A New Organizational Structure Database:

Examining Structure through Top Management Team Compositions

LIST OF ONLINE APPENDICES:

1. Overview of the initial labeling process
2. Details on fine-tuning process for generative AI model
3. Comparison of usage of different LLMs for labelling
4. Industry sector and year breakdown of data
5. Database snapshot and variable dictionary
6. Structural proxy measure development
7. Additional single firm examples from database
8. Further details on database applications
9. Summary of time spent in each stage of the data development process

Online Appendix 1. Overview of the initial labeling process

We developed sets of labeling rules to ensure a consistent role nomenclature across firms to enable firm-with-firm and year-by-year comparisons of firms' structures based on the composition of their executive teams. These rules were developed after iterative explorations of the patterns in the raw data, as well as reviews of the existing literatures suggesting leadership roles should be aligned with value chain responsibilities (primary and supporting activities) of the firm (Porter, 2008; Porter & Van der Linde, 1995), which Chandler (1964) referred to as value creating and administrative. Specifically, we developed two sets of labeling rules as described below: (1) based on the nature of the roles (e.g. primary, support, business units), and (2) based on the rank or level of the roles.

1. **Managerial role categories (each manager can have multiple categories):**

   (1) CEO – captures whether a person is the highest decision-maker of the company. This usually is the CEO, but it can also refer to the President in firm-years where a CEO is not found but a President is.

   (2) CXO – captures whether a person is a Chief Officer of the company (other than the CEO), such as the Chief Financial Officer, or the Chief Operating Officer, etc.

   (3) Primary – captures if the person is directly involved in the primary value chain of a firm, including marketing sales and operations-related activities:

   - Marketing and Sales: activities that help convince a consumer to purchase a company's product or service, as well as activities that help to maintain products and enhance consumer experience. Would include any job titles that make references to activities such as marketing, advertising, promotion, pricing, customer service, maintenance, repair, refund, exchange, warranty, etc.

- Operations: all activities involved in the production/manufacturing process of converting the raw materials into a finished product/service that are in customers' hands. Would include any job titles that make references to receiving raw materials from suppliers, packaging, branding, manufacturing, processing inputs to outputs, storage and distribution, shipping, delivery, etc.

(4) Support – captures if the person oversees activities that support the primary activities within a firm's value chain, including activities such as:

- Finance – accounting, tax, budget, risk management, audit, etc.
- Strategy – alliance, M&A, corporate development, integration, new ventures, planning, strategic transformation, etc.
- Legal – ethics, law, compliance, governance, litigation, patents, policy, regulations, trademarks
- Human Resources (HR) – compensation, personnel, learning, labor, people, culture, diversity, etc.
- Procurement – sourcing, procurement, selecting raw materials, supplier contract management; exclude inbound/outbound logistics or business unit-specific procurement roles
- Technology & Innovation – Research and Development (R&D), innovation, product development, research, etc.
- Information Technology – information systems, digital infrastructure, technical operations, etc.

- Firm Infrastructure – chief administrative officer, all stakeholder management / communication-related activities (i.e. Investor Relations, government communications, external affairs, public affairs, etc.), anything administrative that doesn't fall under Finance/Strategy/Legal/HR, management-type roles.

(5) Business Unit – captures if the person is involved with specific business units, including involvement with specific products, markets, customer groups, or geographies. Usually, business units have profit and loss responsibility for a specific part of a company.

(6) Board – captures whether a TMT member is also part of the board if such data is reported.

2. Hierarchical rank or level categories (each manager can have multiple categories):

Our dataset includes all executives reported in the sample firms' filings (e.g. annual reports, 10-Ks, or proxy statements). We assume that if firms list their executives in such high-profile filings, they play a meaningful part in firms' decision-making. In other words, rather than the researcher placing some arbitrary boundary on the number of executives in the TMT, we capture all executives listed as we believe firms' own reporting reflects each firm's specific view of what constitutes their executive team. However, many of them often have their ranks or levels as part of their job titles as disclosed in the filings. To date, there is relatively little guidance in the literature on TMT hierarchical levels besides the levels of CEO, C-Suite, and President. This is partly due to limited formal rank definitions for other titles, and firms may vary with respect to which hierarchical levels they introduce and which ones they omit. After careful inspection of our raw data, we identified 12 unique rank titles among all firms within our sample. They include:

(1) President – we made sure these are not board roles.

(2) Senior Vice Chairman – we made sure these are not board roles.

(3) Vice Chairman – we made sure these are not board roles.

(4) Senior Executive Vice President (SEVP)

(5) Executive Vice President (EVP)

(6) Senior Vice President (SVP)

(7) Vice President (VP)

(8) Senior Director – we made sure these are not board director roles.

(9) Senior Managing Director (SMD) – we made sure these are not board director roles.

(10) Director – we made sure these are not board director roles.

(11) Managing Director (MD) – we made sure these are not board director roles.

(12) Senior Executive.

Human-led and dictionary-led labeling process

The following provides an overview of our initial attempts at using a human-led and dictionary-led labeling process. However, as we describe in the manuscript and below, we abandoned this approach in favor of a Large Language Model-led approach. Nonetheless, these efforts helped us develop the labeling categories and descriptions that we provide to the Large Language Model during fine-tuning.

Upon developing the above labeling categories for the executive roles through iterative rounds of careful inspections of the raw data and group discussions among the authors, we initially undertook a human-led labeling strategy as described in Stage 3.1 of the main text. That is, we initially used the established technique that involved semi-automated matching of labels with roles based on written matching algorithms in R that would draw on an extensive dictionary of known variations of titles that fall under each of the categories. For each of the six broad categories, we also developed more detailed categories that reflect the specific activities included in each category as described above. Such algorithms would match based on perfect matches and root-word matches:

- Perfect Match: In this approach, the exact title we are looking for must be matched entirely and exclusively. This means the manager's title cannot have any extra words or letters beyond the specified title. This perfect match method is used for functional roles, CEOs of the firm, general management titles, and chairman positions.
- Keyword or Root Search: This method involves searching for specific root words within the title description. This keyword search is performed for functional roles and business unit responsibilities.

