Online Appendix

This appendix is not intended for publication with the rest of the paper. The first section provides detailed information on the algorithm that extracts the family links from the proxy statements. The second section presents some examples of proxy statements to clarify my methodology. The third section reports additional robustness tests not included in the main paper.

A.   Procedure for extracting family links

I measure the number of family links disclosed in the proxy statements filed by U.S. public firms by seeking key words indicating relationships. I take into account only those forms in EDGAR whose filing type is equal to DEF 14A.[10] The key words that I consider are wife, husband, ex-wife, ex-husband, son, son-in-law, daughter, daughter-in-law, father, father-in-law, mother, mother-in-law, brother, brother-in-law, sister, sister-in-law, grandfather, grandmother, cousin, nephew, niece, aunt, and uncle.

In order to extract links, I code an algorithm that operates in the following way. First, it starts from the files downloaded from the EDGAR database and cleans them from any kind of encoding-related wording. For example, any HTML tags, special characters, and text strings originated from the presence of images and charts in the original files. It also drops any dot that appears in the file and, presumably, does not mark the end of a sentence (e.g., "Mr.", "Ms.", "Dr.", "Inc.", "A. Smith", "18.97"). This step is important for my procedure, as I define a sentence as anything that stands between two dots. If the same family link is reported twice or more within the same sentence, the algorithm counts it only once, as it is likely that it refers to the same relationship.

Second, the algorithm removes from the filings all sentences that contain groups of words

---

[10]The word "DEF" stands for definitive, or final, proxy statement. 14A refers to the fact that proxy statements are filed pursuant to Section 14(a) of the Securities Exchange Act of 1934.

from a predefined set of "false positives." These standard wordings contain some of the keywords I use to identify family links in my analysis but do not disclose the actual presence of a relative. One of the most frequent among these false positives is the following formulation: "For purposes of these procedures, immediate family members means any child, stepchild, parent, stepparent, spouse, sibling, mother-in-law, father-in-law, son-in-law, daughter-in-law, brother-in-law or sister-in-law, and any person (other than a tenant or employee) sharing the household with the executive officer, director or 5% beneficial owner." This sentence mentions several family links but does not actually disclose the presence of any, as it is a common phrasing for introducing one of the sections of the proxy statement. I therefore exclude all similar sentences before proceeding with counting the number of links.

Third, my algorithm excludes sentences referring to stock ownership or financial transactions. A threat to the identification arises because, together with the relatives employed by the firm, the proxy statements disclose the ownership of shares of the firm and related financial transactions by family members of current directors (regardless of whether they are employed by the firm or not). For example, the most popular family link in the sample by simple word counting is "wives." However, most of the sentences in the proxy statements regarding wives do not concern employment but stock ownership, as item 404(a) of Regulation S-K requires to disclose it. To avoid that my procedure erroneously classifies stock ownership by a relative as a family link, the algorithm removes also all sentences that include words indicating financial ownership (e.g., *stock*, *shares*, *owned by*, *held by*).

Finally, the algorithm counts the number of times each family link is mentioned in the remaining part of the filings. I manually check one by one a random selection of firms to verify the accuracy of the methodology. In this way, I attempt to minimize the risk of miscounting.

B. Examples of proxy statements

To clarify my methodology to count links, I present in the following some brief extracts from proxy statements filed by three firms. Note that I count the number of relationships rather than the number of employed relatives.

