Centre for Economic Performance

Discussion Paper

ISSN 2042-2695

No.1804

October 2021

Business groups as knowledge-based hierarchies of firms

Carlo Altomonte
Gianmarco Ottaviano
Armando Rungi
Tommaso Sonno

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

30 YEARS CEP

Economic and Social Research Council

Abstract
Hierarchical differentiation is a cornerstone of the organizing process. In this paper, we do three things. First, exploiting a newly assembled dataset, we provide the first worldwide overview of the patterns of hierarchical differentiation across Business Groups (BGs), highlighting the co-existence of different hierarchical shapes. Second, we show how the different shapes can arise as optimal hierarchical structures in a knowledge-based model of BGs when subsidiaries' operations involve ubiquitous problem solving under parents' supervision. Three primitive characteristics of a BG determine its optimal choice of hierarchical structure: production efficiency and two dimensions of problem solving efficiency related to supervising knowledge creation and handling associated communication across subsidiaries. Third, we check the consistency of the model's predictions with the empirical patterns for Europe, the US, and the world. The model successfully passes the consistency test.

Key words: business groups, knowledge hierarchies, multinational enterprises, organisation of production
JEL: D23; L23; F23; L25; G34

This paper was produced as part of the Centre's Trade Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We are indebted with Luis Garicano for initial discussions, and Davide Suverato for constant interactions. We are also grateful to Pol Antràs, Lorenzo Caliendo, Paola Conconi, Esteban Rossi-Hansberg, Vincenzo Scrutinio, as well as seminar participants to the ERWIT-CEPR meeting in St. Gallen and the 2018 Princeton Summer Institute for useful comments. We thank Paolo De Rosa, Giulia Lo Forte, Francesco Losma, Marta Mojoli, Alessandro Pisa, and Davide Zufacchi for excellent research assistance.
Carlo Altomonte, Bocconi University and Baffi Carefin. Gianmarco Ottaviano, Bocconi University, Baffi Carefin, IGIER, CEPR and Centre for Economic Performance, London School of Economics. Armando Rungi, IMT Lucca. Tommaso Sonno, University of Bologna and Centre for Economic Performance, London School of Economics.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

A business group (BG) is an organizational form of economic activity in which at least two legally autonomous firms function as a single economic entity through hierarchical control: a parent company ('headquarter' - HQ) owns, directly or indirectly, the majority of the equity shares of at least one legally independent firm ('subsidiary'). Subsidiaries in the first layer are directly controlled by the HQ; subsidiaries in the second layer are controlled directly by subsidiaries in the first layer, and thus indirectly by the HQ; and so on. Under this definition, multinational enterprises (MNEs) are BGs with at least one subsidiary incorporated in a country different than the HQ's. A few numbers highlight how BGs are crucial components of the global economy. The world's largest corporations by consolidated revenue, as classified in the Fortune 500 list, are all organized as BGs. Around 60% of multinational value added is accounted for by few, large MNEs with more than 100 subsidiaries located in different countries. At least 75% of total US trade can be linked to firms operating in the US as parts of BGs (either as US headquarters or US subsidiaries of foreign groups). In the case of France around 65% of aggregate imports or exports can be attributed to firms that belong to BGs.[1]

A BG's hierarchical structure is defined by the number of subsidiaries, the number of layers and the distribution of subsidiaries across layers. According to UNCTAD (2016) data, top multinational groups feature on average around 300 subsidiaries, and more than 10 hierarchical layers of control. Yet, even BGs with very similar portfolios of activities may exhibit very different hierarchical structures. This is the case, for example, of two well-known BGs, General Motors and Mitsubishi, both operating in the automotive industry. In 2015, the two BGs were quite similar in terms of number of subsidiaries: 685 subsidiaries in 55 countries for General Motors and 528 subsidiaries in 47 countries for Mitsubishi. However, General Motors was organized in 10 hierarchical layers covering 26 sectors (3-digits NAICS 2002), while Mitsubishi was relatively flatter, with 5 hierarchical layers spanning a wider range of 44 sectors.

While hierarchical differentiation is arguably a cornerstone of the organizing process of BGs, it remains largely under-researched, especially in economics. Moreover, in the few existing studies scholars have tended to focus on pyramid-shaped hierarchies (such that a few subsidiaries control all other subsidiaries) under the assumption that those hierarchies are ubiquitous. This state of affairs mirrors the one in the studies on formal hierarchical structures and team performance (as e.g. in Wellman, Applegate, Harlow, & Johnston, 2020).

In this paper we do three things. First, we provide the first worldwide overview of the patterns of hierarchical differentiation across BGs, highlighting the coexistence of different hierarchical shapes. Second, we propose a parsimonious model to help making sense of such heterogeneity in observed hierarchies. Third, we bring the model to the data to assess its empirical relevance.

[1] See UNCTAD (2016), Bureau of Economic Analysis data, and (Altomonte & Rungi, 2013).

Our overview of BGs' hierarchies exploits the Orbis Ownership Database of Bureau van Dijk, which reports ownership control links among more than 200 million entities. To identify the BGs we apply the principle of the Ultimate Controlling Institution, as defined in the OECD FATS Statistics, through original algorithms that iteratively map the information provided on the bilateral ownership links across firms. Running these algorithms we retrieve about 2,9 million HQs controlling, directly or indirectly, more than 5 million subsidiaries worldwide in 2015. We then construct the hierarchical structure of each BG in terms of number of layers and number of subsidiaries on each layer with their location, industry and year of incorporation.[2]

The top layer of a BG's hierarchy consists of subsidiaries in which the HQ has a direct ownership share larger than 50%. We index this layer by 1 as a mnemonic for one degree of separation between the HQ and the corresponding affiliates. Descending from top to bottom, the next layer 2 (two degrees of separation) consists of subsidiaries in which the HQ has an ownership share larger than 50% by direct ownership, or through indirect ownership of affiliates at layer 1. The subsequent layer 3 (three degrees of separation) consists of subsidiaries in which the HQ has an ownership share larger than 50% by direct ownership or through indirect ownership of affiliates at layers 1 and 2, and so on. We then study the distribution of subsidiaries across layers and highlight a set of new stylized facts about BGs' hierarchies. In particular, in the wake of Wellman et al. (2020), as long as a BG has at least two layers of subsidiaries, we characterize its hierarchical differentiation in terms of the skewness of the distribution of its subsidiaries across layers. When skewness is negative, the distribution is left-skewed and subsidiaries are denser at lower layers (those with larger index) leading to a pyramid-shaped hierarchy. When skewness is positive, the distribution is right-skewed and subsidiaries are denser at higher layers (those with smaller index) leading to an inverse-pyramid-shaped hierarchy. The closer skewness is to zero, the more symmetric the distribution is, leading to a diamond-shaped hierarchy.

We find that, while unimodal, the distribution of BGs across skewness levels reveals substantial hierarchical differentiation. In our full global sample skewness ranges approximately from -2 to 3 with a rounded mean of 0.3 and standard deviation 0.9. Hence pyramid-shaped hierarchies are by no means the rule. Classifying hierarchies with skewness half a standard deviation below zero as pyramids, those with skewness half a standard deviation above zero as inverse pyramids, and the rest with skewness within half a standard deviation from zero as diamonds, these three classes contain around 20%, 50% and 30% of all BGs respectively. The dominance of inverse pyramids is more pronounced in BGs with more subsidiaries. There is also substantial cross-country heterogeneity: pyramids are less frequent in the US than in the EU, and even less so in Japan, where relatively 'flat' groups with an inverse-pyramidal shape are dominant, in line with our General Motors and Mitsubishi example. Diamonds are the dominant shape in the US. Despite this heterogeneity, two robust common patterns stand out when it comes to two salient characteristics of subsidiaries at different layers. On the one

[2] To validate our results, we have replicated the entire dataset of BGs and hierarchies through a different algorithm based on network analysis, obtaining very similar outcomes. Details are discussed in Section 2 as well as Appendix A.

hand, subsidiaries tend to be involved in increasingly routinizable activities within BGs, as their hierarchical distance from the parent increases: HQs keep more complex activities relatively 'close' in the chain of control. On the other hand, within BGs there is no correlation between the contractibility of subsidiaries' activities, given the industry in which they are involved and the countries where they are located, and their hierarchical position: variables that typically affect the make-or-buy decision of the HQ seem to be relatively independent of the hierarchical organization.

We explain these stylized facts in terms of a model in which hierarchical differentiation allows BGs to use knowledge efficiently, and to communicate it among subsidiaries in order to minimize the cost of using it as a production input. In the model, a parent owns the 'blueprints' of a large portfolio of off-the-shelf final products, and has to decide how many of these products to produce and how to organize their supply through production units (subsidiaries). The parent has exclusive knowledge of the production process of each blueprint. However, in order to turn these blueprints into actual production, a product-specific problem has to be solved. The problem comes in different versions: more complex production processes are harder to design, and thus require more effort to solve, but allow for production at lower marginal cost. Each subsidiary consists of a problem solver ('executive') and a team of producers ('employees') whose number depends on the amount of output. The executive receives the problem in the version chosen by the parent. If the executive solves the problem, the productivity of employees in her subsidiary is determined by the difficulty of the chosen version. If the executive does not solve the problem, employees in her subsidiary cannot produce and their productivity is zero. Solving more difficult versions of the problem requires higher ability that not all executives have. On top of adequate ability, to solve her subsidiary's problem the executive also needs supervision by the parent. Supervision can be direct, if the subsidiary is directly controlled by the parent, or indirect, through other executives placed in subsidiaries above hers in the hierarchy. The parent and, in turn, the executives operating in subsidiaries have only a limited amount of time they can devote to supervision, and the amount of time needed depends on the difficulty of the problem version to be solved as well as the supervisee's ability. Adding subsidiaries in a further layer of control expands the BG's problem solving ability, but it also entails a fixed cost. As a result, the parent faces a trade-off between (directly or indirectly) supervising several lower ability executives in the solution of easier problem versions and supervising few higher ability executives in the solution of more difficult problem versions.

Three primitive characteristics of the BG determine its optimal choice of hierarchical structure: production efficiency and two dimensions of problem solving efficiency, i.e. efficiency in supervising knowledge creation and efficiency in handling associated communication across subsidiaries. We find that a BG's number of layers ('hierarchical depth') increases with production efficiency, supervision efficiency and communication efficiency. Higher supervision efficiency and communication efficiency also lead to a larger number of subsidiaries per layer. Crucially, the hierarchical shape is determined by the comparison between problem solving efficiency and problem solving difficulty, as these affect the cost and benefits of knowledge creation. If solving more difficult problems gives a large boost to subsidiaries' profits and communication efficiency is high enough, the optimal hierarchical structure is a pyramid with few subsidiaries at a higher layer supervising many subsidiaries at a lower layer. If the boost is contained, a pyramid is still optimal for relatively high supervision efficiency and high communication efficiency. In contrast, an inverse pyramid (with many subsidiaries at a higher layer supervising few subsidiaries at a lower layer) is optimal with relatively low supervision efficiency and low communication efficiency, whereas a diamond is the optimal choice for intermediate supervision efficiency and intermediate communication efficiency. The higher supervision efficiency and communication efficiency are, the larger the pyramidal part of the diamond. Finally, independently from the shape of the hierarchy, the span of control of higher layer subsidiaries over lower layer subsidiaries (an inverse measure of 'hierarchical steepness') depends on the gap between the boost given by solving more difficult problems and the related communication costs: the closer they are to offsetting each other, the steeper the hierarchy. More steepness is associated with a smaller span of control in a pyramid, or in the pyramidal part of a diamond, but larger span of control in an inverse pyramid, or in the inverse-pyramidal part of a diamond.

We then go back to the data to check the consistency of the model's predictions with key empirical patterns. We proceed in two steps. First, we exploit the model's equilibrium equations to specify econometric regressions that allow us to obtain estimates for the parameters regulating the two dimensions of problem solving efficiency by leveraging the routinizability of activities (as a proxy of the difficulty of problem solving) and the number of subsidiaries across the hierarchical layers of a BG. Second, we use these estimates, together with the model's structure and other estimated parameters borrowed from the literature, to calibrate the parameters regulating production efficiency by targeting an additional moment of the data: the number of layers of a BG. In order to have enough variation across hierarchies to identify these parameters, we target BGs with at least 50 subsidiaries. This implies that we have to restrict the analysis to the European, the US and the full world samples as in the Japanese sample such BGs are relatively rare. The model successfully passes the consistency test. The results of the first step indicate that, while US BGs are less efficient than European BGs in supervising knowledge creation, they are more efficient in handling associated communication across subsidiaries. The results of the second step further suggest that US BGs are more efficient than European BGs also in production.

The paper speaks to several strands of literature. Hierarchical differentiation has been studied in the management literature, although often confined to emerging countries (Colpan, Hikino, & Lincoln, 2010; Khanna & Yafeh, 2007). The finance literature has traditionally emphasized the pyramidal structure of BGs built by a shareholder through a chain of equity ties as a way to achieve control of a firm using only a small cash flow stake (Berle & Means, 1932; Graham & Dodd, 1934). Recent works have shown how pyramids can obtain the same objective without relying on the separation of cash flow from voting rights (La Porta, Lopez-de Silanes, & Shleifer, 1999; Almeida & Wolfenzon, 2006). Another strand of literature has looked at tax motives behind BGs' hierarchies (Altshuler & Grubert, 2003; Desai et al., 2004; Mintz & Weichenrieder, 2010; Lewellen & Robinson, 2013) while studies on how firm organizational structure relates to corporate strategy, finance and economics outcomes have highlighted the resource flexibility associated with internal capital and labor markets (Cestone & Fumagalli, 2005; Belenzon & Schankerman, 2013; Cestone et al., 2016; Boutin et al., 2013). These works either do not emphasize hierarchical differentiation or, when they do, they operationalize the hierarchical structure in terms of the number of layers of control, or the extent to which control is concentrated in the HQ. As a result, the possible relevance of alternative hierarchical shapes to pyramids has stayed under the radar.