Next, we trained research assistants (RAs) who then manually reviewed the assigned categories from the R algorithms based on the dictionary instructions to manually check and edit all categories. This has been arguably the longest and most widely utilized technique to label secondary data in the social sciences. This technique is particularly powerful as humans can interpret words within a specific context, which simple and even somewhat more elaborate algorithms cannot reliably accomplish. After the RAs manually inspected all outputs from the R algorithms to check for false positives, and they then iteratively updated the rules for perfect matches and root searches to further eliminate the omission errors. The RAs then undertook manual checks and labeling of those roles with zero labels.

However, we quickly recognized the challenges of high variance in manual labeling among RAs. Furthermore, we cannot easily assess whether more categories should have been assigned to a dictionary-based labelled observation. Given the large size of the data (over 100,000 job titles) and the long list of categories to check per job title, it was also particularly difficult to assess manually how many labels are missed, especially since we may not have included all root words and terminologies in our dictionaries for each of the managerial role and rank categories. As such, after substantial experimentation with this labeling process, we ultimately decided to utilize generative AI for labeling managerial roles and ranks.

Online Appendix 2. Details on fine-tuning process for generative AI model

Fine-tuning large language models involves adjusting the parameters (weights) of a pre-trained model to optimize its performance on a specific task or domain using a user-provided training dataset. Before starting the main fine-tuning process, we conducted a few test runs to familiarize ourselves with the process and its requirements. We utilized OpenAI's available fine-tuning process for its GPT-4o-2024-08-06 model. The fine-tuning process consists of the following three steps.

Step 1: Preparation of fine-tuning dataset

The fine-tuning dataset was prepared using 4,977 cases from our main dataset. The authors labeled and iteratively revised their labeling through multiple rounds of discussions to reach a consensus on the labels for these cases. For both the training process and later the labeling of our dataset using the fine-tuned model, we provided the AI with labeling instructions and one executive’s reported role title at a time. To aid the interpretation of context, we also provided the AI with the company name and reporting year for each role title.

Step 2: Splitting data into training and validation sets

The labeled dataset was split into training and validation sets, with 75% of cases were used for training and 25% used for in-training validation. The training set is used to update the model's parameters during fine-tuning, while the validation set is used to evaluate the model's performance and prevent overfitting.

Step 3: Executing the fine-tuning process

The fine-tuning process was carried out using OpenAI's API, specifying the highest available GPT model for fine-tuning (GPT-4o at the time of this study). We did not specify the number of epochs, allowing the API to determine the optimal number of training iterations.

An epoch refers to one complete pass through the entire training dataset during the fine-tuning process. When the number of epochs is not specified, the OpenAI API automatically determines the optimal number of epochs based on the size of the training dataset and the model's performance on the validation set. In our case, the API chose to train the model for three epochs.

In a last step, we tested the fine-tuned model on an unseen holdout data set that we compiled and coded following the same principles as for the fine-tuning dataset. This holdout dataset comprised 2,010 cases and has no overlap with the 4,977 cases that were used for fine-tuning (training & verification). The fine-tuned model achieved a 98.1% accuracy in our unseen and untrained test data.

Online Appendix 3: Comparison of usage of different LLMs for labelling

In addition to the closed source GPT 4o model that we fine-tuned, we also fine-tuned two open source LLMs. The open-source nature of these models allows us to make these models publicly available for anyone to use, further fine-tune, and download (within the open-source license terms of service of the respective base model providers).

We fine-tuned two open-source large language models for our specific needs: Meta's Llama 3.1 (8 billion parameters) and Mistral's NeMo (12 billion parameters). The training process used the same 75% of our training dataset over three epochs. For this process, we utilized the Unsloth library (https://unsloth.ai), which makes training more efficient but comes with some technical limitations. We also applied Low-Rank Adaptation (LoRA-16), which adjusts only the most essential model weights rather than all parameters. This can significantly reduce computational requirements. The training was optimized for each model's size: Llama 3.1 processed 8 examples per device with immediate updates, while NeMo handled 2 examples per device and accumulated gradients over four steps to manage memory constraints. These settings balanced efficient training with model performance.

When assessing these two models against our holdout sample of 2,010 cases, we arrived at the following accuracy scores:

- Mistral Nemo 12B: We achieved an estimated accuracy of 95.9 %
- Llama 3.1 8B: We achieved an estimated accuracy of 95.3%

The models are made available on Huggingface, a machine learning community platform.

Links to base models (prior to finetuning):

https://huggingface.co/unsloth/Meta-Llama-3.1-8B-bnb-4bit

https://huggingface.co/unsloth/Mistral-Nemo-Base-2407-bnb-4bit

Links to finetuned models (by the authors), hosted on Huggingface:

https://huggingface.co/daresearch/Llama-3.1-8B-bnb-4bit-M-exec-roles

https://huggingface.co/daresearch/mistral-nemo-12b-ft-exec-roles

Online Appendix 4: Industry sector and year breakdown of data

We focused our data collection on firms that were listed in the S&P 500 index at some stage between 2003 and 2007, covering 8 of the 11 GICS sectors that we outline in the table below, as defined by the Compustat database. Our database has complete coverage of all firms that were in the S&P 500 in these years with time-series extending before and after this period for all firms for which we could collect filings (e.g. annual reports, 10-Ks, or proxy statements). This leads to a total of 161,028 executive-year observations between 1993 and 2020, over 11,658 unique firm-years and 521 unique firms as illustrated in Table OA1. It can be seen in Table OA2 that coverage drops prior to the mid-1990s. The former drop in coverage was driven by ease of access to the relevant financial filings of firms and the latter drop is due to when our data collection finished.