MSG Networks, October 27, 2020 (...) Marianne Dolan Weber is the *daughter* of Charles F. Dolan, the *sister* of James L. Dolan and Thomas C. Dolan, the *sister-in-law* of Brian G. Sweeney and Kristin A. Dolan, the *cousin* of Paul J. Dolan and the *aunt* of Charles P. Dolan, Quentin F. Dolan and Ryan T. Dolan. (...) our Board has concluded that Marianne Dolan Weber should serve as a director of the Company. (...) Paul J. Dolan is the *nephew* of Charles F. Dolan, the *cousin* by marriage of Brian G. Sweeney and Kristin A. Dolan and the *cousin* of James L. Dolan, Thomas C. Dolan, Marianne Dolan Weber, Charles P. Dolan, Quentin F. Dolan and Ryan T. Dolan. (...) Mr. [Ryan] Dolan is the *son* of James L. Dolan, the *stepson* of Kristin A. Dolan, the *brother* of Charles P. Dolan and Quentin F. Dolan, the *grandson* of Charles F. Dolan, the *nephew* of Marianne Dolan Weber, Thomas C. Dolan and Brian G. Sweeney and the *cousin* of Paul J. Dolan. (...) Mr. [Thomas] Dolan is the *son* of Charles F. Dolan, the *brother* of James L. Dolan and Marianne Dolan Weber, the *brother-in-law* of Brian G. Sweeney and Kristin A. Dolan, the *cousin* of Paul J. Dolan and the *uncle* of Charles P. Dolan, Quentin F. Dolan and Ryan T. Dolan. (...) Brian G. Sweeney is the *son-in-law* of Charles F. Dolan, the *brother-in-law* of James L. Dolan, Marianne Dolan Weber, Thomas C. Dolan and Kristin A. Dolan, the *cousin* of Paul J. Dolan and the *uncle* of Charles P. Dolan, Quentin F. Dolan and Ryan T. Dolan.

The New York Times, March 17, 2014 Proxy Statement. (...) Arthur Sulzberger, Jr. was employed as Chairman of the Company and Publisher of The New York Times (...) Samuel Dolnick was employed as a staff reporter and deputy sports editor for The New York Times (...) Michael Greenspon, (...) was employed as general manager, news services division

(...) Rachel G. Kirscht was employed as a manager in marketing (...) David Perpich,(...) was employed as vice president, product management and general manager, (...) Arthur Gregg Sulzberger was employed as an assistant editor for The New York Times (...) Samuel Dolnick is the *son* of Michael Golden's *sister*. James Dryfoos and Michael Greenspon are each the *son* of a *cousin* of Arthur Sulzberger, Jr. and Michael Golden, and Michael Greenspon is Carolyn D. Greenspon's *brother*. Rachel G. Kirscht is Michael Golden's *daughter*. David Perpich is the *son* of Arthur Sulzberger, Jr.'s *sister* and Arthur Gregg Sulzberger is Arthur Sulzberger, Jr.'s *son*.