Our model explains the observed heterogeneity in BGs' hierarchical structure borrowing concepts from knowledge-based models of employees' hierarchies with firms (Garicano, 2000; Garicano & Wu, 2012; Garicano & Rossi-Hansberg, 2015). In these models hierarchical differentiation allows firms to use and communicate knowledge efficiently. In particular, a hierarchical structure arises endogenously as the optimal way to organize problem solving when this requires matching problems of heterogeneous difficulty with employees who know how to solve them. By adding layers of problem solvers of different ability to deal with problems of different difficulty, the organization increases problem solvers' specialization and the utilization rate of their specific know-how, thus economizing on knowledge acquisition at the cost of increasing the communication required. The trade-off between communication and knowledge acquisition costs determines the structure of the hierarchy, which in these models is given by the span of control of problem solvers and the number of layers in the organization. The 'hierarchical shape' is always a pyramid. This is were we innovate.

In terms of data sources, the Orbis-Bureau van Dijk data we rely on have already been used in the literature to study BGs in terms of innovation (Belenzon & Berkovitz, 2010), the international transmission of shocks (Cravino & Levchenko, 2017), the effect of managerial culture on firm boundaries (Gorodnichenko et al., 2021). Other studies have used similar data to explore the boundaries of the firms belonging to BGs (e.g. Alfaro & Charlton, 2009; Alfaro, 2017) with respect to BGs' external suppliers, or have started to exploit the Ownership Database of Orbis - Bureau Van Djik to map BGs through the notion of control.[3] Belenzon et al. (2019) uses Bureau Van Djik data on a subset of European BGs to investigate the relationship between BG structure and subsidiary autonomy, finding that the latter increases with the number of intermediate layers between the subsidiary and the HQ. Altomonte and Rungi (2013) use ownership information in a cross-sectional analysis of BGs across countries in 2010. Rungi et al. (2017) adopt a network framework to identify BG-like structures as a function of direct and indirect corporate control, uncovering a strong concentration of corporate power (less than 1% of parent companies has more than 100 subsidiaries). Grosskurth (2019) uses Orbis data from 2000 to 2018 to map the development of the global networks of multinational BGs and their core components over time. Although he uses the number of controlled companies to identify a company's importance within a BG, hierarchical shapes are not the focus of his analysis. Sonno (2017)

[3] Some of the studies on external suppliers rely on data sourced from Dun&Bradstreet (D&B), which is one of the different sources now integrated in the Orbis Ownership database.

creates a world-wide panel of BGs' structures over the period from 2007 to 2018, and uses these data to establish a causal link between multinational activities and episodes of conflict in developing countries. Aminadav and Papaioannou (2020) use ownership data to investigate ownership concentration and types of control across continents. Eppinger and Kukharskyy (2020) compare ownership shares across half a million firm pairs worldwide, including domestic and cross-border ownership links, and find that firms choose higher ownership shares in subsidiaries located in countries with better contracting institutions. We show that our results hold also after controlling for the sources of variation exploited in these studies.

The rest of the paper is structured as follows: Section 2 presents the data sources, discusses the construction of the dataset, and describes the patterns of hierarchical differentiation across BGs. Section 3 develops the knowledge-based model of BGs' hierarchical differentiation. Section 4 checks the consistency of some of its key predictions with the empirical patterns observed in our dataset. Section 5 concludes.

2 Hierarchical Differentiation

We define a Business Group (BG) as a collection of at least two legally autonomous firms functioning as a single economic entity through a common source of hierarchical control. Hierarchy in control implies that a parent company ('headquarters', or simply HQ) owns, partially or totally, the equity shares of a second legally independent firm. Moreover, the HQ may directly own the equity shares of a third firm, thus placing it at the same (first) layer as the second firm in a 'flat' hierarchy. Alternatively, the second firm could control the third one, thus placing it at the second layer of a deeper hierarchy in the BG. In this case the HQ has indirect control of the third firm through the second. Multinational enterprises (MNEs) are a subset of BGs with at least one subsidiary abroad.[4]

2.1 Dataset

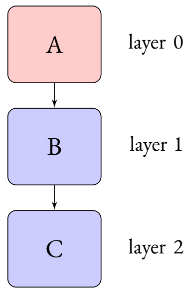
We start from the Historical Ownership Database by Bureau Van Djik. This dataset provides for each company information on all its shareholders, corporate or non-corporate (e.g. individuals or partnerships), and identifies direct and indirect voting rights.[5] We focus on year 2015 and, in line with

[4]When a single legal entity operates more than one productive plant (multi-plant firms), or is organized internally through multiple divisions, those plants or divisions are sometimes considered by Orbis as different entities, and are flagged as branches. Since the focus of this work is on the organizational choices of BGs in terms of the control hierarchy, we focus exclusively on the legally independent entities and drop branches from our dataset.

[5]See Appendix A for details on the data. In general, corporate control can be derived by a direct, indirect or consolidated concentration of voting rights (Faccio & Lang, 2002; Chapelle & Szafarz, 2007; Del Prete & Rungi, 2017). For example, a company H owns 100% of the voting rights (VR) in companies A and B. Company A, also owns 70% of the VR in X and 30% of those in Y. Finally company B owns another 40% of the VR of Y. In this case H is able to control, directly or indirectly, more than 50% of the voting rights in all the other companies: in fact, H enjoys direct control of A and B, indirect control of X trough A and consolidated control of Y trough A and B. This is known as the principle of the Ultimate Controlling Institution in the OECD FATS Statistics (or Ultimate Beneficial Owner in UNCTAD data).

Figure 1: From ownership links to business groups

Subsidiary	Shareholder	Relation
B	A	GUO
B	A	ISH
C	A	GUO
C	B	ISH



Notes: The table on the left shows a simplified example of four links in the Orbis Historical Ownership Database. From the links we can reconstruct the group in the diagram on the right. We observe that A is the parent (GUO) of both B and C, and that A is the direct owner (ISH) of B while B is the direct owner of C. Therefore, we can conclude that A owns C indirectly trough B.

international accounting standards, we limit our definition of control to the case in which a corporate shareholder (i.e. we exclude non-corporate entities) can command the majority (i.e. strictly more than 50%) of the voting rights in another company. As the latter company is majority owned by the former, we call it a 'subsidiary' of the corporate shareholder.[6]

The chosen notion of majority control allows us to identify, for each company in the dataset, the presence (if existing) of a unique corporate 'Global Ultimate Owner' (GUO), that is a controlling parent company. The dataset also reports, for each company, information on the 'Immediate shareholder' (ISH), that is, the corporate entity that directly controls (at least 50% of) the company under analysis. Starting from the two relations of GUO and ISH, we follow Sonno (2017) and create an algorithm that first identifies the boundaries of a BG as the set of all subsidiaries having the same GUO, and then retrieves the hierarchical distance of a subsidiary from its parent. In particular, we first identify the ISH of each subsidiary, and then assign hierarchical layer 0 to the parent (GUO), hierarchical layer 1 to the subsidiaries directly controlled by the parent (i.e. one degree of separation), and analogously layer $\ell + 1$ to the subsidiaries with ISH at generic layer ℓ (i.e ℓ degrees of separation). In the case where a subsidiary is controlled through one or more subsidiaries, we identify it as indirectly controlled by the parent. Layers can be interpreted as a measure of hierarchical distance from the HQ: the higher the number assigned to each layer, the higher the hierarchical distance. The number assigned to the most distant layer L defines the depth of the hierarchy with layers indexed $\ell = 0, ..., L$. Figure 1 provides a simple illustrative example of the construction of the hierarchy of a BG starting from the ownership links. It identifies A as a parent company, because the company is reported both as a GUO and as the

[6]The notion of control stems from OECD (2005), Guidelines for Multinational Enterprises; UNCTAD (2009), Training Manual on Statistics for FDI and the Operations of TNCs; Eurostat (2007), Reccomendations Manual on the Production of Foreign Affiliates Statistics (FATS). This notion of control neglects cases in which affiliates are *de facto* controlled through minority ownership, as well as cases in which control derives from market advantage (e.g. monopsony) or government regulations (e.g. 'golden share'). This notion of control applies equally to domestic and multinational business groups, and it allows for a straightforward comparison with official statistics, as it is commonly used for foreign subsidiaries (Eurostat or OECD FATS) and for international taxation (IAS, IFRS).

Table 1: Sample size

Number of groups	2,901,466
Number of groups with NAICS	2,192,303
Number of garents with only one sub	2,182,934
Number of subs	5,676,289
Number of subs with layer	5,653,433
Number of subs with NAICS	4,580,899
Number of subs with routinizability	4,550,351
Number of subs with contractibility	554,510

Notes: Breakdown of the sample size with details on the number of observations for which we find layer, NAICS sector, routinizability and contractibility.

ISH of company B, which in turn can be located at hierarchical layer $\ell = 1$. Subsidiary C, instead, can be located at layer $\ell = 2$ of the BG with company A as parent given that its ISH is a company located at layer 1 in that BG. The depth of the BG's hierarchy is $L = 2$.

This procedure generates a dataset of 2,901,466 parent companies controlling 5,676,289 subsidiaries in 2015. For our analysis we also require additional details on the type of ownership links, which might yield different subsamples of the dataset depending on the information available in the data. Using the Orbis dataset, also produced by Bureau Van Djik, we merge information on the country of incorporation of each company, in order to distinguish international links from domestic ones, and to map the geographic distribution of the BGs. In order to analyse the hierarchical layer distribution of subsidiaries, we also require the exact position in the hierarchical layer to be computable. These constraints on data availability slightly reduce the number of subsidiaries to 5,653,433. Moreover, to characterize the BGs we need information on the industry (at 3-Digits NAICS revision 2002) of each entity in our database. Orbis does not contain complete information on the industry of activity for all entities, and hence we lose some 20% of observations in the match with NAICS 3-digit sectors. This is the sample we use except for descriptive statistics, which are based on the broader sample.[7] A breakdown of the different sample compositions can be found in Table 1.

Some combinations of cross-holding of shareholders might yield an ambiguous position of subsidiaries in the hierarchical layer of control. Yet, thanks to the algorithm's efficiency, we are able to unequivocally assign a layer to approximately 99.6% of the subsidiaries. Nevertheless, to evaluate the robustness of our algorithm in identifying BGs and their hierarchical structure, we also experiment with an alternative approach embedded in network theory. The method has been developed by Rungi et al. (2017) and uses data from the Ownership section of the Orbis Database, which is a different section with respect to the Historical Ownership Database we use as our main starting point, and thus does not rely on the bilateral ownership links at the source of the aforementioned ambiguity. For each company recorded, Orbis collects all information available on its ownership structure. We arrange this infor-

[7] Going from the broader sample to the restricted sample we use does not induce any self-selection in terms of relevant characteristics, in particular hierarchical differentiation and country of incorporation.

Table 2: Geographic coverage

Region or Country	Business Groups All	%	Multinational	%	Subsidiaries All	%	Foreign	%
Africa	6,698	0.24	4,852	2.23	41,928	0.74	26,175	2.55
Japan	5,357	0.19	4,069	1.87	27844	0.49	4,345	0.42
Other Asia	81,855	2.90	23767	10.92	367,740	6.50	173,615	16.94
Australia	83,677	2.96	3,789	1.74	188,346	3.33	28,329	2.76
EU 28	718,057	25.43	115,611	53.10	1,899,408	33.56	506,157	49.38
Other Europe	62,051	2.20	16,258	7.47	154,332	2.73	37,880	3.70
Latin America	31,892	1.13	19,183	8.81	95,995	1.70	65,223	6.36
Russia	53,472	1.89	1,051	0.48	159,973	2.83	53,632	5.23
USA	1,743,710	61.74	22,209	10.20	2,628,317	46.44	99,213	9.68
Rest of the world	37,423	1.33	6,923	3.18	96,042	1.70	30,510	2.98
Total	2,824,192	100.00	217,712	100.00	5,659,925	100.00	1,025,079	100.00
Country not assigned	77,274				16,364			

Notes: The table details the number of observations of subsidiaries and business groups in a list of regions or countries. It also specifies the number of business groups that control at least one subsidiary abroad (multinational), and that of subsidiaries that are controlled by a foreign parents. For regional aggregation we have used the online version of the United Nations publication Standard Country or Area Codes for Statistical Use. Under the label 'Country not assigned' we classify the observations for which Orbis does not identify a country. Percentages are computed excluding observations located in these countries.

mation in matrix form, and then launch the authors' algorithm, simulating a voting rule in presence of interlocking assemblies of shareholders. The algorithm detects concentrations of voting rights that allow for strategic coordination. This way we are able to delimit the boundaries of each BG, to identify the headquarters, and to assign each subsidiary to its hierarchical position.[8] Reassuringly, even though the two approaches start from different initial datasets, and use different methods to identify the structure of BGs, they both produce remarkably similar results.[9]

2.2 Descriptive Statistics

Table 2 describes the geographic coverage of the dataset, which spans all countries in the world. We report the number of BGs (which is equivalent to the number of HQs) and the number of subsidiaries located in each country or geographical area. In addition, we specify the number of BGs that are multinational (i.e. the number of parents that control also subsidiaries abroad), and the number of subsidiaries that are ultimately controlled by a foreign entity or a domestic parent. Most parents and subsidiaries recorded in the data are incorporated in the US or in Europe. Moreover, unsurprisingly, Europe as a geographical area displays a larger share of multinational groups given that around half of the subsidiaries of European BGs are located in predominantly European foreign countries.

[8] A peculiarity of Rungi et al. (2017) is that, starting from all the ownership paths through which a management decision could run, the approach correctly manages cases of cross-holdings, ownership cycles and consolidation of voting rights across otherwise fragmented webs of equity stakes. In this framework, the hierarchical distance between a parent company and any of its subsidiaries (layer) is thus defined as the shortest ownership path in the network that connects them through.

[9] Appendix A presents the two approaches in more detail and compares them in terms of descriptive statistics, while the Online Appendix B replicates our main stylized facts on the sample of BGs identified through the Rungi et al. (2017) algorithm.

