Working Paper 9318*.

Cusolito, A. (2020). *Source code and use instructions for generating financial and ownership datasets using Orbis data.* WBG.

Cusolito, A. (2021). *WBG Clean Orbis Database: Methodoloy and User's Guide.* WBG.

Dall'Olio, A., Goodwin, T., Patino, F., Sanchez-Navarro, D., Drodz, M., & Alonso, A. (2022). *SOE sector taxonomy paper.* World Bank Group.

EBDR. (2021). *Transtition Report 2020-21: The state strikes back.*

EBRD. (2020). *Economic performance of state-owned enterprises in emerging economies: a cross-country study.* EBRD.

European Commission. (2016). *State-Owned Enterprises in the EU: Lessons Learnt and Ways Forward in a Post-Crisis Context.* European Commission Institutional Paper 31.

Freund, C., & Sidhu, D. (2017). Global Competition and the rise of China. *Peterson Institte for International Economics. Working Paper 17-3*.

Gal, P. (2013). *Measuring Total Factor Productivity.* OECD: Working Paper (2013) 41.

Gaspar, V. P. (2020, May 7). State-Owned Enterprises in the Time of COVID-19. *IMF Blog*.

Ginting, E., & Naqvi, K. (2020). *Reforms, opportunities, and challenges for state-owned enterprises.* Asian Development Bank.

Harrison, A., Meyer, M., Wang, P., Zhao , L., & Zhao, M. (2019). Can a Tiger Change Its Stripes? Reform of Chinese State-Owned Enterprises in the Penumbra of the State. *NBER Working Papers Series. WP 25475*.

Herrera Dappe, M., Musacchio, A., Pan, C., Semikolenova, Y. V., Turkgulu, B., & Barboza, J. (2022). *Infrastructure State-Owned Enterprises : A Tale of Inefficiency and Fiscal Dependence.* Washington, D.C.: World Bank.

IMF. (2019). *Reassesing the Role of State-Owned Enterprises in Central, Eastern, and Southeastern Europe.*

IMF. (2020). *Fiscal Monitor 2020: Policies to Support People during the COVID-19 pandemic.* Washington: Interantional Monetary Fund. Fiscal Affairs Department.

IMF. (2021). *State-Owned Enterprises in Middle East, North Africa, and Central Asia: Size, Costs, and Challenges.* IMF.

Kalemli-Ozcan, S., Sorensen, B., Villegas-Sanchez, C., Volosovych, V., & Yesiltas, S. (2015). How to construct nationally representative firm level data from the ORBIS global database: new facts and aggregate implications. *NBER Working Papers*, 1-113.

La Porta, R., Lopez-de-Silanes, L., & i Shleifer, A. (2002). Government Ownership of Banks. *The Journal of Finance*, 265-301.

Musaccio, A., & Pineda, E. (2019). *Fixing State-Owned Enterprises: New policy solutions to old problems.* Washington: Inter-American Development Bank.

OECD. (2009). *SOEs operating abroad: An application of the OECD Guidelines on Corporate Governance of State-Owned Enterprises to the cross-border operations of SOEs.* Paris: OECD.

OECD. (2017). *Size and sectoral distribution of state-owned enterprises.* Paris.

OECD. (2018). *Indicators of Product Market Regulation .* Paris: OECD.

ORBIS. (2011). *ORBIS: User Guide.* Bureau van Dijk.

PPIAF/The World Bank. (2017). *Who Sponsors Infrastructure Projects? Disentangling public and private contributions.* Washington, D.C.: ublic-Private Infrastructure Advisory Facility (PPIAF)/The World Bank.

Sanchez-Navarro, D., Goodwin, T., & Kikeri, S. (2021). *SOE CPSD Knowledge note.* Washington: IFC-WBG.

Sanchez-Navarro, D., Martinez-Licetti, M., & Perrotet, J. (2020). *Support to systematically large firms in hard-hit sectors: The case of airlines state-support programs amid COVID-19.* WB EFI Policy notes. Retrieved from https://worldbankgroup.sharepoint.com/sites/gge/Documents/COVID-19%20Response%20Documents/FCISupportSLF-HHS-Tourism_Airlines.pdf

WBG. (2020). *Support to systematically large firms in hard-hit sectors: The case of airlines state-support programs amid COVID-19.* Washington: WBG-EFI notes.

World Bank . (2019k). *Creating Markets in Rwanda: Transforming for the jobs of tomorrow.* Washington: World Bank.

World Bank. (2016). Republic of Mozambique: Assessing fiscal risks from the Public Corporation sector.

World Bank. (2019). *Integrated State-Owned Enterprises Framework.* Washington: WB.

World Bank. (2019). *iSOEF: Corporate governance. Module 4.* Washington: World Bank.

World Bank. (2022). *Who's the BOSs : Shedding New Light on Businesses of the State.* Washington, DC: World Bank.