John B Sanfilippo & Son, September 28, 2011 Proxy Statement. (...) Michael J. Valentine, Chief Financial Officer (...) is the *son* of Mathias A. Valentine, a director of our company, the *brother* of James A. Valentine, an executive officer of our company, the *nephew* of Jasper B. Sanfilippo, a director of our company, and *cousin* of Jasper B. Sanfilippo, Jr. and Jeffrey T. Sanfilippo, both of whom are executive officers and directors of our company. Michael J. is also a first *cousin* by marriage of Timothy R. Donovan, a director of our company. (...) Roseanne Christman, Director of Creative Services and Customer Solutions of the company, is the *sister-in-law* of Timothy R. Donovan. (...) Jasper B. Sanfilippo is the *father* of (and Mathias A. Valentine is the *uncle* of) Jeffrey T. Sanfilippo and Jasper B. Sanfilippo, Jr. Mathias A. Valentine is the *father* of (and Jasper B. Sanfilippo is the *uncle* of) Michael J. Valentine and James A. Valentine.
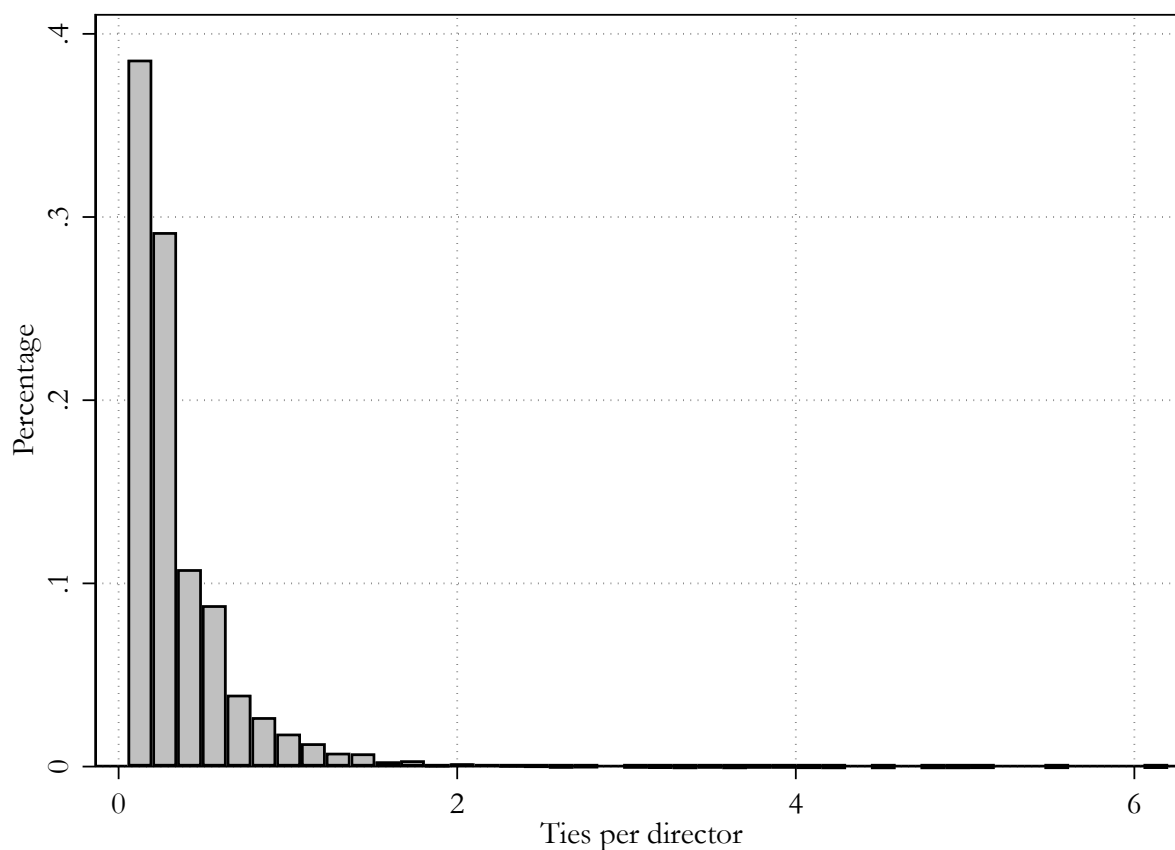
C. Additional Tables and Figures

Figure A.1: Distribution of links

Notes: This figure illustrates the distribution of family links scaled by the number of directors for firms that disclose at least one family link.

Table A.1: Alternative definitions of family-run firms

	ROA (1)	ROA (2)	ROA (3)	ROA (4)
LinksDirector	0.058*** (0.007)			
I(Links)>0		0.031*** (0.004)		
Links			0.006*** (0.001)	
FF90				0.050*** (0.005)
Controls	Y	Y	Y	Y
Sector FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Observations	55,672	55,672	55,672	55,672
Adj R2	0.241	0.240	0.240	0.241
Mean D. Var.	-0.040	-0.040	-0.040	-0.040

Notes: **This table shows regressions of firm ROA on different proxies of family-run firms.** ***LinksDirector*** **is the number of family links scaled by the number of directors;** ***I(Links)>0*** **is a dummy variable that takes a value of 1 if there is at least one family link disclosed;** ***Links*** **is the number of family links unadjusted; and** ***FF90*** **is a dummy variable that takes a value of 1 if the number of disclosed family links scaled by the number of possible links is in the top 10% of the annual distribution. Control variables include** *Size*, *Tangibility*, **and** *Leverage*. **All accounting variables are winsorized at the 1% level and defined in Table A. The bottom row reports the mean of the dependent variable. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.**

Table A.2: Possible family links and q

	q (1)	q (2)	q (3)
Plinks	0.001 (0.001)		
High plinks		-0.012 (0.029)	
Low plinks			-0.049 (0.033)
Controls	Y	Y	Y
Sector FE	Y	Y	Y
Time FE	Y	Y	Y
Observations	55,672	55,672	55,672
Adj R2	0.130	0.130	0.130
Mean D. Var.	2.084	2.084	2.084