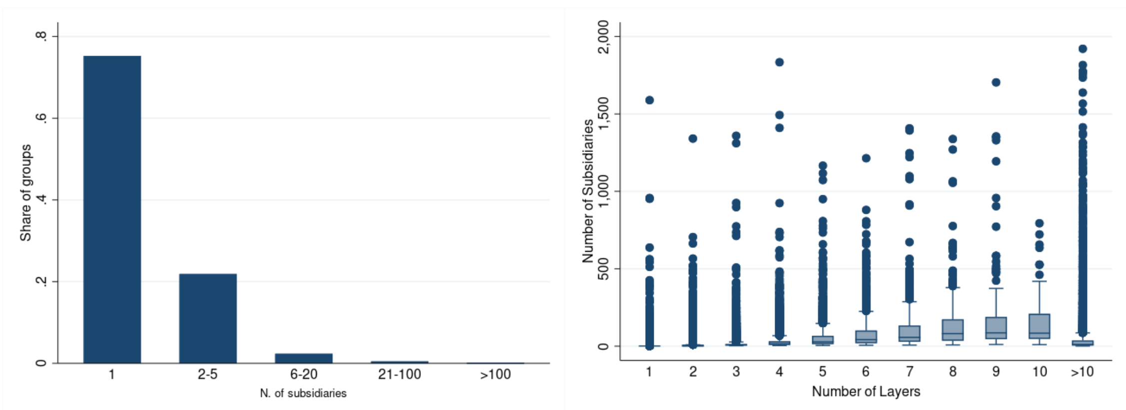
Table 3: Hierarchical distance

Layer	Domestic subsidiaries	%	Foreign subsidiaries	%	All subsidiaries	%
1	2,000,151	76.13	357,147	42.39	2,357,298	67.92
2	417,260	15.88	233,214	27.68	650,474	18.74
3	129,417	4.93	121,626	14.44	251,043	7.23
4	46,818	1.78	60,661	7.2	107,479	3.1
5	18,551	0.71	31,742	3.77	50,293	1.45
6	7,571	0.29	17,170	2.04	24,741	0.71
7	3,878	0.15	9,726	1.15	13,604	0.39
8	1,867	0.07	4,593	0.55	6,460	0.19
9	1,128	0.04	2,563	0.3	3,691	0.11
10	582	0.02	1,515	0.18	2,097	0.06
> 10	735	0.03	2,608	0.31	3,343	0.10
Total	2,627,223	100.00	842,565	100.00	3,470,523	100.00

Notes: The table shows the number of subsidiaries per layer, distinguishing between domestic and foreign subsidiaries. It excludes business groups with only one subsidiary.

Figure 2: Hierarchical description

(a) (b)



Notes: The left panel shows the distribution of groups' dimension in terms of number of subsidiaries. The right panel presents eleven boxplots of groups' dimension conditional on the number of layers their hierarchies display. We have excluded the 13 business groups with more than 2,000 subsidiaries so as to make the figure more legible.

In Table 3 we start analyzing the hierarchical structure of the groups, as determined by the number of subsidiaries, the number of layers, and the distribution of subsidiaries across layers. The vast majority of subsidiaries (68%) appear to be located in the first layer of control, although this is a feature especially prevalent in domestic subsidiaries (76% of them) rather than foreign ones (with 42% and 28% of them at the first and second layers respectively). On average, the first four layers of control contain 97% of all the subsidiaries, although again multinational groups appear to be distributed across relatively deeper structures: 91% in the first four layers, and 97% in the first six layers. The skewness in the distribution of control hierarchies is also apparent when we look at the average number of subsidiaries per group. The left panel of Figure 2 shows that around 75% of groups in our data have very simple structures, with one parent controlling only one subsidiary whereas around 20% of the groups have between 2 and 5 subsidiaries. On the other tail of the distribution, around 0.1% of the parents in the sample have more than 100 subsidiaries each. The right panel of Figure 2 shows that, on average,

Table 4: Number of affiliates per layer across business groups

	Maximum layer of the BG										
Avg. subs per layer	1	2	3	4	5	6	7	8	9	10	> 10
1	1.37	3.26	6.24	11.35	20.76	30.97	41.74	45.91	47.57	68.38	81.75
2		2.18	6.13	11.44	19.7	29.01	38.75	43.86	62.43	40.15	63.72
3			2.8	8.36	16.52	23.65	34.91	40.39	45.03	44.08	54.14
4				3.06	9.32	17.17	26.21	30.09	36.17	28.34	43.35
5					3.32	9.24	17.27	21.95	27.94	18.64	37.97
6						3.36	9.32	14.41	19.74	16.09	32.62
7							3.8	8.11	13.58	14.03	31.37
8								3.05	8.57	10.46	19.32
9									3.4	7.54	16.68
10										2.64	13.46
> 10											24.22
N. of BG	2,739,149	126,168	24,172	6,811	2,571	1,157	618	350	207	91	138

Notes: The table shows the average number of subsidiaries per layer for business groups having different maximum layers. If business groups with only one subsidiary were excluded, the mean number of affiliates in the first layer for business groups with only one layer would be 2.81.

groups organized in a larger number of hierarchical layers (reported on the horizontal axis) tend to have more subsidiaries (see the central bar of the box plot, i.e. the median in the distribution of the number of subsidiaries). Yet, as shown by the outliers of the box plots, there exist relevant exceptions: some groups with only few layers of control have more than 100 subsidiaries, while some groups with fewer subsidiaries arrange them across several layers.

Table 4, looks at the average number of subsidiaries per layer, across BGs characterized by different numbers of layers. BGs organized with one layer of control have on average 1.37 subsidiaries on their first (and unique) layer. Groups organized with two layers tend to have 3.26 subsidiaries on average in the first layer, and 2.18 in the second, and so on. Most naturally, as BGs organized in more layers tend to have more subsidiaries, the average number of subsidiaries per layer increases as we move along the rows in the table, i.e. along BGs with more layers of control.

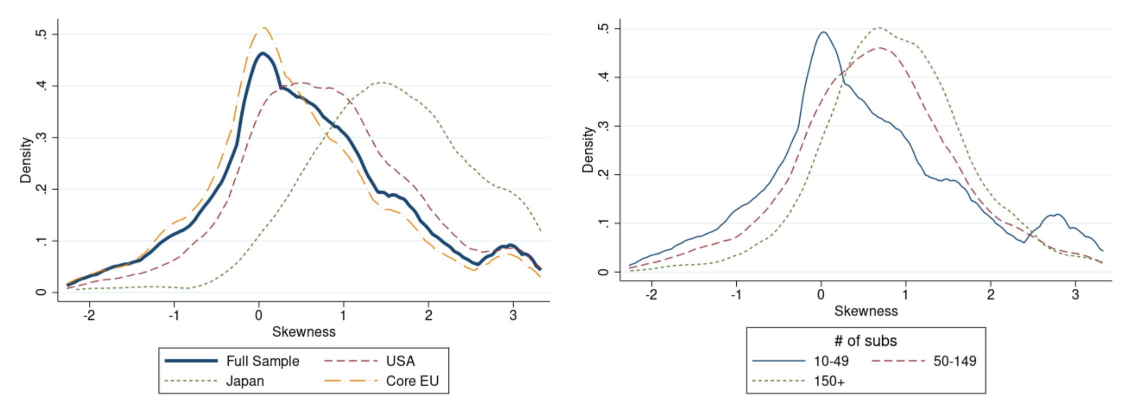
2.3 Stylized Facts

In the wake of Wellman et al. (2020), as long as a BG has at least two layers of subsidiaries, we characterize its hierarchical differentiation in terms of the skewness of the distribution of its subsidiaries across layers. When skewness is negative, the distribution is left-skewed and subsidiaries are denser at lower layers (those with larger ℓ) leading to a pyramid-shaped hierarchy. When skewness is positive, the distribution is right-skewed and subsidiaries are denser at higher layers (those with smaller ℓ) leading to an inverse-pyramid-shaped hierarchy. The closer skewness is to zero, the more symmetric the distribution is, leading to a diamond-shaped hierarchy.

First, according to Figure 3, the distribution of BGs across skewness levels is unimodal, with substantial hierarchical differentiation. In particular, in the full global sample skewness ranges continuously from -2.26 to 3.32 with mean 0.19 and standard deviation 0.87.

Figure 3: Skewness distribution

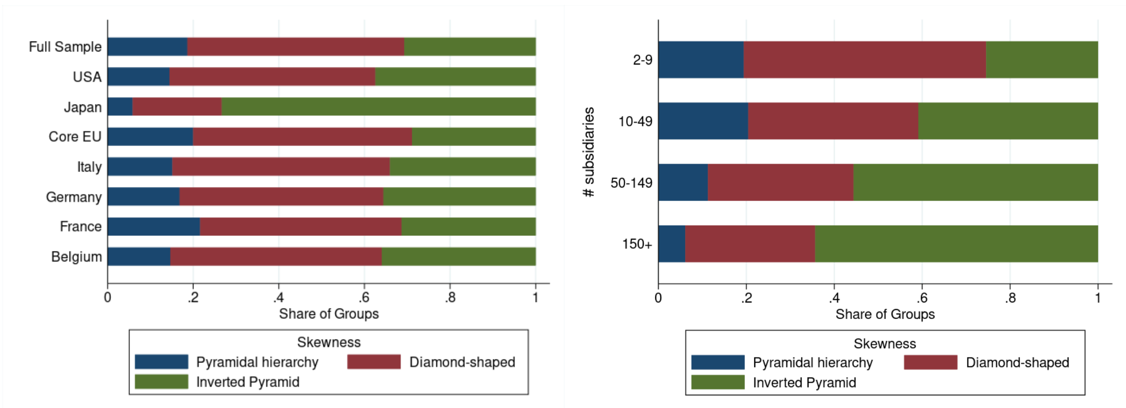
(a) (b)



Notes: The figure shows the distribution of the skewness of BGs' hierarchies. In the left panel BGs are grouped by geographical area, while in the right panel they are grouped by size. BGs with less than 2 layers or 10 subsidiaries are dropped because their skewness can only take few values and thus alter the densities.

Figure 4: Skewness bars

(a) (b)

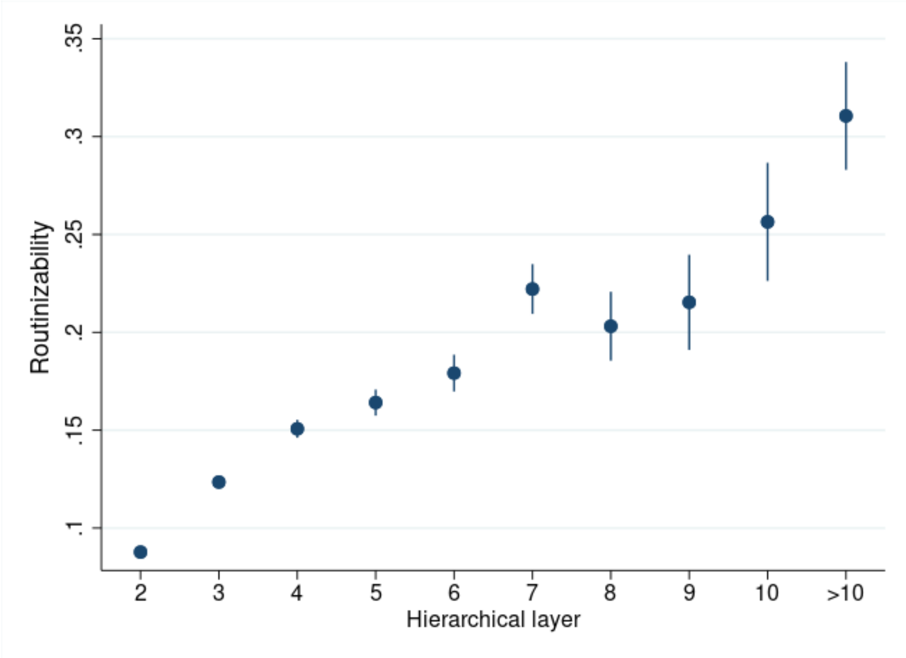


Notes: The figure shows the prevalence of the three types of hierarchical structures. In the left panel BGs are grouped by geographical area, while in the right panel they are grouped by size. BGs with less than 2 layers are excluded because their skewness can only take value 0 and thus alter the graphs.

Second, pyramid-shaped hierarchies (those with negative skewness) are by no means the rule. This is most readily seen by classifying hierarchies with skewness half a standard deviation (0.53 in our sample) below zero as pyramids, those with skewness half a standard deviation above zero as inverse pyramids, and the rest with skewness within half a standard deviation from zero as diamonds. Then, the left panel of Figure 4 shows that in the full sample those three classes contain around 20%, 50% and 30% of all BGs respectively.

Third, the right panel of Figure 4 highlights that the dominance of inverse pyramids is more pronounced in BGs with more subsidiaries. For example, the hierarchies of BGs with $50 - 149$ subsidiaries are around 10% pyramids, 30% diamonds and 60% inverse pyramids; those with more than 150 subsidiaries are around 5% pyramids, 20% diamonds and 75% inverse pyramids. These patterns are reflected in the mean and the standard deviation of skewness (the right panel of Figure 3). In BGs

Figure 5: Routinizability over hierarchies



Notes: The graph shows the coefficients and 95% confidence intervals obtained from a regression of the routinizability index of each subsidiary on the hierarchical layer of the same subsidiaries, including group and host country FE, and using robust standard errors. Regression results are reported in Appendix Table A4.

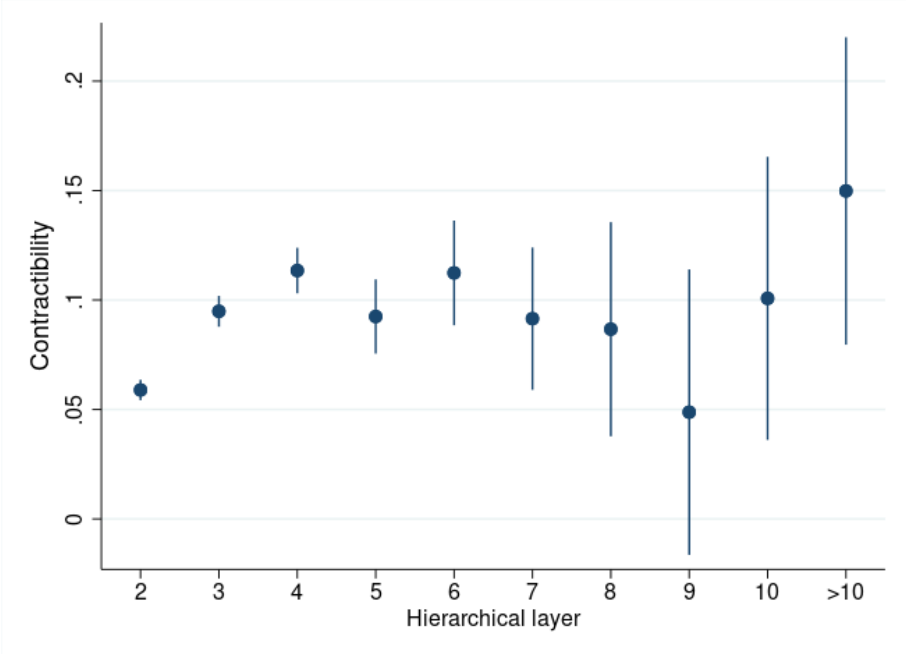
with $50 - 149$ subsidiaries mean skewness equals 0.65 with standard deviation 0.96, whereas in BGs with more than 150 subsidiaries mean skewness equals 0.88 with standard deviation 0.94, compared with 0.19 and 0.87 in the global sample.[10]

Fourth, considering again the left panel of Figure 4, we see that there is substantial cross-country heterogeneity. Pyramids are less frequent in the US (18%), and even less so in Japan (6%), than in the full global sample. However, diamonds are the dominant shape in the US (45%) while in Japan (17%) the lion share goes to inverse pyramids (72%). In core European countries BGs exhibit a more balanced distribution across shapes, roughly in line with the full global sample, though the frequency of pyramids is slightly higher than in the latter. These patterns are reflected in the mean and the standard deviation of skewness. In the US and the European mean skewness equals 0.19 and 0.14 with standard deviation 0.87 and 0.85, whereas in Japan mean skewness equals 1.15 with standard deviation 1.10. Interestingly, hierarchical differentiation varies also across European countries, with France featuring a larger frequency of pyramids than Germany, Italy and Belgium.