Table OA1: Coverage of database by GICS Industry Sectors

Sector	GICS Code	# Firms	# Firm-years	# Observations
Industrials	20	64	1,635	23,811
Consumer Discretionary	25	86	2,014	26,697
Consumer Staples	30	43	1,024	20,291
Healthcare	35	69	1,471	20,126
Financial Services	40	99	2,120	31,284
Information Technology	45	94	1,933	18,562
Communications and Media	50	29	592	5,825
Utilities	55	37	869	14,432
Total		**521**	**11,658**	**161,028**

Table OA2: Coverage of database by fiscal year

Year	# Firms in year	# Observations	Percentage	Cumulative Percentage
1993	244	4,010	2.49	2.49
1994	315	4,796	2.98	5.47
1995	364	5,411	3.36	8.83
1996	438	6,321	3.93	12.75
1997	463	6,990	4.34	17.10
1998	470	7,065	4.39	21.48
1999	487	7,396	4.59	26.08
2000	491	7,217	4.48	30.56
2001	498	7,446	4.62	35.18
2002	499	7,422	4.61	39.79
2003	495	7,404	4.6	44.39
2004	490	7,253	4.5	48.89
2005	484	7,253	4.5	53.40
2006	470	6,769	4.2	57.60
2007	460	6,460	4.01	61.61
2008	438	6,023	3.74	65.35
2009	433	5,738	3.56	68.92
2010	423	5,601	3.48	72.39
2011	418	5,441	3.38	75.77
2012	406	5,045	3.13	78.91
2013	403	5,037	3.13	82.03
2014	393	4,787	2.97	85.01
2015	377	4,564	2.83	87.84
2016	361	4,402	2.73	90.57
2017	349	4,144	2.57	93.15
2018	341	3,824	2.37	95.52
2019	330	3,652	2.27	97.79
2020	318	3,557	2.21	100.00
Total	**11,658**	**161,028**		

Online Appendix 5: Database snapshot and variable dictionary

In Table OA3 below, we provide the list of variables in our dataset and the associated explanations on how each variable has been coded.

Table OA3: Variable dictionary

Variable	Description
year	Fiscal Year
company	Company Name in Compustat
GV_KEY	Compustat gvkey identifier for firm
ticker	Company Ticker
cusip	Company Cusip (nine-digit)
GICGroups	Global Industry Classification (GIC) Groups
GICIndustries	GIC Industries
GICSectors	GIC Sectors
GICSubIndustries	GIC Sub Industries
sic_code	Standard Industry Classification
CIK	CIK (Central Index Key) identifier
role	Title of Executive in Financial Filing
last_name	Surname of Executive
first_name	First name of Executive
full_name	Full Name of Executive
uniqueid	Unique identifier for executive in a specific firm-year (unit of analysis of database
TMT Source	Where the executive data was collected: 10K – SEC 10-K Annual filing DEF14A – SEC filed company proxy statement AR – Company Annual Report
vp	1 if rank of executive is Vice President
svp	1 if rank of executive is Senior Vice President
evp	1 if rank of executive is Executive Vice President
sevp	1 if rank of executive is Senior Executive Vice President
dir	1 if rank of executive is Director
sdir	1 if rank of executive is Senior Director
md	1 if rank of executive is Managing Director
smd	1 if rank of executive is Senior Managing Director
se	1 if rank of executive is Senior Executive
vc	1 if rank of executive is Vice Chairman

Variable	Description
svc	1 if rank of executive is Senior Vice Chairman
president	1 if rank of executive is President
board	1 if an executive has a board role
ceo	1 if executive is the Chief Executive Officer of the entire firm
cxo	1 if Chief Officer e.g., COO, CFO
primary	1 if executive has a primary value chain activity role
support	1 if executive has a support role
bu	1 if executive has a business unit role

Excerpt from database

A screenshot of the database in Stata is provided for the example of Adobe in 2018 in Figure OA1. The first 11 fields identify the firm, industry, and year of a specific observation. The 12th field has the role description as extracted from the relevant financial filing. Fields 13-15 provide the specifics of the individual in the relevant role, namely first name, last name, and full name. Field 16 has the unique id for the executive in that year. Field 17 outlines the data source used to provide the pertinent observation (e.g., 10k). Fields 18-29 have dummies (0/1) that indicate whether an individual has a certain rank in their role description. Fields 30-35 have dummies (0/1) that indicate whether an individual has a specific role e.g., primary or support roles.

Figure OA1: Except from Database for Adobe in 2018

	year	company	GV_KEY	ticker	cusip	GICGroups	GICIndustri...	GICSectors	GICSubInd...	sic_code	CIK	role	last_name	first_name	full_name	uniqueid	TMTS
106049	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	executive vice president and chief marketing officer	lewnes	ann	ann lewnes	2503	10K
106097	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	chief product officer and executive vice president, creative cloud	belsky	scott	scott belsky	2500	10K
106124	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	executive vice president, worldwide field operations	thompson	matthew	matthew thompson	2510	10K
106133	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	chief human resources officer and executive vice president, employee experience	morris	donna	donna morris	2504	10K
106165	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	executive vice president and general manager, digital media	lamkin	bryan	bryan lamkin	2502	10K
106188	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	executive vice president, general counsel and corporate secretary	rao	dana	dana rao	2508	10K
106217	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	chairman, president and chief executive officer	narayen	shantanu	shantanu narayen	2506	10K
106233	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	vice president, chief accounting officer and corporate controller	garfield	mark	mark garfield	2501	10K
106260	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	executive vice president and general manager, digital experience	rencher	bradley	bradley rencher	2509	10K
106261	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	executive vice president and chief technology officer	parasnis	abhay	abhay parasnis	2507	10K
106262	2018	adobe inc	12540	ADBE	00724F101	4510	451030	45	45103010	7372	796343	executive vice president and chief financial officer	murphy	john	john murphy	2505	10K

	last_name	first_name	full_name	uniqueid	TMTSo...	vp	svp	evp	sevp	dir	sdir	md	smd	se	vc	svc	president	board	ceo	cxo	primary	support	bu
106049	lewnes	ann	ann lewnes	2503	10K	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
106097	belsky	scott	scott belsky	2500	10K	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
106124	thompson	matthew	matthew thompson	2510	10K	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
106133	morris	donna	donna morris	2504	10K	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
106165	lamkin	bryan	bryan lamkin	2502	10K	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
106188	rao	dana	dana rao	2508	10K	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
106217	narayen	shantanu	shantanu narayen	2506	10K	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
106233	garfield	mark	mark garfield	2501	10K	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
106260	rencher	bradley	bradley rencher	2509	10K	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
106261	parasnis	abhay	abhay parasnis	2507	10K	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
106262	murphy	john	john murphy	2505	10K	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0

Online Appendix 6. Structural proxy measure development

Description of variables

We develop two sets of structural variables to illustrate the potential use cases to examine different attributes of firms' structures using the database developed in this paper.