Notes: This table shows a regression of firm q on the number of possible family links (*plinks*). Possible family links are defined as $p = \dfrac{n!}{2!(n-2)!}$, where n is the number of directors. The independent variable is a dummy that takes value of 1 if *Plinks* is in the top (bottom) quintile of the annual distribution in Column 2 (Column 3). Control variables include *Size*, *ROA*, *Tangibility*, and *Leverage*. All accounting variables are winsorized at the 1% level and defined in Table A. The bottom row reports the mean of the dependent variable. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.3: Wage differential

	Wages (1)
FF	-0.094***
	(0.031)
Observations	3,443

Notes: This table shows a *t*-test for the hypothesis that FFs pay the same wages to their workers as non-FFs. *FF* is a dummy variable that takes a value of 1 if the number of disclosed family links scaled by the number of possible links is in the top 20% of the annual distribution. Wages is staff expenses scaled by the number of employees. *** indicates statistical significance at the 1% level.

Table A.4: Pre-dynastic firms

	ROA (1)	R&D (2)	COGS (3)	Cost of debt (4)	Productivity (5)	Payout (6)
$FF \times Founder\ in$	-0.007	-0.012	0.256	-0.009	-0.269	0.002
	(0.011)	(0.010)	(0.170)	(0.015)	(0.313)	(0.002)
FF	0.035***	-0.021***	-0.303***	-0.006	-0.198	0.002***
	(0.005)	(0.005)	(0.075)	(0.006)	(0.208)	(0.001)
$Founder\ \ in$	-0.020***	0.040***	-0.100	0.016*	0.188	-0.003***
	(0.007)	(0.006)	(0.107)	(0.009)	(0.172)	(0.001)
Controls	Y	Y	Y	Y	Y	Y
Sector FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Observations	35,647	22,557	35,647	28,423	35,500	35,647
Adj R2	0.187	0.330	0.086	0.061	0.188	0.037
Mean D. Var.	-0.040	0.105	1.189	0.121	4.027	0.011

Notes: **This table shows regressions of firm outputs on** FF **and** $Founder\ in$. FF **is a dummy variable that takes a value of 1 if the number of disclosed family links scaled by the number of possible links is in the top 20% of the annual distribution.** $Founder\ in$ **is a dummy variable that takes a value of 1 if the founder holds a position in the firm. Control variables include** $Size$, $Tangibility$, **and** $Leverage$. **All accounting variables are winsorized at the 1% level and defined in Table A. The bottom row reports the mean of the dependent variable. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.**

Table A.5: Matching

Panel A: Family-run firms

	ROA (1)	q (2)	R&D (3)	Payout (4)	Mkt share (5)
FF	0.045***	-0.113*	-0.032***	0.001	-0.000
	(0.008)	(0.059)	(0.007)	(0.001)	(0.001)
Controls	Y	Y	Y	Y	Y
Sector FE	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y
Observations	22,245	22,245	15,585	22,245	22,245
Adjusted R2	0.291	0.144	0.340	0.018	0.242
Mean D. Var.	-0.081	2.358	0.148	0.011	0.006

Panel B: Founder-CEO firms

	ROA (1)	q (2)	R&D (3)	Payout (4)	Mkt share (5)
FCEO	0.000	0.040	0.010	0.000	-0.001
	(0.013)	(0.102)	(0.010)	(0.001)	(0.001)
Controls	Y	Y	Y	Y	Y
Sector FE	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y
Observations	6,621	6,621	5,293	6,621	6,621
Adj R2	0.342	0.146	0.402	0.031	0.319
Mean D. Var.	-0.069	2.783	0.173	0.007	0.003

Notes: This table shows regressions of firm outputs on *FF* (Panel A) and *Founder CEO* (Panel B). *FF* is a dummy variable that takes a value of 1 if the number of disclosed family links scaled by the number of possible links is in the top 20% of the annual distribution. *FCEO* is a dummy variable that takes a value of 1 if the founder is the current CEO. Samples are based on a coarsened exact matching estimator that selects control firms among those that do not disclose any family link (Panel A) or those in which the current CEO is not the founder (Panel B), operate in the same industry defined using SIC 4-digit codes and are the same age. If multiple control firms exist, only the one closest in size is retained. Control variables include *Size*, *Tangibility*, and *Leverage*. All accounting variables are winsorized at the 1% level and defined in Table A. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.