Despite all this heterogeneity two robust common patterns stand out when it comes to two salient characteristics of subsidiaries at different layers. On the one hand, Figure 5 highlights that, within

[10] Several subsidiaries in the first layer could be financial shells or regional HQs. To check the sensitivity of the skewness patterns reported in Figure 4 to the thickness of the first layer, we have restricted our sample to BGs with at least three layers, and then excluded from our computations the subsidiaries in the first layer. While we observe a decrease in pyramidal shapes and an increase in both inverse pyramids and diamonds, the country and size patterns in Figure 4 are confirmed no matter whether we include or exclude the subsidiaries in the first layer.

Figure 6: Contractibility over hierarchies



Notes: The graph shows the coefficients and 95% confidence intervals obtained from a regression of the contractibility index of each subsidiary (liberal definition, using inputs that are neither sold on an organized exchange nor reference priced) on the hierarchical layer of the same subsidiaries, including group FE, and using robust standard errors. Host-country FE are omitted because contractibility incorporates rule of law that is country-specific. Regression results are reported in Appendix Table A5.

BGs, subsidiaries tend to be involved in increasingly routinizable activities as their hierarchical distance from the parent increases.[11] The figure plots the coefficients with the corresponding 95% confidence intervals estimated regressing for each subsidiary in the sample, its index of routinizability on a set of eleven layer dummies (for $\ell = 1, ..., 10$ and $\ell > 10$) plus parent and country fixed effects. The index is constructed exploting the specific question Q.25b in the survey questionnaire developed for the Princeton Data Improvement Initiative (PDII) as described in Blinder and Krueger (2009, 2013). This question asks respondents: "How much of your workday involves carrying outshort, repetitive tasks? Would you say ..." There are four possible answers: "Almost all the time - 1", "More than half of the time - 2", "Less than half of the time - 3", "Almost none of the time - 4". We compute the mean answer by the 3-digit NAICS 2002 sector in which respondents are employed, and use it as an index of routinizability of the industry in which a subsidiary is active. For easier interpretation we reverse the original ranking to associate a higher value of the index with a higher degree of routinizability. We estimate routinizability for 4,550,351 subsidiaries in our sample, obtaining a mean of 2.44, a standard deviation of 0.47, a minimum of 1.33 and a maximum of 4.00. Figure 5 thus shows that the routinizability of subsidiaries' activities increase with their hierarchical distance from the parent.

On the other hand, Figure 6 shows that, within BGs, there is no correlation between the contractibil-

[11] Table A4 in Appendix B reports the econometric model and the regression results behind Figure 5.

ity of subsidiaries' activities and their hierarchical layers.[12] We use the contractibility index developed by Nunn (2007). This index combines a measure of the quality of contract enforcement for each country with a measure of contractual intensity for each final good in the manufacturing sector (i.e., the share of inputs requiring relationship-specific investments).[13] Contractibility varies by country and industry: a high value of contractibility is associated with a high level of judicial quality combined with a high level of contract intensity. We define the contractibility of a subsidiary's activities as the contractibility index associated with its industry and its country of incorporation. Then, for each subsidiary, we regress the contractibility of its activities on its layer within its group hierarchy plus parent and country fixed effects. Figure 6 plots the coefficients of this regression with their 95% confidence intervals and does not reveal any significant pattern of correlation between contractibility and hierarchical distance from the parent.

3 Knowledge-Based Hierarchies

Our dataset reveals rich patterns of hierarchical differentiation that are hard to rationalize in terms of the existing approaches to BGs discussed in the introduction, as most of them have not been devised to explain any specific shape, or have focused on the pyramid as the only relevant shape. In particular, any approach based on the idea that hierarchies help the ultimate owner ***[to]*** mechanically multiply its span of control are inherently bound to predict the emergence of pyramids rather **than** inverse pyramids. We now show that alternative shapes can arise as optimal hierarchical structures in a parsimonious knowledge-based model of BGs when subsidiaries operations involve ubiquitous problem solving under parents' supervision.

Following Garicano and Rossi-Hansberg (2015), we conceptualize knowledge as the ability to solve the problems that naturally arise in any production process. In a BG their solution requires supervision time by the parents, which comes in limited amount. Some problems are harder to solve than others. Their solution has higher upside potential in terms of the BG's performance, but it also absorbs more supervision time. The BG's optimal hierarchy arises from the parent's profit-maximizing allocation of supervision time across subsidiaries solving problems of different difficulty. While the parent's efficiency in production determines the number of subsidiaries of a BG, two primitive parameters play a crucial role in determining the optimal hierarchy: the parent's efficiency in supervising knowledge creation, and its efficiency in handling the associated communication across subsidiaries.

[12] Table A5 in Appendix B reports the econometric model and the regression results behind Figure 6.

[13] Specifically, we measure contractibility as $RoL \times z(a, b)$, where RoL is the 'Rule of Law' index from the Worldwide Governance Indicators (Group, Kaufmann, Kraay, and Mastruzzi (2010)), and $z(a, b)$ is a variable from Nunn (2007), which measures the level of contract intensity of a specific industry; a represents the Rauch (1999) industry definition (liberal or conservative); b represents the methodology adopted to define relationship-specific inputs. The latter in turn can vary between inputs that are neither sold on an organized exchange nor reference priced, and those that are not sold on an organized exchange but are reference priced (see Nunn (2007) for details). Hence, the final measure of contractibility encompasses four different indexes: the two measures of proportion of intermediate inputs and the two estimates by Rauch (1999)

3.1 Knowledge Creation

Consider an integrated global market in which a large number of BGs compete by selling imperfectly substitutable products. Each BG can produce only a countable number of products, which is a subset of measure zero of the total mass of products available in the market. This supports a monopolistic competition outcome even though BGs are multi-product suppliers, as in Mayer, Melitz, and Ottaviano (2014). For each of its products a BG faces isoelastic demand $y = Ap^{-\sigma}$, where y is ouput, p is price, $\sigma > 1$ is the cross-price as well as the own-price demand elasticity, and $A > 0$ is a demand shifter. Both σ and A capture market characteristics that are common to all products, and thus all BGs.

Each BG consists of a parent and an endogenous number of subsidiaries. The parent owns the 'blueprints' of a large portfolio of off-the-shelf final products and has to decide how many of these products to produce, and how to organize their supply through subsidiaries. The parent has exclusive knowledge of the production possibilities of each blueprint but, in order to turn them into actual production, a product-specific problem has to be solved. The problem comes in different versions, indexed $\wp = 1, ..., P$ in decreasing order of difficulty, and the parent decides which version to tackle in any given subsidiary. A subsidiary consists of a problem solver ('executive') and a team of producers ('employees') whose number depends on the amount of output. Accordingly, problem solving and production entail a fixed and a variable costs respectively.

The executive is assigned the problem by the parent in the version the parent decides. If the executive solves the problem, the productivity of employees in her subsidiary is determined by the difficulty of the chosen version. If she solves version $\wp$, her employees' productivity is $\omega\theta_{\wp}$, where $\omega > 0$ is a parent-specific component while $\theta_{\wp} = e^{-\theta\wp}$, with $\theta > 0$, is a subsidiary-specific component. If the executive does not solve the problem, employees in her subsidiary cannot produce and their productivity is zero. The subsidiary-specific productivity component is determined by the problem version $\wp$ chosen by the parent, and by the rate θ at which the solutions of more difficult problem versions (with lower $\wp$) are associated with higher employees' productivity, but at which such solutions are also harder to find. As σ and A, the problem solving difficulty rate θ captures market characteristics that are common to all products, and thus to all BGs. For instance, problem versions can be interpreted in terms of routinizability of the adopted technology: less standard production processes are harder to design, but they typically allow for production at lower marginal cost.

All employees have the same skills. Their wage is normalized to 1 and at this wage their supply is infinitely elastic. There are, instead, different ability types of executives. Solving more difficult versions requires higher ability that not all executives have. Executive types are indexed $\wp = 1, ..., P$ in decreasing order of ability, so that $\wp$ refers indifferently to the difficulty of a problem's version and to the ability type of the executives who can solve it. Executives have only a limited amount of time they can devote to problem solving. In this amount of time an executive of ability $\wp$ can solve at most one problem of corresponding or lower difficulty (i.e. indexed $\wp$ or above). The ability differential between executives is reflected in different hiring costs, with $w\theta_\wp$ denoting the fixed cost of hiring an executive of type $\wp$ whose problem-solving ability allows employees in her subsidiary to achieve productivity $\omega\theta_\wp$. At $w\theta_\wp$ the supply of executives of ability $\wp$ is also infinitely elastic. Employees' wage and executives' hiring cost per efficiency unit w are market characteristics.

On top of adequate ability, to solve her subsidiary's problem the executive also needs supervision by the parent, which can be direct or indirect through executives in other subsidiaries. In the latter case, however, the executive cannot be supervised by executives of equal or lower ability so that indirect supervision by the parent of an executive of ability $\wp$ must go through executives of higher ability (i.e. indexed $\wp - 1$ or below). Supervision is time consuming for the supervisor and the amount of time needed depends on the difficulty of the problem version to be solved as well as the supervised executive's ability. Specifically, in order to solve a problem version of difficulty $\wp$, an executive of ability $\wp$ requires $\varphi_\wp\theta_\wp$ units of supervision time where $\varphi_\wp = e^{\varphi\wp}$, with $\varphi > 0$, is a communication cost that inversely captures efficiency in generating the required information flow. Accordingly, φ is the rate at which communication becomes more efficient as the supervised executive's ability rises (i.e. $\wp$ falls). Taking stock, the chosen functional form for $\varphi_\wp$ implies that the higher is the supervisee's ability (smaller $\wp$), the lower is the communication cost (smaller $\varphi_\wp$). The multiplicative form for supervision time $\varphi_\wp\theta_\wp$ implies, while supervising the solution of more difficult versions absorbs more time (larger $\theta_\wp$), less time is needed for communication (smaller $\varphi_\wp$) as more difficult versions can be tackled only by higher ability supervisees. Given $\varphi_\wp\theta_\wp = e^{(\varphi-\theta)\wp}$, the former aspect dominates when the rate at which communication efficiency increases with executives' ability is smaller than the rate at which more able executives improve their employees' productivity by solving more difficult problems ($\varphi < \theta$). Vice versa, the latter aspect prevails when the opposite holds ($\varphi > \theta$).

The amount of effective supervision time is the same for the parent and the executives. It is equal to $\tau_\wp = e^\tau$ for all $\wp$'s, where τ measures efficiency in supervising knowledge creation. Supervision is the only activity of the parent, hence e^τ is its total amount of time available. Differently, for a executive e^τ is extra time in addition to the amount of time she has for problem solving. For simplicity, the executive's supervision and problem-solving amounts of time are not substitutable.

3.2 Optimal Hierarchy

We are interested in understanding how BGs' organizational structures vary as a function of the three parameters capturing their efficiency in production (ω), in supervision (τ) and communication (φ). A BG's structure is defined by the number of subsidiaries (hence of executives), the output level (hence the number of employees) of each subsidiary, the ability of executives in all subsidiaries, and the way executives of different ability are arranged in hierarchical layers of supervision across subsidiaries. For example, with layers indexed $\ell = 0, ..., L$, in a 'pyramid' structure the parent at layer $\ell = 0$ supervises directly a smaller number of subsidiaries at top layer $\ell = 1$ and indirectly a rising number of subsidiaries at increasingly lower layers $\ell > 1$; in an 'inverse pyramid' structure the parent at layer $\ell = 0$ supervises directly a larger number of subsidiaries at top layer $\ell = 1$ and indirectly a diminishing number of subsidiaries at increasingly lower layers $\ell > 1$. We assume that the activation of a hierarchical layer bears an administrative cost $wF > 0$, which is a market characteristic common to all BGs.

The optimal organizational structure is then determined by the total number of layers L as well as by the number n_ℓ and the output level y_ℓ of subsidiaries at each layer $\ell = 1, ..., L$ that jointly maximize the BG's overall profit

$$\Pi = \sum_{\ell=1}^{L} \left(n_\ell \Pi_\ell - wF \right), \tag{1}$$

with subsidiary profit Π_ℓ equal to revenues minus the remunerations of its employees and executive

$$\Pi_\ell = A^{\frac{1}{\sigma}} \left(y_\ell \right)^{\frac{\sigma-1}{\sigma}} - \frac{y_\ell}{\omega\theta_\ell} - w\theta_\ell, \tag{2}$$

subject to $n_0 = 1$ (as there is only one parent) and to the recursive supervision time constraint

$$\tau_{\ell-1} n_{\ell-1} = \varphi_\ell \theta_\ell n_\ell \tag{3}$$

for $\ell = 1, ..., L$.

The solution of the BG's profit maximization can be characterized recursively going down layer by layer starting from the parent ($\ell = 0$). For ease of exposition, we focus on a 'contiguous' organizational structure, such that executives of ability $\wp$ directly supervise executives of ability $\wp + 1$ and are directly supervised by executives of ability $\wp - 1$. We assume that such contiguity is an equilibrium outcome and then determine the conditions under which this is indeed the case.