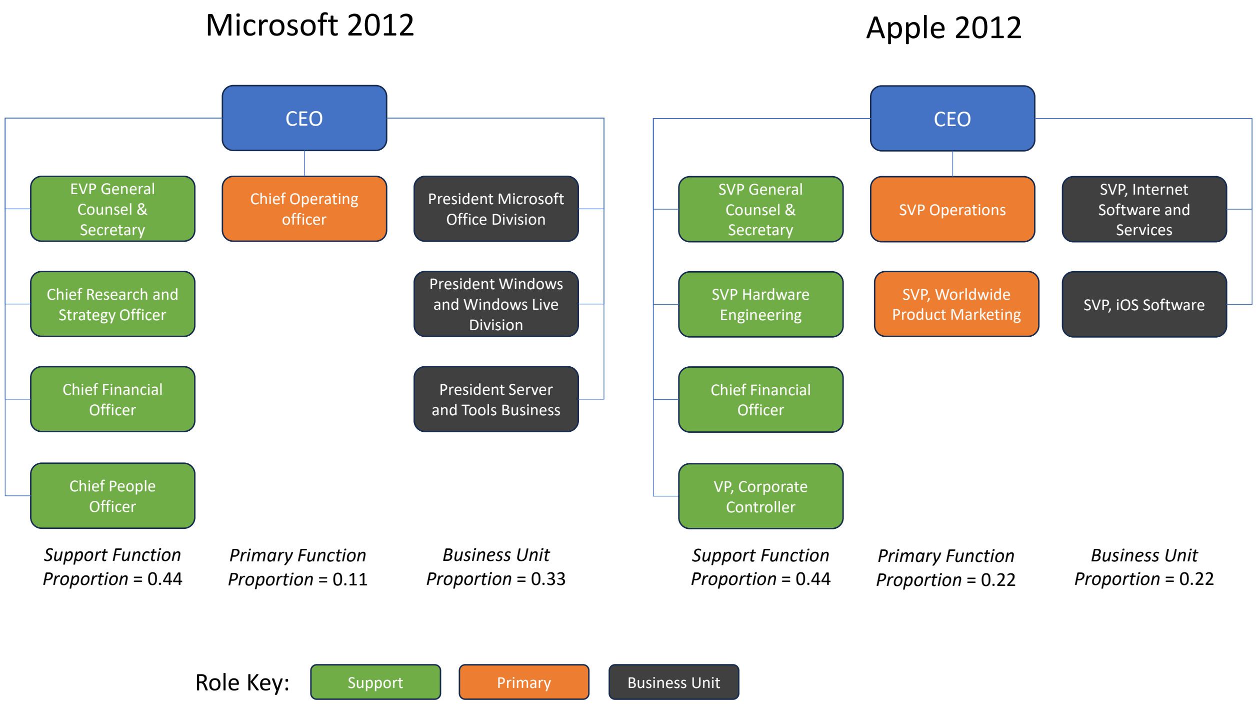
First, we measure the degree of vertical hierarchy within a senior management team (*TMT Hierarchy*). We leverage the count of the number of distinct job title levels within a management team excluding the CEO. We code the management roles into six groups in descending order of hierarchy: (1) Chief Officer (e.g. Chief Operating Officer or Chief Financial Officer); (2) President; (3) Senior Executive Vice President; (4) Executive Vice President; (5) Senior Vice President; (6) Vice President, which represent the six most commonly occurring levels among the 12 rank categories. We then count how many of these levels have an executive within the management team. For example, in 2004, retailer Lowe's has five of these levels on their executive team as illustrated in Table OA4. We also examine whether Chief Officers are in primary roles (e.g. operations) or support roles (e.g., finance).

Table OA4: Lowe's Ranks in 2004 executive team

Level	Example Role in Lowe's in 2004
Chief Officer	chief information officer
Senior Executive Vice-President	senior executive vice president — merchandising/marketing
Executive Vice-President	executive vice president, business development
Senior Vice-President	senior vice president – logistics
Vice-President	vice president, internal audit

Second, we examine the proportion of managers within a management team that relate to different roles (i.e. primary, support, business unit). Building on the work of Guadalupe, Li, and Wulf (2014), we define three types of roles within a management team. As described in Online Appendix 1, primary roles pertain to the primary value chain activities within a firm and include activities such as operations and sales and marketing. Support roles include finance, research and development, information technology etc. Finally, executive teams also include business unit roles that have defined responsibility (profit and loss related) of a distinct part of the business associated with the focal firm. We divide the number of each of these types of roles by the total size of the executive team to create the variables: *Primary Function Proportion, Support Function Proportion,* and *Business Unit Proportion*. To illustrate this measure, we compare two prominent firms: Apple and Microsoft in 2012 in Figure OA2.

Figure OA2: Microsoft and Apple Structures in 2012



Both firms have similar structures, the key difference is that Microsoft has more business unit roles but fewer primary roles.

Online Appendix 7: Additional single firm examples from database

Beyond the two examples of Hewlett Packard and General Electric presented in the main paper, we also present two additional firm examples to illustrate how the database can be used to investigate individual level firm attributes.

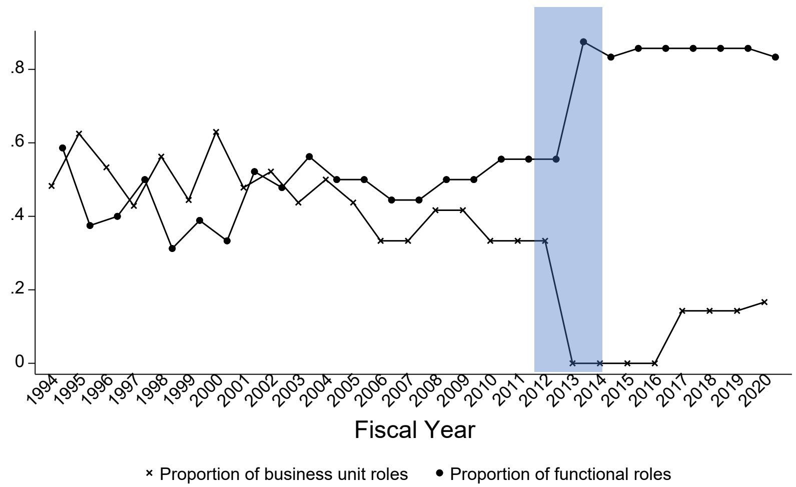
First, there is Microsoft which underwent a major reorganization in 2013 from a more decentralized, divisional structure to a functional structure. This was undertaken as part of a "One Microsoft" vision. Prior to 2013, Microsoft suffered from its largely autonomous business units operating in silos, often competing intensely with each other. Moving to a functional structure had the goal of eliminating this competition and trying to integrate Microsoft's efforts more effectively. This change can be observed in our dataset and is illustrated by the shaded area in Figure OA3(a). We also observe that after Satya Nadella became CEO in 2014, business units were reintroduced in 2017 as Microsoft increasingly focused on cloud services.[1]

Second, we also examine Disney which underwent a major structural change in 2005 when Bob Iger replaced Michael Eisner as the CEO. Iger centralized functions such as human resources, government relations and communications to gain the advantages of its scale, while still providing the firm's business units with significant autonomy. These changes can be observed in our dataset and are illustrated by the shaded area in Figure OA3 (b).

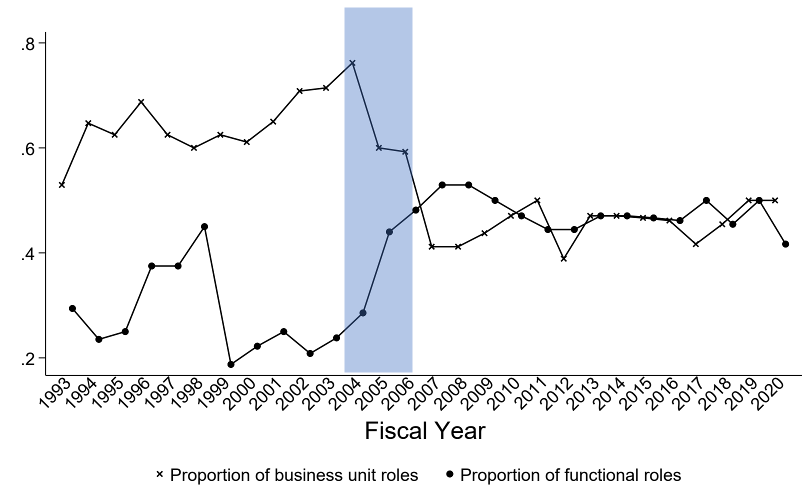
[1] https://www.linkedin.com/pulse/decade-later-why-microsofts-2013-restructuring-necessary-michael-hyzy-e62vc/

Figure OA3: Variation of (a) Microsoft; (b) Disney proportion of business unit roles and proportion of functional (support and primary) over period 1993-2020.

(a)



(b)



Online Appendix 8: Further details on database applications

Application 1: Guadalupe et al. (2014)

Guadalupe et al. (2014) utilize a proprietary dataset from a human resource consulting firm (Hewitt Associates) that has role and compensation data for senior managers in Fortune 500 firms. According to the authors (p 830), their sample "*spans the 1987–1999 period and includes approximately 300 firms, of which 69% are in manufacturing and 31% are in services*." The authors examine the relationship between diversification and the composition of firms' executive management teams. They argue that greater diversification is associated with larger management teams consisting of more business unit general managers and fewer primary managers, as such managers find it increasingly difficult to span multiple business categories.

We initially attempt to quasi-replicate the results of Table 3 in the original paper. To do so we used a subsample of our dataset spanning all firms and industries between 1993 and 1999 with 2539 observations. We use most of the controls in the original paper, replacing CAO by whether the firm has a CXO (0 or 1), and COO by whether the firm has a COO (0 or 1). We do not include the control "PCs per employee" as we do not have access to this data. Table OA5 shows the main results. We find similar results to the original study with respect to the magnitude of the *Diversification* coefficient but not the degree of statistical significance. Specifically we observe for each dependent variable—where our study provides the first coefficient and Guadalupe et al. (2014) provides the second: TMT Size 0.809 ($p=0.68$) vs. 0.741 ($p<0.10$), BU Roles 0.592 ($p=0.66$) vs. 0.625 ($p<0.05$), Functional 1.444 ($p=0.37$) vs. 0.116 ($p>0.10$), Primary -0.326 ($p=0.66$) vs. -0.262 ($p<0.05$), Support 1.770 ($p=0.10$) vs. 0.378 ($p>0.10$). For TMT size, BU Roles, and Primary the coefficients for *Diversification* are very similar in magnitude in our study to those of Guadalupe et al. (2014) but of lower statistical significance. Interestingly, the coefficients for

Functional and Support are much larger in our study. This difference may be because our Support roles cover more roles than the six roles used by Guadalupe et al. (2014: 828). The difference in our results is likely to stem from the larger sample in our study consisting of 433 firms between 1993 and 1999 as opposed to 290 firms between 1987 and 1999.

As an extension, we used the dependent variable *TMT Hierarchy* that measures the number of different hierarchical levels in a management team based on their role titles as discussed in Section 4 of the main paper. We find that the presence of a Chief Operating Officer is associated with a flatter management team (Table OA6). A Chief Operating Officer is like the right-hand manager to the CEO in an executive team potentially negating the need for a variety of more junior managers in an executive team to support the CEO (Zhang, 2006). Interestingly, the presence of any Chief Officer is associated with a more hierarchical management team, and this seems to be directionally driven by Chief Strategy and Chief Business Development Officer roles.

Table OA5: Replicating Guadalupe et al. (2014)

DV	TMT Size	BU Roles	Functional	Primary	Support
Diversification	0.809	0.592	1.444	-0.326	1.770^{+}
	(1.978)	(1.339)	(1.608)	(0.742)	(1.061)
Segment Count	0.078	0.024	-0.242	-0.195^{+}	-0.047
	(0.201)	(0.134)	(0.184)	(0.116)	(0.094)
CXO	0.368	1.028*	2.230**	0.241	1.989**
	(1.774)	(0.505)	(0.534)	(0.180)	(0.410)
COO	0.398	0.192	0.902^{+}	1.123**	-0.221
	(0.760)	(0.391)	(0.513)	(0.295)	(0.277)
% non-US Rev.	-0.638	-1.907	-1.993	-0.632	-1.360
	(2.719)	(1.627)	(1.964)	(0.558)	(1.689)
Log(Revenues)	1.820^{+}	1.308*	1.327*	0.297	1.030**
	(1.055)	(0.517)	(0.584)	(0.277)	(0.389)
R&D/Revenue	-1.633	-1.020*	-0.835	0.281	-1.116
	(1.641)	(0.488)	(1.113)	(0.329)	(0.934)
Margin	-1.203	-0.810*	-0.636	0.146	-0.783
	(1.108)	(0.328)	(0.728)	(0.219)	(0.612)
Constant	-0.177	-6.693	-3.129	0.302	-3.431
	(9.556)	(4.424)	(5.175)	(2.625)	(3.331)
Year Fixed Effect	Y	Y	Y	Y	Y
Firm Fixed Effect	Y	Y	Y	Y	Y
Number of firms	433	433	433	433	433
N	2539	2539	2539	2539	2539
R^2	0.648	0.678	0.689	0.694	0.709

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Table OA6: Extension to Guadalupe et al. (2014) what predicts *TMT Hierarchy*

DV	TMT Hierarchy 1993-1999	TMT Hierarchy 1993-2020
Diversification	0.187	0.073
	(0.170)	(0.116)
Segment Count	0.005	0.001
	(0.018)	(0.011)
CXO	**1.061****	**1.052****
	(0.084)	**(0.102)**
COO	**-0.336****	**-0.251****
	(0.058)	**(0.032)**
% non-US Rev.	-0.876*	-0.064
	(0.375)	(0.200)
Log(Revenues)	0.046	-0.003
	(0.068)	(0.036)
R&D/Revenue	-0.143	-0.094*
	(0.159)	(0.037)
Margin	-0.049	0.006
	(0.106)	(0.015)
Constant	2.505**	2.576**
	(0.594)	(0.326)
Year Fixed Effect	Y	Y
Firm Fixed Effect	Y	Y
N	2539	11233
R^2	0.692	0.516

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Application 2: Zhang (2006)

Drawing from the behavioral theory of the firm, Zhang (2006) examines how prior performance is associated with the degree of strategic change made by the senior managers of a firms. Further, Zhang (2006) highlights that this negative correlation between prior performance and the degree of strategic change is negatively moderated by the presence of a COO. Namely, a COO will magnify the degree of strategic change made by a CEO and their team for under-performing firms.

We initially attempt to quasi-replicate the results of Table 2 in the original paper with most of the controls used in the original paper and using firm and year fixed effects (see Table OA7). The initial study utilized a sample of 187 non-diversified manufacturing firms between 1993 and 1998. We attempt to replicate this sample by focusing on firms between 1993 and 1998, with *Diversification* less than 0.64 (less diversified firms) and exclude the IT and Financial Services industries from our dataset. This provides us with a sample of 684 firm-years. As in the original study, we lagged all independent variables 1 year from the dependent variable. Table OA8 illustrates our results. We observe that lagged performance has a negative association with the degree of strategic change, consistent with Zhang (2006). The Coefficient for *Prior Firm Performance* is -0.204 ($p<0.01$) in the original study and -0.457 ($p=0.035$) in this study. However, we do not observe the negative interaction between *Chief Operating Officer in TMT* and *Prior Firm Performance* in our replication study, instead observing a positive coefficient that is not significantly different from zero. When we used a contemporaneous COO variable, we do observe a negative interaction between *Chief Operating Officer in TMT* and *Prior Firm Performance* (-0.299, $p=0.267$ for the non-diversified sample; -0.298, $p=0.092$ for all sample firms in period 1993-1998). This may be due to differences in how Zhang's (2006) data sources of COO data (e.g., the Dun & Bradstreet Reference Book of Corporate Management) have applied timestamps compared to our data, which mainly comes from annual reports linked to financial years.

In Table OA8, we now examine how the presence of a Chief Strategy Officer in a management team can shape a firm's degree of strategic change in response to prior performance. We use our full available data sample between 1993-2020 to undertake this analysis. Interestingly, we observe that the presence of a Chief Strategy Officer positively moderates the relationship between prior firm performance and strategic change (see Model 2). This suggests that a Chief Strategy Officer dampens the magnitude of strategic change when firms under-perform and magnifies it when they over-perform. This is a valuable insight, potentially CSOs help management teams avoid over-reacting to poor performance and ensure that firms do not rest on their laurels when firms are performing well.

Table OA7: Analysis of relationship between strategic change and performance with presence of chief operating officer (Zhang, 2006)

.

DV = Strategic Change	1	2
Prior Firm Performance (lag ROA)	**-0.457***	**-0.475***
	(0.215)	**(0.227)**
Chief Operating Officer in TMT	**0.004**	**0.004**
	(0.015)	**(0.015)**
Prior Firm Performance x Chief Operating Officer in TMT		**0.094**
		(0.316)
Dual CEO Chairman	-0.014	-0.014
	(0.015)	(0.015)
CEO tenure	0.001	0.001
	(0.001)	(0.001)
CEO age64up	-0.117**	-0.118**
	(0.018)	(0.018)
Size (log assets)	-0.032	-0.033
	(0.058)	(0.058)
# board members in TMT	0.016	0.016
	(0.015)	(0.015)
# TMT members	-0.000	-0.000
	(0.001)	(0.001)
Diversification	0.311*	0.315*
	(0.152)	(0.152)
Constant	-0.332	-0.334
	(0.339)	(0.340)
Firm fixed effects	Y	Y
Year fixed effects	Y	Y
N	684	684
R^2	0.555	0.555

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Table OA8: Analysis of relationship between strategic change and performance with presence of chief strategy officer

DV = Strategic Change	1	2
Prior Firm Performance (lag ROA)	**-0.267****	**-0.305****
	(0.050)	**(0.061)**
Chief Strategy Officer in TMT	**-0.086**	**-0.078**
	(0.057)	**(0.057)**
Prior Firm Performance x Chief Strategy Officer in TMT		**0.158**$^{+}$
		(0.087)
Dual CEO Chairman	0.042	0.043
	(0.041)	(0.041)
CEO tenure	-0.002	-0.002
	(0.002)	(0.002)
CEO age6163	0.136	0.136
	(0.085)	(0.085)
CEO age64up	0.041	0.042
	(0.058)	(0.058)
Size (log assets)	0.025	0.025
	(0.031)	(0.031)
# board members in TMT	0.000	0.001
	(0.014)	(0.014)
# TMT members	0.001	0.001
	(0.001)	(0.001)
Diversification	0.032	0.032
	(0.052)	(0.052)
Constant	0.017	0.017
	(0.122)	(0.122)
Firm fixed effects	Y	Y
Year fixed effects	Y	Y
N	9413	9413
R^2	0.425	0.425

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Application 3: Menz and Scheef (2014)

In this paper the authors argue and find that firms that undertake a greater number of strategic actions, such as acquisitions and are more diversified are more likely to have a Chief Strategy Officer in their executive teams. The general theoretical arguments are if firms are more complex and undertake more corporate strategies then Chief Strategy Officers will be needed to support the CEO.

The original study sample consisted of 147 randomly selected firms from the S&P500 (excluding financial services) over the period 2004 to 2008. Our sample includes all firms in our dataset between 2004 and 2008 excluding financial services firms. Table OA9 illustrates our quasi-replication of Table 2 Models 1 and 2 in Menz and Scheef (2014). Focusing on column 2, we observe only that firms in lower performing industries are less likely to have CSOs. This is different to the results observed by Menz and Scheef (2014) who find that firms that are more diversified, undertake more acquisitions, have greater TMT role interdependence (equivalent to our TMT functional diversity measure), have dual CEO-chairman, less tenured CEOs, do not have COOs, and in more recent years are more likely to have a Chief Strategy Officer. These differences are likely to arise due to differences in samples used with our sample consisting of a larger number of observations. In Model 3 in Table OA9, we use our complete sample over this period representing 506 unique firms. Now we observe that lower performing firms are associated with the absence of a CSO and CSOs have increased in prevelance over time which is consistent with the original study.

In extending the work of Menz and Scheef (2014) we examine other factors that are associated with the likeihood of a firm having a Chief Strategy Officer. As can be seen in Table OA10, having a greater proportion of the TMT on a firm’s board is associated with a lower likelihood of a CSO. This association is much stronger if we use our full sample and do not lag the *Board Proportion* variable as was undertaken in the original study. Model 1 is using our closest sample to that used in the original study of all non-financial service S&P500 firms between 2004 and 2008. Model 1 uses our complete sample with a one-year lagged *Board Proportion* variable. Model 2 uses our full sample and a lagged value of *Board Proportion.* Model 3 uses our complete sample with a non- lagged *Board Proportion* variable. We suggest that this results arises because if more executives in a firm's top management team are on the board they will have more access to directors in developing their strategies thereby reducing the need for a CSO.

We also leverage our dataset to extend the analysis by focusing on firms' corporate development executives (or Corporate Development Officers -CDOs) responsible for managing the M&A process (Table OA11). Corporate development involves merger and acquisition strategy as well as exploring alliance opportunities, which are often part of the Chief Strategy Officer role. We find that firms that undertake more acquisitions and that have less functionally diverse management teams are more likely to have such corporate development roles in a firm's TMT. These findings highlight opportunities for future work examining the interactions between Chief Strategy Officers and corporate development officers, and their presence on the TMT over time as firms' strategies evolve.

Table OA9: Logit analysis of determinants of the presence of a Chief Strategy Officer (CSO) in a firms' top management teams (Menz & Scheef, 2014).

DV= CSO	1	2	3
Diversification		**-0.247**	**0.149**
		(1.191)	**(0.587)**
Acquisition Activity		**0.235**	**0.102**
		(0.444)	**(0.170)**
Firm Size		**0.072**	**0.040**
		(0.227)	**(0.118)**
TMT Functional Diversity		**-0.602**	**-1.083**
		(1.874)	**(0.812)**
Industry Performance	-18.131	-18.737^{+}	-2.987
	(11.613)	(10.933)	(3.565)
Firm Performance	-0.757	-0.652	-0.823^{*}
	(1.600)	(1.671)	(0.338)
CEO duality	0.264	0.227	-0.450
	(0.575)	(0.577)	(0.286)
CEO position tenure	-0.048	-0.045	-0.007
	(0.040)	(0.041)	(0.017)
COO	-0.718	-0.690	-0.123
	(0.589)	(0.601)	(0.262)
Year	0.112	0.099	0.116^{**}
	(0.179)	(0.199)	(0.018)
Constant	-228.460	-201.652	-237.128^{**}
	(359.404)	(398.785)	(35.850)
N	1844	1829	9882
Log Likelihood	-119.618	-118.881	-1033.755

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Table OA10: Logit analysis of other determinants of the presence of a Chief Strategy Officer (CSO) in a firms' top management teams (Menz & Scheef, 2014).

DV= CSO	1	2	3
Board proportion	**-0.610**	**-3.106**	**-4.559***
	(3.388)	**(2.147)**	**(1.988)**
Diversification	-0.262	0.122	0.123
	(1.231)	(0.590)	(0.590)
Acquisition Activity	0.237	0.117	0.127
	(0.451)	(0.171)	(0.172)
Firm Size	0.068	0.034	0.036
	(0.233)	(0.118)	(0.118)
TMT Functional Diversity	-0.523	-0.624	-0.565
	(2.069)	(0.941)	(0.894)
Industry Performance	-18.732^{+}	-3.249	-3.284
	(10.901)	(3.523)	(3.492)
Firm Performance	-0.653	-0.850*	-0.850*
	(1.662)	(0.355)	(0.339)
CEO duality	0.270	-0.219	-0.172
	(0.551)	(0.336)	(0.306)
CEO position tenure	-0.045	-0.005	-0.004
	(0.041)	(0.016)	(0.017)
COO	-0.694	-0.146	-0.159
	(0.592)	(0.259)	(0.261)
Year	0.099	0.116**	0.115**
	(0.199)	(0.018)	(0.018)
Constant	-202.761	-236.355**	-234.054**
	(399.811)	(35.966)	(36.264)
N	1829	9882	9882
Log Likelihood	-118.864	-1029.949	-1025.183

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Table OA11: Corporate Development Officer (CDO) Prediction using a logit model.

DV= CDO	1	2
Diversification	**0.036**	**0.345**
	(0.603)	**(0.446)**
Acquisition Activity	**0.365^{+}**	**0.269^{+}**
	(0.204)	**(0.143)**
Firm Size	**-0.141**	**-0.062**
	(0.113)	**(0.070)**
TMT Functional Diversity	**-3.388****	**-3.478****
	(0.976)	**(0.549)**
Industry Performance	6.627	2.353
	(5.413)	(1.763)
Firm Performance	-0.073	0.321
	(1.031)	(0.541)
CEO duality	-0.187	-0.142
	(0.232)	(0.150)
CEO position tenure	0.000	0.013
	(0.021)	(0.012)
COO	0.226	0.108
	(0.233)	(0.132)
Year	-0.131**	-0.034**
	(0.051)	(0.010)
Constant	264.499**	68.475**
	(101.532)	(19.899)
N	1829	9882
Log Likelihood	-700.029	-3500.148

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Summary of variables used in application analyses

Table OA12: Description of variables used in application analyses

Variable	How estimated	Data sources
Application 1: Guadalupe et al. (2014)		
TMT Size	Number of executives in firms' TMTs in firm-year	This database
BU Roles	Number of business unit roles in executive team in firm-year	This database
Functional	Number of functional (primary and support) roles in executive team in firm-year	This database
Primary	Number of primary functional roles in executive team in firm-year	This database
Support	Number of support functional roles in executive team in firm-year	This database
Diversification	Estimate using a Herfindahl-Hirschman measure of firms' sales across different industry segments (using SIC codes). This Herfindahl-Hirshman index is then subtracted from one to obtain diversification measure	Compustat
Segment Count	Count of number of operating segments	Compustat
CXO	A binary variable indicating whether a firm has a Chief Officer (e.g. Chief Financial officer) in its executive team	This database
COO	Whether firm has a Chief Operating Officer in its executive team in firm-year	This database
% non-US Rev.	Percentage of a firms' sales from non-US geographical segments	Compustat
Log(Revenues)	Log of a firm's annual sales	Compustat
R&D/Revenue	Annual research and development expenditure divided by annual revenues	Compustat
Margin	Annual net income divided by annual revenues	Compustat
TMT Hierarchy	Measures the number of different hierarchical levels in a management team based on their role titles. This can vary between 1 and 6.	This database
Application 2: Zhang (2006)		
Strategic Change	Developed using 6 resource allocation measures (e.g. research and Development) as outlined in Zhang (2006).	Compustat
Prior firm performance	One-year-lagged return on assets (ROA)	Compustat
Chief Operating Officer in TMT	Whether firm has a Chief Operating Officer in its executive team in firm-year	This database
Dual CEO Chairman	Whether CEO is also the Chairman of a firm in the pertinent year	This database
CEO tenure	Number of years CEO has been in the CEO role in the focal year	ExecuComp
CEO age6163	Binary variable stating if CEO is 61-63 or not	ExecuComp

Variable	How estimated	Data sources
CEO age64up	Binary variable stating if CEO is 64 or older	ExecuComp
Size (log assets)	Natural log of firms' total asset value	Compustat
# board members in TMT	Number of executives in TMT that are also board members	This database
# TMT members	Number of executives in firms' TMTs in firm-year	This database
Diversification	Estimate using a Herfindahl-Hirschman measure of firms' sales across different industry segments (using SIC codes). This Herfindahl-Hirshman index is then subtracted from one to obtain diversification measure	Compustat
Chief Strategy Officer in TMT	Whether firm has a Chief Strategy Officer in its executive team in firm-year	This database
Application 3: Menz and Scheef (2014)		
CSO	Whether firm has a Chief Strategy Officer in its executive team in firm-year	This database
Diversification	Estimate using a Herfindahl-Hirschman measure of firms' sales across different industry segments (using SIC codes). This Herfindahl-Hirshman index is then subtracted from one to obtain diversification measure	Compustat
Acquisition Activity	Log of the count of completed majority owned M&A deals in focal year and prior two years	SDC Platinum
Firm Size	Natural log of firms' annual revenues	Compustat
TMT Functional Diversity	Estimate using a Herfindahl-Hirschman measure of executive teams' proportion of executives in BU, primary functional, and support functional roles. This Herfindahl-Hirshman index is then subtracted from one to obtain TMT Functional Diversity measure	This database
Industry Performance	Mean annual ROA across industry as defined using 2-digit GICS sector codes.	Compustat
Firm Performance	Return on Assets in focal year	Compustat
CEO duality	Whether CEO is also the Chairman of a firm in the pertinent year	This database
CEO position tenure	Number of years CEO has been in the CEO role in the focal year	ExecuComp
COO	Whether firm has a Chief Operating Officer in its executive team in firm-year	This database
Year	Year of data observation e.g. 2012	Compustat
Board Proportion	Proportion of executives in TMT that are also board members	This database
CDO	Presence of Chief Development Officer in TMT in firm-year	This database

Online Appendix 9: Summary of time spent in each stage of the data development process

This dataset is unique in that it requires significant perseverance to collect, as firms tend to report their executive teams in a variety of different filings such as their 10-Ks, proxy statements or annual reports and in a variety of different format within each type of filing. As such, the data development process required a non-trivial amount of time and resources. It took 5 years (about 5,800 hours) to collect and develop this database, with the research team spending over USD 60,000 in RA and AI-related expenses. Table OA13 below summarizes the approximate time spent for each stage of the process.

Table OA13: Time spent in each stage of the data development process

Stage	**Hours spent**
1. Study Plan	~60 hours
2. Data Collection	~3,000 hours
3.1 Initial human- and dictionary-led Labeling	~400 hours
3.2 Generative AI Labeling	~100 hours
4. Analysis and Data Complementation	~2,200 hours
Total	**~5,800 hours**

Online Appendix References

Chandler, A. D. 1964. Strategy and structure. In N. Foss, J. (Ed.), ***Resources, Firms, and Strategies***: 40-51. Oxford, UK: Oxford University Press.

Guadalupe, M., Li, H. Y., & Wulf, J. 2014. Who Lives in the C-Suite? Organizational Structure and the Division of Labor in Top Management. ***Management Science***, 60(4): 824-844.

Menz, M., & Scheef, C. 2014. Chief strategy officers: Contingency analysis of their presence in top management teams. ***Strategic Management Journal***, 35(3): 461-471.

Porter, M. E. 2008. ***Competitive advantage: Creating and sustaining superior performance***. NYC: Simon and Schuster.

Porter, M. E., & Van der Linde, C. 1995. Toward a new conception of the environment-competitiveness relationship. ***Journal of economic perspectives***, 9(4): 97-118.

Zhang, Y. 2006. The presence of a separate COO/president and its impact on strategic change and CEO dismissal. ***Strategic Management Journal***, 27(3): 283-300.