First, at layer $\ell = 0$ there is only the parent and no profit is generated at that layer as the parent's time can be used for supervision but not for problem solving. Second, as the parent cannot produce without opening at least one subsidiary, the minimum number of layers of an active BG is two ($\ell = 0$ and $\ell = 1$). Third, the first order condition with respect to y_ℓ pins down the profit maximizing output of a subsidiary at layer $\ell \geq 1$ to

$$y_\ell = \left(\frac{\sigma-1}{\sigma} \omega\theta_\ell \right)^{\sigma} A,$$

which implies that the subsidiary's maximized profit evaluates to

$$\Pi_\ell = \left[a \left(\theta_\ell \right)^{\sigma-2} - w \right] \theta_\ell \tag{4}$$

with bundling parameter $a \equiv \omega^\sigma (A/\sigma) \left[\sigma/(\sigma-1) \right]^{1-\sigma}$. If we make the natural assumption that Π_ℓ is increasing in executive ability θ_ℓ (which is the case for $\sigma > 2$), at layer $\ell = 1$ the parent has an incentive to appoint only executives with the highest ability, i.e. those of ability $\wp = 1$. This implies $\ell = \wp = 1$ and thus $\theta_1 = e^{-\theta}$, with the parent receiving profits from each first-layer subsidiary equal to

$$\Pi_1 = \left[ae^{-\theta(\sigma-2)} - w\right] e^{-\theta}.$$

Fourth, due to the time constraint (3), the number of subsidiaries that can be opened at layer $\ell = 1$ equals

$$n_1 = \frac{\tau_0 n_0}{\varphi_1 \theta_1} = e^{\tau+\theta-\varphi},$$

so that the total profits received by the parent from its subsidiaries at layer $\ell = 1$ evaluate to

$$n_1\Pi_1 - wF = \left[ae^{-\theta(\sigma-2)} - w\right] e^{\tau-\varphi} - wF.$$

It then follows that layer $\ell = 1$ is activated at all if and only if $n_1\Pi_1 \geq 0$. Consider next layer $\ell = 2$. Given (4), also profits generated by subsidiaries at layer $\ell = 2$ are an increasing function of executive ability, which implies that the parent appoints again only executives with the highest feasible ability. This is $\wp = 2$ as, if assigned to layer $\ell = 2$, executives of ability $\wp = 1$ cannot be supervised by the executives of the same ability assigned to level $\ell = 1$, and thus cannot solve any problem. We therefore have $\ell = \wp = 2$ with the profit of each subsidiary equal to

$$\Pi_2 = \left[ae^{-2\theta(\sigma-2)} - w\right] e^{-2\theta}.$$

Due to the time constraint (3), the number of subsidiaries that can be opened at layer $\ell = 2$ equals

$$n_2 = e^{2\tau+3(\theta-\varphi)}$$

with total profit

$$n_2\Pi_2 - wF = \left[ae^{-2\theta(\sigma-2)} - w\right] e^{2\tau+\theta-3\varphi} - wF.$$

Layer $\ell = 2$ is activated if and only if $n_2\Pi_2 - wF \geq 0$. This condition is more stringent than $n_1\Pi_1 - wF \geq 0$ as long as $n_1\Pi_1 > n_2\Pi_2$ holds, which always happens for $\tau + 2(\theta - \varphi) < 0$, as in this case we have $n_2 < n_1$ and $\Pi_2 < \Pi_1$ given that Π_ℓ is increasing in executive ability (decreasing in $\wp$). Differently, for $\tau + 2(\theta - \varphi) > 0$ and thus $n_2 > n_1$, $n_1\Pi_1 > n_2\Pi_2$ holds if and only if φ satisfies

$$\varphi > \varphi_2 \equiv \frac{1}{2}\left(\tau + \theta + \ln \frac{ae^{-2\theta(\sigma-2)} - w}{ae^{-\theta(\sigma-2)} - w}\right). \tag{5}$$

As long as this restriction holds, a necessary condition for $\ell = 2$ to be worth activating is that $\ell = 1$ is itself worth activating. Vice versa, a sufficient condition for $\ell = 1$ to be worth activating is that $\ell = 2$ is itself worth activating. In other words, when restriction (5) holds, the hierarchy is contiguous as initially assumed. Intuitively, given $\varphi_\wp = e^{\varphi\wp}$, contiguous hierarchies better contain communication costs, which is important when these costs increase rapidly along layers (i.e. when φ is large).

As long as the hierarchy is contiguous, the foregoing results, obtained for $\ell = 1$ and $\ell = 2$, can be extended by induction to the generic layer ℓ, for which we have have $\ell = \wp$ and subsidiary profit

$$\Pi_\wp = \left[ae^{-\wp\theta(\sigma-2)} - w\right] e^{-\wp\theta}. \tag{6}$$

Solving the recursion (3) for $\ell = \wp$ with the chosen functional forms and $n_0 = 1$ yields the number of subsidiaries

$$n_\wp = e^{\wp\tau + \frac{1}{2}\wp(\wp+1)(\theta-\varphi)}. \tag{7}$$

Layer $\ell = \wp$ is activated if it breaks at least even: $n_\wp \Pi_\wp - wF \geq 0$. Hence, the lowest active layer ('cutoff layer') L in the hierarchy corresponds to the largest integer $\ell = \wp$ such that $n_\wp \Pi_\wp \geq 0$, i.e.

$$L = \wp^* \equiv \sup_\wp \{n_\wp \Pi_\wp - wF\} \geq 0. \tag{8}$$

For $\tau + \wp(\theta - \varphi) < 0$ with $\wp = 2, ..., \wp^*$, the hierarchy is always contiguous as in this case we have not only $\Pi_\wp < \Pi_{\wp-1}$ but also $n_\wp < n_{\wp-1}$; whereas for $\tau + \wp(\theta - \varphi) > 0$ and thus $n_\wp > n_{\wp-1}$, the hierarchy is contiguous if and only if we have

$$\varphi > \varphi_\wp \equiv \frac{1}{\wp}\left[\tau + (\wp - 1)\theta + \ln \frac{ae^{-\wp\theta(\sigma-2)} - w}{ae^{-(\wp-1)\theta(\sigma-2)} - w}\right] \tag{9}$$

for $\wp = 2, ..., \wp^*$. Hence, an inverse pyramid is always contiguous as $\tau + \wp(\theta - \varphi) < 0$ for all $\wp$'s implies $\varphi > \tau/\wp + \theta > \tau/\wp + \theta(\wp - 1)/\wp > \varphi_c$ given that the argument of the logarithm is smaller than 1.

Finally, the BG's total number of subsidiaries and its overall profits are given by $N = \sum_{\ell=0}^{L} n_\ell = \sum_{\wp=1}^{\wp^*} n_\wp$ and $\Pi = \sum_{\ell=0}^{L} (n_\ell \Pi_\ell - wF) = \sum_{\wp=1}^{\wp^*} (n_\wp \Pi_\wp - wF)$ respectively.

3.3 Pyramids and Diamonds

Despite its parsimony, the model has rich implications. As contiguous hierarchies are such that $n_\wp \Pi_\wp < n_{\wp-1}\Pi_{\wp-1}$ holds, (8) leads to:

Proposition 1. ***(Existence and uniqueness)*** *Assume (9) holds. A cutoff layer $\wp^*$ exists and it is unique.*

The cutoff layer determines the depth of the hierarchy with larger $\wp^*$ describing a deeper hierarchy with a larger number of layers. As a is an increasing function of ω, the cutoff condition (8) together with layer profit (6) implies that a BG's hierarchical depth increases with its production efficiency. Depth depends also on the rate at which layer profit $n_\wp \Pi_\wp$ falls as $\wp$ rises, which is higher for lower supervision efficiency (smaller τ) and lower communication efficiency (larger φ) as inferred from (6) and (7). Accordingly, a BG's hierarchical depth increases with its supervision efficiency and communication efficiency. Hence we have:

Proposition 2. ***(Number of layers)*** *Assume (9) holds. The cutoff layer $\wp^*$ is an increasing function of ω and τ whereas it is a decreasing function of φ.*

Through (7) supervision efficiency and communication efficiency also affect the number of subsidiaries placed at each layer:

Proposition 3. ***(Subsidiaries per layer)*** *Assume (9) holds. The number of subsidiaries $n_\wp$ at layer $\wp$ is an increasing function of τ and a decreasing function of φ for all $\wp = 1, ..., \wp^*$.*

The same parameters determine the shape of the hierarchy. In particular, the model can generate pyramids (i.e. left-skewed hierarchies), inverse pyramids (i.e. right-skewed hierarchies) and mixed structures ('diamonds'). This is due to the fact that (7) implies that a generic layer $\wp \geq 2$ has more (fewer) subsidiaries than the layer $\wp - 1$ above it if and only if $\tau + \wp(\theta - \varphi) > (<)0$. Therefore, there exists a threshold value $\wp_d \equiv \tau/(\varphi - \theta)$ such that the hierarchy is a pyramid for $\wp < \wp_d$ and an inverse pyramid for $\wp > \wp_d$. This leads to:

Proposition 4. ***(Shape)*** *Assume (9) holds. For $\varphi < \theta$ the optimal hierarchical structure of a BG is a pyramid. For $\varphi > \theta$ the optimal shape of a BG's hierarchy is: a pyramid iff $\tau/(\varphi - \theta) \leq 2$, a diamond iff $2 < \tau/(\varphi - \theta) < \wp^*$, or an inverse pyramid iff $\tau/(\varphi - \theta) \geq \wp^*$.*

As a corollary, while a sufficient condition for a pyramid is $\theta > \varphi$, a sufficient condition for an inverse pyramid is $\tau + (\theta - \varphi) < 0$ given that it implies $\tau + \wp^*(\theta - \varphi) < 0$ for any $\wp^* \geq 1$. Proposition 4 implies that, for given intermediate production efficiency gains (θ), a pyramid maximizes the overall profit of BGs with high supervision efficiency (large τ) and high communication efficiency (small φ). In contrast, an inverse pyramid maximizes the overall profit of BGs with low supervision efficiency (small τ) and low communication efficiency (large φ). A diamond is the best option for intermediate supervision efficiency and communication efficiency. The larger these are, the larger the pyramidal part of the diamond.

Finally, (7) sheds light on how steep pyramids or the pyramidal components of diamonds are. If we measure steepness by the absolute value of the difference between $n_\wp$ and $n_{\wp-1}$, then this is smaller the closer φ is to θ, leading to:

Proposition 5. ***(Steepness)*** *Assume (9) holds. The gap in the number of subsidiaries between layers $|n_\wp - n_{\wp-1}|$ is an increasing function of $|\theta - \varphi|$.*

We have seen that supervising the solution of more difficult versions absorbs more time; however, as more difficult versions can be tackled only by higher ability supervisees, less time is needed for communication. Then Proposition 5 states that pyramids or the pyramidal components of diamonds are steeper when those two opposing effects are close to offsetting each other (which happens for $\theta = \varphi$).

4 Model Validation

To validate the model, we check the consistency of some of its key predictions with the empirical patterns observed in our dataset. We proceed in two steps. First, we exploit the model's equilibrium equations to specify econometric regressions that allow us to obtain estimates for the parameters reg-

Table 5: Reduced form estimates of θ

	(1)	(2)	(3)
Dependent variable	log(Routinizability)		
Estimation method	OLS	OLS	OLS
Layer	0.0207***	0.0252***	0.0156***
	(0.000846)	(0.000609)	(0.000784)
Parent FE	Yes	Yes	Yes
Sub country FE	Yes	Yes	Yes
Cluster	BG	BG	BG
Sample	Excl. 1to1	Excl. 1to1	Excl. 1to1
		Layers 1-6	# Subs ≥ 10
Observations	2,687,600	2,666,804	1,121,548
R-squared	0.536	0.541	0.360

Notes: OLS estimations. *log*(*Routinizability*) is the logarithm of the routinizability of the layer's activity, as from the Princeton Data Improvement Initiative (PDII) at the 3-digit NAICS 2002 sector level; *Layer* is the degree of separation between the HQ and the corresponding subsidiary. The constant is omitted from the table. All specifications include parent company FE and subsidiary's country FE, and are estimated on the full world sample, excluding BGs with only one subsidiary. Specification (2) only includes the subsidiaries in the first six layers, specification (3) only includes subsidiaries belonging to BGs with ten or more subsidiaries. Standard errors clustered at the HQ level are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

ulating the difficulty of problem solving (θ), the efficiency in handling communication (φ), and the efficiency in supervising knowledge creation (τ). Second, we use these estimates, together with the model's structure, to calibrate the parameters regulating production efficiency (ω), using the BGs' number of layers as the target moment of the data while relying for the remaining parameters on normalizations and existing estimates borrowed from the literature.

Specifically, we are interested in checking whether the calibrated model can reproduce the patterns of hierarchical differentiation across the four groups of countries emphasized in Section 2: Europe, US, Japan and the full world sample. The model loads this cross-country variation on the heterogeneity of ω, φ and τ, while assuming that all other parameters (A, F, w, θ, A) are the same for all BGs. For calibration, we normalize $A = F = w = 1$ and set $\sigma = 5$ following Head and Mayer (2014). We then constrain φ and τ to be the same for all BGs in each country group. This implies that, given its definition $a \equiv \omega^{\sigma} (A/\sigma) [\sigma/(\sigma - 1)]^{1-\sigma}$, the only remaining source of variation in the bundling parameter a across BGs within each country group is ω.

4.1 Problem Solving Efficiency

In the case of θ, we recall its definition in $\theta_{\wp} = e^{-\theta \wp}$ as not only the rate at which the solutions of more difficult problem versions (those with lower $\wp$) are increasingly associated with higher employees' productivity, but also the rate at which such solutions become harder to find. Given that, in equilibrium, more difficult problems are allocated to higher layers, the model implies that θ is also the rate at which problem difficulty falls as $\wp$ increases. Based on this equilibrium property, we proxy the difficulty $\theta_{\wp}$ of problems addressed in equilibrium at layer $\wp$ by the inverse of the routinizability of the layer's

Table 6: Reduced form estimates of α_1 and α_2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	$\log(n_\wp)$				$\log(n_\wp)$			
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Country/Area	World	USA	Japan	C. Europe	World	USA	Japan	C. Europe
Layer	0.666***	0.848***	1.313***	0.625***	0.833***	0.968***	1.591***	0.732***
	(0.00851)	(0.0103)	(0.0492)	(0.0147)	(0.00159)	(0.00226)	(0.0271)	(0.00414)
Layer^2	-0.0650***	-0.0814***	-0.185***	-0.0721***	-0.141***	-0.154***	-0.298***	-0.121***
	(0.00333)	(0.00525)	(0.0200)	(0.00638)	(0.000923)	(0.00196)	(0.0123)	(0.00232)
Cluster	BG	BG	BG	BG	BG	BG	BG	BG
Sample	Excl. 1to1	Excl. 1to1	Excl. 1to1	Excl. 1to1	Excl. 1to1, layers 1-6	Excl. 1to1, layers 1-6	Excl. 1to1, layers 1-6	Excl. 1to1, layers 1-6
Observations	939,062	384,629	4,768	191,700	935,812	383,794	4,738	191,330
R-squared	0.512	0.708	0.535	0.459	0.541	0.729	0.571	0.473

Notes: OLS estimations. *log*($n_\wp$) is the log number of subsidiaries at layer $\wp$; *Layer* is the degree of separation between the HQ and the corresponding subsidiary. Together with the estimated θ from Table 5, the estimate coefficients of *Layer* and *Layer*^{2} allow for the identification of the values of τ and φ. Model estimated for 4 groups of countries (world, USA, Japan, core Europe) always excluding BGs with only one subsidiary. Specifications (5) to (8) restrict the sample to only include subsidiaries in the first six layers. Standard errors clustered at the HQ level are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

activities. As taking $\theta_\wp = e^{-\theta\wp}$ in logs and inverting gives

$$\ln(\theta_\wp^{-1}) = \theta\wp, \tag{10}$$

we thus estimate θ by regressing the log routinizability of a layer $\ln(\theta_\wp^{-1})$ on the layer's index $\wp$.

The parameter θ thus corresponds to the average semi-elasticity of routinizability to the layer index within groups, estimated controlling for BG fixed effects and subsidiary host country fixed effects, and clustering standard errors at the HQ level. Routinizability is the same variable already used in Figure 5 and retrieved from Blinder and Krueger (2009, 2013). The estimation results are reported in Table 5 for the full world sample. The table shows that, as one moves down the hierarchy by one layer, routinizability increases by around 2% on average.

As for τ and φ, they are estimated starting from equation (7) as follows. Taking logs and bundling the primitive parameters as $\alpha_1 \equiv \tau + (\theta - \varphi)/2$ and $\alpha_2 \equiv (\theta - \varphi)/2$ gives

$$\ln(n_\wp) = \alpha_1\wp + \alpha_2\wp^2, \tag{11}$$

which implies that α_1 and α_2 can be estimated by regressing the log number of subsidiaries at layer $\wp$ on the layer's index.[14] Then, using the estimated α_1 and α_2 together with the previously estimated θ, the corresponding values of τ and φ can be backed out by inverting the definitions of α_1 and α_2 as $\tau = \alpha_1 - \alpha_2$ and $\varphi = \theta - 2\alpha_2$. The estimates of α_1 and α_2 from regression (11) are displayed in Table 6 for the four groups of countries. Both coefficients are consistently and significantly estimated across the four samples. A sensitivity check (columns 5 to 8) performed excluding BGs with more

[14] As BGs with only one affiliate are uninformative for this regression, we exclude them from the estimation.

Table 7: Bootstrap for $\hat{\tau}$ and $\hat{\phi}$

	$\hat{\tau}$			$\hat{\phi}$		
Country/Area	5^{th} p.	50^{th} p.	95^{th} p.	5^{th} p.	50^{th} p.	95^{th} p.
World	0.67688	0.74745	0.817205	0.1185	0.1594	0.2023
USA	0.871795	0.96475	1.043	0.14786	0.2055	0.2668
Japan	1.3159	1.812	2.5111	0.2747	0.5707	1.0567
C. Europe	0.61225	0.7457	0.831	0.11442	0.194	0.2508

Notes: Estimates are conducted on 1000 bootstrapped samples covering 5% of the BGs in each Country/Area. $\hat{\phi}$ is computed using $\hat{\theta} = 0.0207$ which is our preferred estimate in Table 5, Column (1).

than 6 layers shows that the estimates are not overly sensitive to extreme distributions in the shape of the BGs. To estimate τ and φ together with their standard errors, we bootstrap our estimates of α_1 and α_2 on 1,000 random subsamples, each including 5% of the relevant BGs. We then compute τ and ϕ using θ from Table 5, Column (1).

The bootstrapped results are reported in Table 7 for the four country groups, taking the median across the subsamples as point estimate, and the 5^{th} and 95^{th} percentiles of the distribution as confidence intervals. The estimated parameters indicate that US BGs are less efficient than European BGs in supervising knowledge creation; they are, however, more efficient in handling associated communication across subsidiaries. Moreover, while Japanese BGs are more efficient than European and US BGs in supervising knowledge creation, they are less efficient in handling associated communication across subsidiaries.

4.2 Production Efficiency

Turning to ω, we consider cutoff layer condition (8), under the normalized values of A, F and w, and the value of σ borrowed from the literature. We iterate on alternative values of ω until condition (8) delivers the observed cutoff layers for the four groups of countries. In order to have enough variation across hierarchies to identify ω, we target BGs with at least 50 subsidiaries. This implies that we have to restrict the analysis to our European, US and full world samples as in our Japanese sample the number of observations is significantly reduced.[15] For this target, the cutoff layers are 10 in the European and full world samples, and 12 in the US sample. The resulting estimated ω's are 2.28 for both the European and the full world samples, and 2.61 for the US sample. Hence, for BGs with at least 50 subsidiaries, production efficiency is revealed to be around 15% higher in US BGs than in European and world ones.

4.3 Predicted Hierarchies

Figure 7 depicts the hierarchies predicted by the calibrated model for BGs with at least 50 subsidiaries across the three country groups. All hierarchies look like diamonds. Skewness is 0.08 for the European sample, 0.10 for the US sample, and 0.07 for the full world sample. This reveals the overall dominance of right-skewed structures, closer to inverted pyramids than pyramids.

The calibrated model can be used to study the comparative statics of optimal hierarchies through the comparison of simulated counterfactual scenarios with the baseline outcomes reported in Figure 7. Figure 8 Panel (a) compares the baseline outcome (blue bars) with the simulated outcome (yellow bars) obtained when US BGs with at least 50 subsidiaries are imputed the same problem solving efficiency as their European counterparts.[16] The figure shows that in the simulated outcome US BGs lose depth with respect to the baseline, with the number of layers falling from 12 to 11. Moreover, the simulated distribution of subsidiaries across layers is (first-order) stochastically dominated by the baseline one, with larger shares of subsidiaries at higher layers in the former than in the latter distributions. Lastly, in the simulation US BGs are more right-skewed than in the baseline, with a 14% increase in skewness from 0.10 to 0.12.

Analogously, Figure 8 Panel (b) compares the baseline outcome (blue bars) with the simulated outcome (yellow bars) obtained when European BGs with at least 50 subsidiaries are imputed the same problem solving efficiency as their US counterparts.[17] The figure shows that in the simulated outcome European BGs gain depth with respect to the baseline, with the number of layers rising from 10 to 12. Moreover, the simulated distribution of subsidiaries across layers (first-order) stochastically dominates the baseline one, with larger shares of subsidiaries at lower layers in the former than in the latter distributions. Yet, in the simulation European BGs are more right-skewed than in the baseline, with a 30% increase in skewness from 0.08 to 0.10.

5 Conclusions

Hierarchical differentiation is a cornerstone of the organizing process. While research has traditionally focused primarily on pyramid-shaped hierarchies, alternative hierarchical shapes may be at least as relevant. In this paper we have done three things. First, exploiting a newly assembled dataset, we have provided the first worldwide overview of the patterns of hierarchical differentiation across Business Groups (BGs), highlighting the coexistence of different hierarchical shapes. Second, we have shown how the different shapes can arise as optimal hierarchical structures in a parsimonious knowledge-based model of BGs when subsidiaries operations involve ubiquitous problem solving under parents'

[15] We account for 5,992 BGs with more than 50 subsidiaries in the full world sample, of which 1,517 US groups, 1,138 BGs in the European sample and 236 Japanese ones.

[16] This is equivalent to the counterfactual in which European BGs are imputed the production efficiency of US BGs.

[17] This is equivalent to the counterfactual in which US BGs are imputed the production efficiency of European BGs.

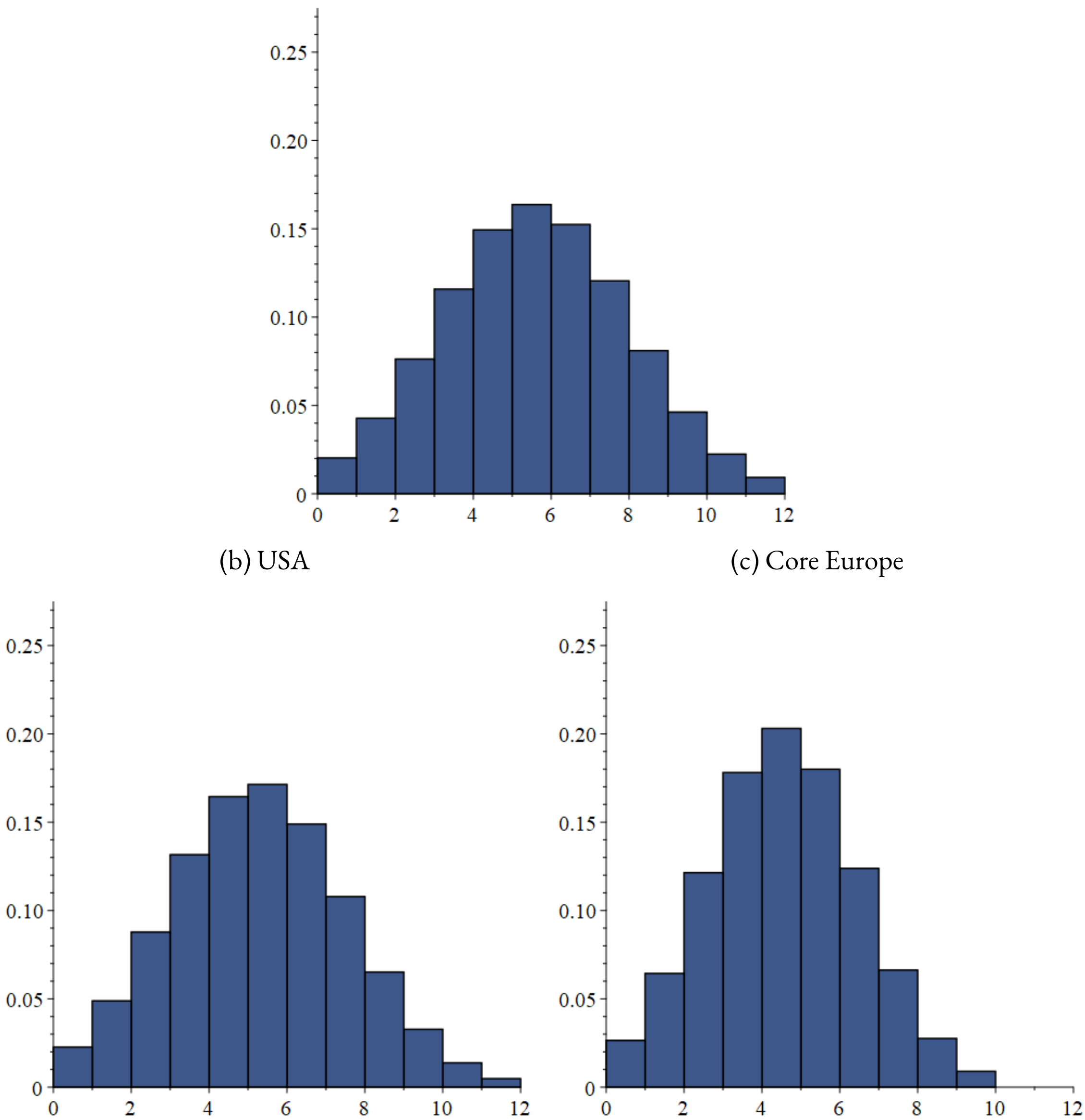
supervision. Third, we have checked the consistency of some of its key predictions with the empirical patterns observed in our dataset for three groups of countries: core Europe, US, and the full world sample. The model has successfully passed the consistency test. The estimation of its parameters has revealed that, while US BGs are less efficient than European ones in supervising knowledge creation, they are more efficient in handling associated communication across subsidiaries. It has also revealed that US BGs are more efficient than European BGs in production too.

Figure 7: Predicted Hierarchies - Density

(a) World

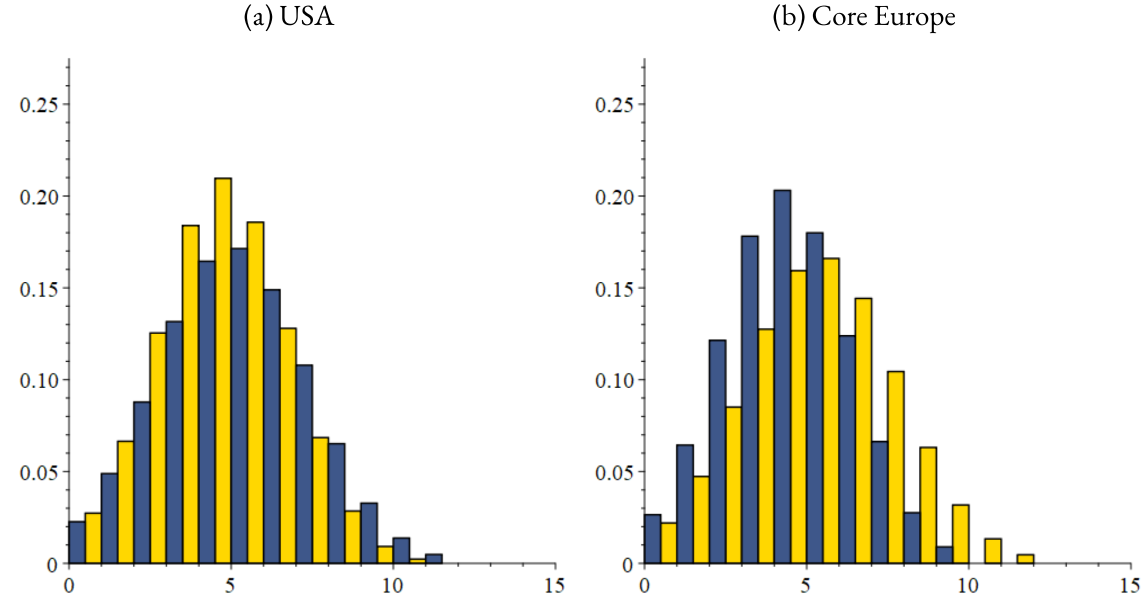
(b) USA

(c) Core Europe



Notes: Predicted hierarchies given $\sigma = 5; a = 5; A = 1; w = 1; F = 1$, with $\hat{\theta} = 0.0112; \hat{\tau}_w = 1.0695; \hat{\phi}_w = .1742; \hat{\tau}_{us} = 1.1319;$ $\hat{\phi}_{us} = .193; \hat{\tau}_{eu} = 1.391; \hat{\phi}_{eu} = .2632$, retrieved from estimation on groups with at least 50 subsidiaries (not shown).

Figure 8: Predicted Hierarchies - Counterfactual



Notes: Predicted hierarchies (as above, Figure 7) are shown in blue, counterfactual hierarchies are in yellow. The latter are computed given $\sigma = 5$; $a = 5$; $A = 1$; $w = 1$; $F = 1$; with $\hat{\theta} = 0.0112$; $\hat{\tau}_{us} = 1.391$; $\hat{\phi}_{us} = .2632$; $\hat{\tau}_{eu} = 1.1319$; $\hat{\phi}_{eu} = .193$.

References

Alfaro, L. (2017). Gains from foreign direct investment: Macro and micro approaches. The World Bank Economic Review, 30(Supplement˙1), S2-S15.

Alfaro, L., & Charlton, A. (2009). Intra-industry foreign direct investment. American Economic Review, 99(5), 2096–2119.

Almeida, H. V., & Wolfenzon, D. (2006). A theory of pyramidal ownership and family business groups. The Journal of Finance, 61(6), 2637–2680.

Altomonte, C., & Rungi, A. (2013). Business groups as hierarchies of firms: determinants of vertical integration and performance (Tech. Rep.). ECB Working Paper.

Altshuler, R., & Grubert, H. (2003). Repatriation taxes, repatriation strategies and multinational financial policy. Journal of Public Economics, 87(1), 73–107.

Aminadav, G., & Papaioannou, E. (2020). Corporate control around the world. The Journal of Finance, 75(3), 1191-1246. doi: https://doi.org/10.1111/jofi.12889

Belenzon, S., & Berkovitz, T. (2010). Innovation in business groups. Management Science, 56(3), 519-535.

Belenzon, S., Hashai, N., & Patacconi, A. (2019). The architecture of attention: Group structure and subsidiary autonomy. Strategic Management Journal, 40(10), 1610–1643.

Belenzon, S., & Schankerman, M. (2013). Spreading the word: Geography, policy, and knowledge spillovers. The Review of Economics and Statistics, 95(3), 884-903.

Berle, A. A., & Means, G. C. (1932). The modern corporation and private property. Commerce Clearing House.

Blinder, A. S., & Krueger, A. B. (2009, August). Alternative measures of offshorability: A survey approach (Working Paper No. 15287). National Bureau of Economic Research.

Blinder, A. S., & Krueger, A. B. (2013). Alternative measures of offshorability: a survey approach. Journal of Labor Economics, 31(S1), S97–S128.

Boutin, X., Cestone, G., Fumagalli, C., Pica, G., & Serrano-Velarde, N. (2013). The deep-pocket effect of internal capital markets. Journal of Financial Economics, 109(1), 122-145.

Cestone, G., & Fumagalli, C. (2005). The strategic impact of resource flexibility in business groups. RAND Journal of Economics, 36(4), 193-214.

Cestone, G., Fumagalli, C., Kramaz, F., & Pica, G. (2016, June). Insurance Between Firms: The Role of Internal Labor Markets (Tech. Rep. No. 11336).

Chapelle, A., & Szafarz, A. (2007). Control consolidation with a threshold: An algorithm. IMA Journal of Management Mathematics, 18(3), 235–243.

Colpan, A., Hikino, T., & Lincoln, J. (2010). The oxford handbook of business groups. doi: 10.1093/oxfordhb/9780199552863.001.0001

Cravino, J., & Levchenko, A. A. (2017, November). The distributional consequences of large devaluations. American Economic Review, 107(11), 3477-3509.

Del Prete, D., & Rungi, A. (2017). Organizing the global value chain: a firm-level test. Journal of

International Economics, 109, 16–30.

Desai, M. A., Foley, C. F., & Hines Jr, J. R. (2004). A multinational perspective on capital structure choice and internal capital markets. The Journal of finance, 59(6), 2451–2487.

Eppinger, P. S., & Kukharskyy, B. (2020). Contracting institutions and firm integration around the world (Tech. Rep.). University of Tübingen Working Papers in Economics and Finance.

Faccio, M., & Lang, L. H. (2002). The ultimate ownership of western european corporations. Journal of financial economics, 65(3), 365–395.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. Journal of Political Economy, 108(5), 874–904.

Garicano, L., & Rossi-Hansberg, E. (2015). Knowledge-based hierarchies: Using organizations to understand the economy. Annual Review of Economics, 7(1), 1-30.

Garicano, L., & Wu, Y. (2012). Knowledge, communication, and organizational capabilities. Organization Science, 23(5), 1382–1397.

Gorodnichenko, Y., Kukharskyy, B., & Roland, G. (2021). Cultural Distance, Firm Boundaries, and Global Sourcing (Tech. Rep.). UC Berkeley.

Graham, B., & Dodd, D. L. (1934). Security analysis. Whittlesey House, McGraw-Hill Book Co.

Grosskurth, P. (2019). Mne and where to find them: An intertemporal perspective on the global ownership network (Tech. Rep.). Ruhr Economic Papers.

Group, W. B., Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). Worldwide governance indicators. World Bank Group.

Khanna, T., & Yafeh, Y. (2007). Business groups in emerging markets: Paragons or parasites? Journal of Economic Literature, 45(2), 331–372.

La Porta, R., Lopez-de Silanes, F., & Shleifer, A. (1999). Corporate ownership around the world. The journal of finance, 54(2), 471–517.

Lewellen, K., & Robinson, L. A. (2013). Internal ownership structures of us multinational firms. Available at SSRN 2273553.

Mayer, T., Melitz, M. J., & Ottaviano, G. I. P. (2014, February). Market size, competition, and the product mix of exporters. American Economic Review, 104(2), 495-536.

Mintz, J. M., & Weichenrieder, A. J. (2010). The the indirect side of direct investment: Multinational company finance and taxation. MIT press.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. The Quarterly Journal of Economics, 122(2), 569–600.

Rauch, J. E. (1999). Networks versus markets in international trade. Journal of international Economics, 48(1), 7–35.

Rungi, A., Morrison, G., & Pammolli, F. (2017). Global ownership and corporate control networks. IMT Lucca EIC WP Series, 7.

Sonno, T. (2017). Globalisation and conflicts: the good, the bad, and the ugly of corporations in africa. Essays on Globalisation and Conflicts, Populism and Firm Heterogeneity, 131.

UNCTAD. (2016). World investment report 2016 - investor nationality: Policy challenges. United

Nations.
Wellman, N., Applegate, J. M., Harlow, J., & Johnston, E. W. (2020). Beyond the pyramid: Alternative formal hierarchical structures and team performance. Academy of Management Journal, 63(4), 997-1027.

Appendix

A Ownership data and robustness

In this Appendix, we provide some more information on the basic ownership data and how we use them for the purpose of our analyses. In particular, we describe more in detail the characteristics and the differences in the determination of BGs that can be retrieved from Bureau Van Djik data either following the approach proposed by Sonno (2017) or by Rungi et al. (2017). These firm-level databases collect original information from a variety of national and international registries, regulatory bodies, companies' annual reports, websites and specialized press. For our purpose, we extract information on shareholding activity for companies active in more than 200 countries in year 2015. The starting point of the two approaches is different, however, how we will see in this section, they both retrieve a comparable dataset of BGs worldwide.

In the paper we follow Sonno (2017) who starts from the Historical Ownership Database by Bureau Van Dijk. This dataset provides for each company information on all its shareholders and identifies several types of relations. The author proposes an algorithm that retrieves the hierarchical distance of a company from its parent company using two types of relations: the corporate Global Ultimate Owner with at least 50.01% of voting rights (GUO 50C hereafter), that is the highest corporate independent shareholder in the shareholding structure of a company, and the Immediate Shareholder (ISH hereafter), that is the first shareholder in the path from an affiliate to its GUO. Combining the definition of GUO 50C and ISH, with the fact that each shareholder is reported more than once depending on its role in the shareholding structure of a subsidiary, it is possible to create a routine that counts the steps leading a subsidiary to its parent. Throughout this approach we rely on the definition of direct or indirect majority ($\geqslant 50.01\%$) of voting rights provided by Bureau Van Djik. Therefore, we use an exogenous definition of control. Using this procedure, it is possible to obtain a dataset of $2,901,466$ parent companies controlling $5,676,289$ subsidiaries for 2015, as we have excluded branches from our analysis. Also note that as we require additional details on the type of ownership links, the latter yields different subsamples of the dataset depending on the information available to us. As an example, to document the sector of parents and affiliates (allocated in a specific layer within the BG) we rely on a sample of 2.2M parents and 4.6M affiliates. A detailed breakdown of the different sample compositions can be found in Table 1.

We test the robustness of this algorithm for the definition of BGs, and the characterization of their hierarchical structure, replicating our results on a dataset of BGs obtained following the methodology proposed by Rungi et al. (2017), more rooted in the network theory approach. The authors use data from the Ownership section of the Orbis Database, where Orbis collects all the information available on its ownership structure. For each company, we have a list of all (individual, corporate or state) shareholders. Any time a company invests in the equity of another company, an ownership network is generated such that voting rights can be separated from cash rights. In modern economies, corporate ownership structures can become very sophisticated (see for example, La Porta et al., 1999) and

Table A1: Geographic coverage - Robustness

	Business Groups				Subsidiaries			
Region or Country	All	%	Multinational	%	All	%	Foreign	%
Africa	5,102	0.22	4,169	2.07	30,346	0.64	17,088	2.27
Asia	105,449	4.45	19,142	9.51	316,014	6.67	99,624	13.24
Australia	58,788	2.48	2,771	1.38	136,189	2.87	14,750	1.96
EU 28	600,829	25.35	111,522	55.41	1,625,508	34.29	387,006	51.44
Other Europe	36,073	1.52	14,089	7.00	84,045	1.77	22,441	2.98
Latin America	30,058	1.27	18,247	9.07	83,227	1.76	51,693	6.87
Russia	29,741	1.25	974	0.48	110,232	2.33	50,541	6.72
USA	1,435,218	60.56	22,511	11.18	2,138,025	45.10	63,220	8.40
Rest of the World	68,634	2.90	7,847	3.90	216,766	4.57	45,992	6.11
Total	2,369,892	100.00	201,272	100.00	4,740,352	100.00	752,355	100.00

Notes: The table refers to the data-set constructed following Rungi et al. (2017). It details the number of observations of subsidiaries and business groups in a list of regions or countries. It also specifies the number of business groups that control at least one subsidiary abroad (multinational), and that of subsidiaries that are controlled by a foreign parents. For regional aggregation we have used the online version of the United Nations publication Standard Country or Area Codes for Statistical Use.

the identification of ultimate parent companies can become very difficult, especially in the case of MNEs (UNCTAD, 2016). The authors model a backward solution for a voting rule across interlocking assemblies of shareholders. Assemblies of shareholders interlock when (individual and corporate) shareholders generate cross-holdings, ownership cycles and multiple ownership paths that connect companies through equity. When shareholding activity interlocks, a coordination effort is required across different ownership paths to enforce management decisions starting from headquarters. This approach is able to extract hierarchies of firms made of parent companies and their subsidiaries, all ordered on hierarchical layers, by considering all the ownership paths that can connect any two nodes (companies/shareholders) in the entire ownership space. Starting from a basic ownership matrix including all the shareholding links between companies and shareholders, the authors simulate corporate control after assuming that the latter entails cases of: (i) *direct control*, when the parent company holds the absolute majority of voting rights in a subsidiary; (ii) *indirect control by transitivity*, when the parent company has direct control of a subsidiary that in turn has direct control over another subsidiary, in a sequence; (iii) *indirect control by consolidation* of voting rights, when a parent company is able to control a subsidiary by summing up to a majority of capital shares held in her portfolio and/or in the portfolio of other subsidiaries; (iv) *dominant shareholding*, when a parent company is able to control a company with just a minority stake, because other minority shareholders are too much fragmented to form an opposing coalition. For the purpose of this paper, we limit our analyses only to the first three cases, excluding control by dominant minorities, although all of them do find a correspondence in international accounting standards (IFRS, 2011; OECD, 2015; UNCTAD, 2009; OECD, 2008). See Rungi et al. (2017) on the solution of the voting rule starting from the original global ownership matrix.

As we can see by comparing relevant descriptive statistics, the results obtained with the two approaches are highly consistent with each other. Table A1, for example, shows the geographic coverage of the data obtained following Rungi et al. (2017). Although the figures are slightly lower than those obtained

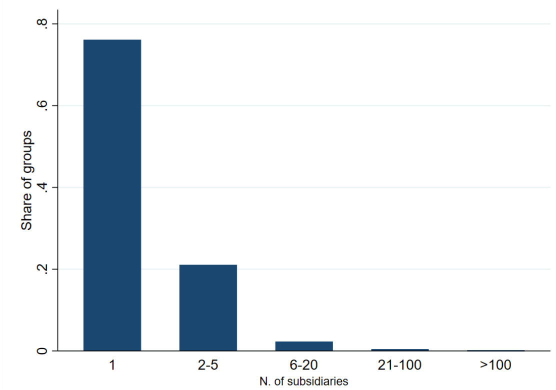
Table A2: Hierarchical distance - Robustness

Layer	Domestic subsidiaries	%	Foreign subsidiaries	%	All subsidiaries	%
1	1,579,504	73.51	228,542	36.11	1,808,046	65.00
2	376,511	17.52	186,062	29.40	562,573	20.22
3	123,021	5.73	104,063	16.44	227,084	8.16
4	43,079	2.00	55,414	8.76	98,493	3.54
5	15,354	0.71	28,135	4.45	43,489	1.56
6	5,934	0.28	14,182	2.24	20,116	0.72
7	2,518	0.12	8,132	1.28	10,650	0.38
8	1,321	0.06	3,765	0.59	5,086	0.18
9	600	0.03	2,104	0.33	2,704	0.10
10	268	0.01	987	0.16	1,255	0.05
> 10	580	0.03	1552	0.25	2132	0.08
Total	2,148,690	100.00	632,938	100.00	2,781,628	100.00

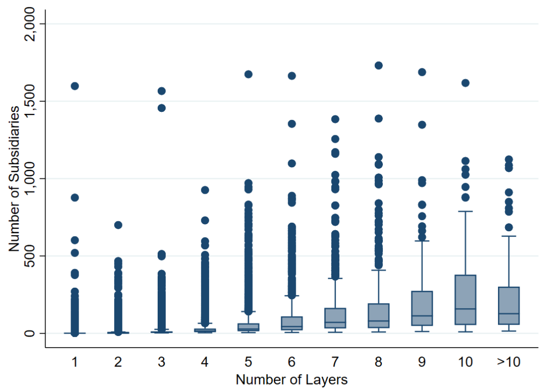
Notes: The table refers to the data-set constructed following Rungi et al. (2017). It shows the number of subsidiaries per layer, distinguishing between domestic and foreign subsidiaries. It excludes business groups with only one subsidiary.

Figure A1: Hierarchical description - Robustness

(a)



(b)



Notes: The graph refers to the data-set constructed following Rungi et al. (2017). The left panel shows the distribution of groups' dimension in terms of number of subsidiaries. The right panel presents eleven box-plots of groups' dimension conditional on the number of layers their hierarchies display. We have excluded the 14 business groups with more than 2,000 subsidiaries so as to make the figure more legible.

following Sonno (2017) and displayed in table 2, the correlation between the two tables is over 0.99 for every category considered, even if we break down the table further into 21 countries and regions.[18] Similarly, looking at tables A2 and 3 that detail the distribution of domestic and foreign subsidiaries among hierarchical layers, we observe slightly different numbers but very similar shares. Again, the figures in the two tables correlate at more than 0.99. Finally, looking at the left panel of figure A1 we observe that the distribution of groups in terms of the subsidiaries they control is extremely skewed, with more than 75% of the groups controlling only one subsidiary, and only 0.08% of them having

[18]In following the approach proposed by Sonno (2017), we use the 2019 release of the Historical Ownership Database of Bureau van Dijk, while the dataset elaborated by Rungi et al. (2017) uses the 2016 release of the Ownership section of the Orbis database by Bureau van Dijk. This discrepancy partially explains the higher number of observations found by the Sonno (2017) approach.

Table A3: Number of affiliates per layer across business groups - Robustness

Layer	Maximum layer of the BG: 1	2	3	4	5	6	7	8	9	10	> 10
1	1.3	3.2	7.3	16.3	39.3	60.5	86.1	75.6	80.3	91.8	118.6
2		2.6	7.1	15.6	27.2	37.2	49.0	97.9	71.3	89.3	78.6
3			3.5	9.9	21.3	42.1	45.0	77.0	168.7	72.1	145.2
4				4.2	11.3	23.1	35.1	32.4	82.0	92.4	79.7
5					4.0	11.4	21.2	23.4	35.5	36.0	32.4
6						5.2	13.3	25.7	34.3	55.0	28.9
7							5.5	13.1	21.7	39.7	58.8
8								3.5	12.6	22.8	19.6
9									4.7	9.8	18.4
10										2.5	15.2
> 10											10.7
N. of BGs	2,367,786	122,697	24,609	6,829	2,491	1,108	561	278	150	83	117

Notes: The table refers to the data-set constructed following Rungi et al. (2017). It shows the average number of subsidiaries per layer for business groups having different maximum layers. If business groups with only one subsidiary were excluded, the mean number of affiliates in the first layer for business groups with only one layer would be 2.8.

more than 100 subsidiaries. Once more, this is very similar to what we observed in figure 2 using Sonno (2017)'s approach.

One feature that is fundamental for our analysis is the hierarchical structure of BGs. To assess any potential difference between the two sample in this respect, we look at tables 4 and A3, that refer, respectively, to the samples following Sonno (2017) and Rungi et al. (2017), and report on the average composition of business groups, conditional on the number of layers of their hierarchy. Even if the two tables are qualitatively very similar, and both distinctly show groups shaped as 'inverted pyramids' (groups with more subsidiaries in the first layers and fewer in the lowest ones), it is worth noticing one main difference: the data-set obtained following Sonno (2017) show a slightly higher number of subsidiaries in the first hierarchical layer. This is because the approach of Rungi et al. (2017) focuses more on *cross shareholding structures* by some shareholder thus potentially adding, in the case of complex hierarchical structures, an additional layer to the BG structure. Indeed, there is a decrease of BGs with at most 1 layer in favor of the other categories.

B Stylized facts and robustness

In this section we present the regression behind our graphs in Section 2. Figure 5 on routinizability is backed by regression Table A4. Figure 6 on contractibility is backed by regression Table A5.

Table A4: Routinizability

Dependent variable	Routinizability
Estimation method	OLS
Layer 2	0.088***
	(0.001)
Layer 3	0.123***
	(0.002)
Layer 4	0.151***
	(0.002)
Layer 5	0.164***
	(0.003)
Layer 6	0.179***
	(0.005)
Layer 7	0.222***
	(0.006)
Layer 8	0.203***
	(0.009)
Layer 9	0.215***
	(0.012)
Layer 10	0.256***
	(0.015)
Layer > 10	0.311***
	(0.014)
Parent FE	Yes
Country FE	Yes
Observations	2,687,600
R-squared	0.568

Notes: OLS estimations. *Routinizability* is the routinizability of the layer's activity, as retrieved from the Princeton Data Improvement Initiative (PDII) at the 3-digit NAICS 2002 sector level; *Layer* is the degree of separation between the HQ and the corresponding subsidiary. The constant is omitted from the table. All specifications include parent company FE and subsidiary's country FE. The model is estimated on the full world sample, excluding BGs with only one affiliate. Heteroskedasticity robust standard errors at the HQ level are in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

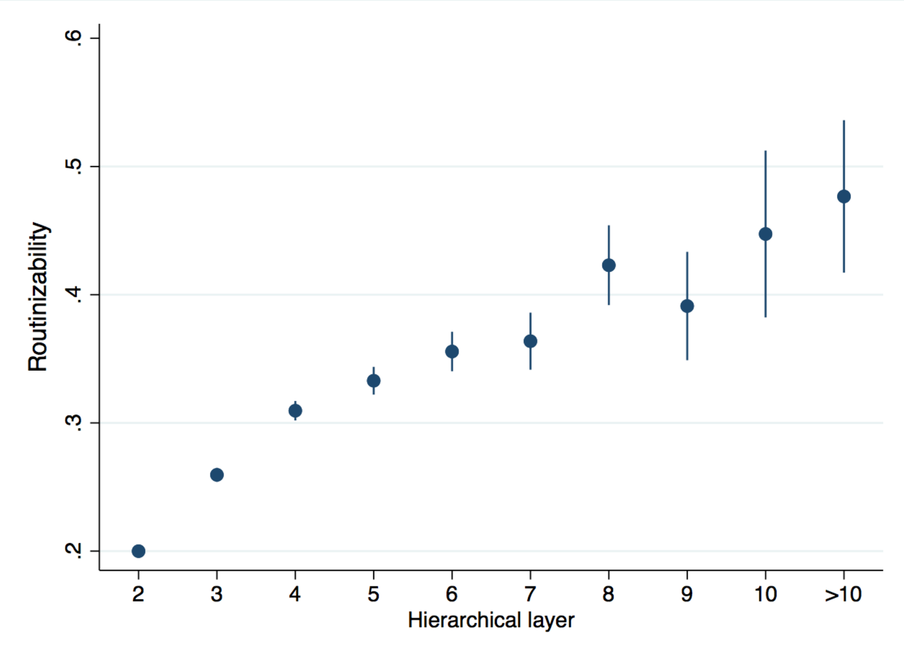
Table A5: Contractibility

Dependent variable	Contractibility
Estimation method	OLS
Layer 2	0.059***
	(0.002)
Layer 3	0.095***
	(0.004)
Layer 4	0.113***
	(0.005)
Layer 5	0.092***
	(0.009)
Layer 6	0.112***
	(0.012)
Layer 7	0.091***
	(0.017)
Layer 8	0.087***
	(0.025)
Layer 9	0.049
	(0.033)
Layer 10	0.101***
	(0.033)
Layer > 10	0.150***
	(0.036)
Parent FE	Yes
Country FE	-
Observations	316,700
R-squared	0.693

Notes: OLS estimations. *Contractability* is the contractability index developed by Nunn (2007). The index combines a measure of the quality of contract enforcement for each country with a measure of contractual intensity for each final good in the manufacturing sector. Contractability varies by country and sector: a high value of contractability is associated with a high level of judicial quality combined with a high level of contract intensity. *Layer* is the degree of separation between the HQ and the corresponding subsidiary. The constant is omitted from the table. All specifications include parent company FE. The model is estimated on the full world sample, excluding business groups with only one affiliate. Heteroskedasticity robust standard errors at the HQ level are in parentheses, * $p< 0.05$, ** $p< 0.01$, *** $p < 0.001$.

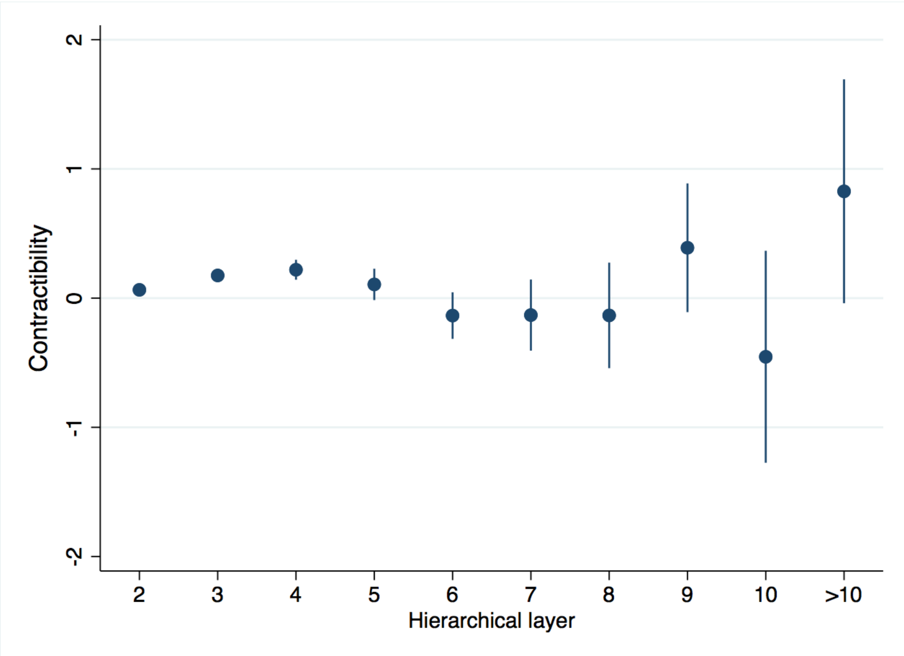
We then replicate the stylized facts using the data-set constructed following Rungi et al. (2017). Specifically, Figure A2 replicates Figure 5, while Figure A3 replicates Figure 6.

Figure A2: Routinizability over hierarchies - Robustness



Notes: The graph refers to the data-set constructed following Rungi et al. (2017). It shows the coefficients and 95% confidence intervals obtained from a regression of the routinizability index of each subsidiary on a series of binary variables for the hierarchical layer of the subsidiaries themselves, including group and country FE, and using robust standard errors.

Figure A3: Contractibility over hierarchies - Robustness



Notes: The graph refers to the data-set constructed following Rungi et al. (2017). It shows the coefficients and 95% confidence intervals obtained from a regression of the contractibility index index of each subsidiary (liberal definition using inputs that are neither sold on an organized exchange nor reference priced) on a series of binary variables for the hierarchical layer of the subsidiaries themselves, including group FE, and using robust standard errors. Country FE are omitted because contractibility incorporates rule of law that is country-specific

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1803	Nikhil Datta Stephen Machin	Living wages and discontinuities for low-wage workers
1802	Holger Breinlich Dennis Novy J.M.C. Santos Silva	Trade, gravity and aggregation
1801	Ari Bronsoler Joseph Doyle John Van Reenen	The impact of healthcare IT on clinical quality, productivity and workers
1800	Italo Colantone Gianmarco Ottaviano Piero Stanig	The backlash of globalization
1799	Jose Maria Barrero Nicholas Bloom Steven J. Davis	Internet access and its implications for productivity, inequality and resilience
1798	Nicholas Bloom Tarek Alexander Hassan Aakash Kalyani Josh Lerner Ahmed Tahoun	The diffusion of disruptive technologies
1797	Joe Fuller Stephen Hansen Tejas Ramdas Raffaella Sadun	The demand for executive skills
1796	Stephen Michael Impink Andrea Prat Raffaella Sadun	Communication within firms: evidence from CEO turnovers
1795	Katarzyna Bilicka Daniela Scur	Organizational capacity and profit shifting
1794	Monica Langella Alan Manning	Income and the desire to migrate

1793	Nicholas Bloom Arjun Ramani	The donut effect of Covid-19 on cities
1792	Brian Bell Nicholas Bloom Jack Blundell	This time is not so different: income dynamics during the Covid-19 recession
1791	Emanuel Ornelas Patricia Tovar	Intra-bloc tariffs and preferential margins in trade agreements
1790	Jose Maria Barrero Nicholas Bloom Steven J. Davis	Why working from home will stick
1789	Scott R. Baker Nicholas Bloom Steven J. Davis Marco Sammon	What triggers stock market jumps?
1788	Nicolas Bloom Robert S. Fletcher Ethan Yeh	The impact of Covid-19 on US firms
1787	Philippe Aghion Antonin Bergeaud Matthieu Lequien Marc J. Melitz Thomas Zuber	Opposing firm-level responses to the China shock: horizontal competition versus vertical relationships
1786	Elsa Leromain Gonzague Vannoorenberghe	Voting under threat: evidence from the 2020 French local elections
1785	Benny Kleinman Ernest Liu Stephen J. Redding	Dynamic spatial general equilibrium
1784	Antonin Bergeaud Clément Malgouyres Clément Mazet-Sonilhac Sara Signorelli	Technological change and domestic outsourcing

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE