

Appendices

Contents

A	Supplemental empirical analysis	2
A.1	Expanded regression results and sample robustness checks	2
A.2	Correlates of network concentration	9
A.3	Observable and unobservable variables	15
A.4	The role of initial firm size	18
A.5	The role of prior employer size	21
A.6	Occupation compositional effects	23
A.7	Sectoral differences	24
A.8	Top manufacturing industries	31
A.9	Additional robustness checks	34
A.10	HHI and share-based concentration measures - Examples	37
A.11	RAIS data cleaning and sample selection	39

A Supplemental empirical analysis

A.1 Expanded regression results and sample robustness checks

Table A1: Expanded results – specification A from table 3

	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
	(1)	(2)	(3)	(4)	(5)
HHI>0	-.14 (.01)***	-.09 (.02)***	.01 (.005)**	-.06 (.006)***	-.007 (.003)**
<i>Human capital controls</i>					
Share w/ formal experience	-.18 (.02)***	-.09 (.03)***	.34 (.01)***	.23 (.02)***	-.07 (.006)***
Share from unemployment	.30 (.02)***	.13 (.03)***	-.11 (.009)***	-.14 (.02)***	.01 (.006)**
Share from same sector	-.19 (.03)***	-.17 (.04)***	.05 (.01)***	.02 (.01)**	-.01 (.004)***
Mean employee age (/10)	.008 (.009)	.01 (.01)	.06 (.006)***	-.02 (.006)***	-.05 (.004)***
Share close to retirement	-.09 (.06)	-.01 (.13)	-.06 (.03)**	-.12 (.04)***	.13 (.02)***
Mean years of schooling	-.01 (.003)***	-.001 (.003)	.02 (.003)***	.008 (.002)***	-.006 (.0006)***
Log mean previous wages	-.04 (.007)***	-.02 (.01)*	.32 (.01)***	.05 (.005)***	-.05 (.003)***
Share female workers	.02 (.02)	.08 (.03)**	-.06 (.008)***	-.02 (.02)	.03 (.006)***
<i>Prior employer size controls</i>					
Unobserved	.02 (.02)	.07 (.04)*	.07 (.004)***	.05 (.01)***	-.01 (.008)
Medium (10 to 100)	.006 (.009)	-.002 (.02)	-.02 (.003)***	.07 (.007)***	.01 (.003)***
Large (100 to 1,000)	.08 (.01)***	.03 (.02)*	-.03 (.004)***	.10 (.009)***	.01 (.004)***
Very large (1,001+)	.06 (.02)***	.02 (.04)	-.04 (.006)***	.07 (.01)***	.01 (.004)**
<i>Initial size controls</i>					
3 to 4 initial empl.	-.11 (.009)***	-.02 (.02)	.05 (.003)***	-.13 (.007)***	-.002 (.003)
5 to 10 initial empl.	-.19 (.02)***	-.02 (.02)	.10 (.004)***	-.28 (.01)***	-.007 (.004)*
11 to 20 initial empl.	-.19 (.03)***	-.03 (.03)	.14 (.006)***	-.44 (.02)***	-.01 (.004)***
21 to 50 initial empl.	-.22 (.04)***	.05 (.05)	.18 (.01)***	-.58 (.03)***	-.02 (.005)***
51 to 100 initial empl.	-.29 (.05)***	.04 (.05)	.23 (.02)***	-.73 (.05)***	-.02 (.01)*
Cohort FE	Yes	Yes	Yes	Yes	Yes
Sector/ Industry FE	Yes	Yes	Yes	Yes	Yes
Obs.	552,559	163,575	520,602	195,986	195,986
R ²			.34	.07	.03

Table A2: Expanded results – specification A from table 3, excluding fully connected new firms (HHI=1)

	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
	(1)	(2)	(3)	(4)	(5)
HHI > 0	-.16 (.01)***	-.08 (.02)***	.01 (.005)***	-.05 (.007)***	-.007 (.003)***
<i>Human capital controls</i>					
Share w/ formal experience	-.18 (.02)***	-.09 (.03)***	.35 (.01)***	.24 (.02)***	-.07 (.005)***
Share from unemployment	.30 (.02)***	.12 (.03)***	-.11 (.009)***	-.16 (.02)***	.01 (.006)*
Share from same sector	-.20 (.03)***	-.16 (.03)***	.05 (.01)***	.03 (.01)***	-.01 (.004)***
Mean employee age (/10)	.009 (.009)	.01 (.01)	.06 (.006)***	-.02 (.006)***	-.05 (.004)***
Share close to retirement	-.07 (.05)	-.01 (.14)	-.06 (.03)**	-.14 (.04)***	.14 (.02)***
Mean years of schooling	-.01 (.003)***	-.001 (.003)	.02 (.002)***	.008 (.002)***	-.006 (.0006)***
Log mean previous wages	-.04 (.007)***	-.02 (.01)*	.31 (.01)***	.05 (.005)***	-.05 (.003)***
Share female workers	.02 (.02)	.08 (.03)**	-.06 (.008)***	-.02 (.02)	.03 (.006)***
<i>Prior employer size controls</i>					
Unobserved	-.003 (.02)	.07 (.04)	.07 (.004)***	.04 (.01)***	-.02 (.008)**
Medium (10 to 100)	.009 (.01)	-.002 (.02)	-.02 (.003)***	.06 (.006)***	.01 (.003)***
Large (100 to 1,000)	.08 (.01)***	.04 (.02)**	-.03 (.005)***	.09 (.009)***	.01 (.004)***
Very large (1,001+)	.06 (.02)***	.03 (.03)	-.05 (.004)***	.07 (.01)***	.01 (.005)**
<i>Initial size controls</i>					
3 to 4 initial empl.	-.11 (.009)***	-.02 (.02)	.05 (.003)***	-.13 (.006)***	-.003 (.003)
5 to 10 initial empl.	-.18 (.02)***	-.02 (.02)	.10 (.005)***	-.28 (.01)***	-.009 (.004)**
11 to 20 initial empl.	-.18 (.03)***	-.03 (.03)	.14 (.006)***	-.44 (.02)***	-.02 (.004)***
21 to 50 initial empl.	-.21 (.04)***	.05 (.05)	.17 (.009)***	-.59 (.03)***	-.02 (.005)***
51 to 100 initial empl.	-.27 (.05)***	.04 (.05)	.23 (.02)***	-.74 (.05)***	-.02 (.01)**
Cohort FE	Yes	Yes	Yes	Yes	Yes
Sector/ Industry FE	Yes	Yes	Yes	Yes	Yes
Obs.	524007	154421	493560	184483	184483
R ²			.33	.07	.03

Notes: Standard errors are clustered at the industry level. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table A3: Expanded results – specification A from table 3, excluding firms wherein the top network encompasses more than half of the traced employees

	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
	(1)	(2)	(3)	(4)	(5)
HHI > 0	-.14 (.01)***	-.07 (.02)***	.01 (.004)***	-.04 (.007)***	-.006 (.003)**
<i>Human capital controls</i>					
Share w/ formal experience	-.16 (.03)***	-.07 (.03)**	.34 (.01)***	.25 (.02)***	-.07 (.006)***
Share from unemployment	.28 (.02)***	.09 (.03)***	-.10 (.009)***	-.17 (.02)***	.009 (.006)
Share from same sector	-.19 (.02)***	-.14 (.03)***	.05 (.01)***	.04 (.01)***	-.01 (.004)***
Mean employee age (/10)	.01 (.009)	.02 (.01)	.06 (.006)***	-.01 (.006)**	-.05 (.004)***
Share close to retirement	-.06 (.05)	.04 (.13)	-.06 (.03)**	-.14 (.04)***	.14 (.02)***
Mean years of schooling	-.01 (.003)***	-.001 (.003)	.02 (.002)***	.008 (.002)***	-.006 (.0006)***
Log mean previous wages	-.04 (.007)***	-.02 (.01)*	.30 (.01)***	.05 (.005)***	-.05 (.003)***
Share female workers	.02 (.02)	.09 (.03)***	-.06 (.008)***	-.02 (.02)	.03 (.006)***
<i>Prior employer size controls</i>					
Unobserved	.004 (.02)	.08 (.04)*	.07 (.004)***	.04 (.01)***	-.02 (.008)**
Medium (10 to 100)	.01 (.009)	-.003 (.02)	-.02 (.003)***	.06 (.006)***	.01 (.003)***
Large (100 to 1,000)	.08 (.01)***	.03 (.02)*	-.03 (.004)***	.08 (.009)***	.01 (.004)***
Very large (1,001+)	.07 (.01)***	.06 (.03)**	-.05 (.004)***	.06 (.01)***	.01 (.005)**
<i>Initial size controls</i>					
3 to 4 initial empl.	-.11 (.009)***	-.02 (.02)	.05 (.003)***	-.13 (.006)***	-.003 (.003)
5 to 10 initial empl.	-.18 (.02)***	-.01 (.02)	.10 (.005)***	-.28 (.01)***	-.009 (.004)**
11 to 20 initial empl.	-.17 (.03)***	-.01 (.03)	.13 (.006)***	-.44 (.02)***	-.02 (.004)***
21 to 50 initial empl.	-.19 (.04)***	.07 (.05)	.17 (.009)***	-.59 (.03)***	-.02 (.005)***
51 to 100 initial empl.	-.26 (.05)***	.05 (.06)	.21 (.02)***	-.76 (.05)***	-.02 (.01)*
Cohort FE	Yes	Yes	Yes	Yes	Yes
Sector/ Industry FE	Yes	Yes	Yes	Yes	Yes
Obs.	497973	145603	468274	172969	172969
R ²			.31	.07	.03

Notes: Standard errors are clustered at the industry level. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table A4: Expanded results – specification A' from table 3

	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
	(1)	(2)	(3)	(4)	(5)
HHI > 0	-.09 (.01)***	-.006 (.02)	-.007 (.004)*	-.03 (.008)***	-.005 (.003)
HHI	-.17 (.04)***	-.25 (.05)***	.06 (.009)***	-.09 (.02)***	-.02 (.009)***
<i>Human capital controls</i>					
Share w/ formal experience	-.25 (.04)***	-.13 (.05)***	.42 (.01)***	.29 (.03)***	-.07 (.008)***
Share from unemployment	.39 (.03)***	.12 (.04)***	-.10 (.01)***	-.17 (.03)***	.005 (.009)
Share from same sector	-.27 (.04)***	-.22 (.05)***	.05 (.02)***	.03 (.02)*	-.01 (.006)*
Mean employee age (/10)	.03 (.01)**	.02 (.02)	.04 (.007)***	-.03 (.008)***	-.04 (.005)***
Share close to retirement	-.12 (.10)	-.33 (.17)*	-.09 (.03)***	-.08 (.06)	.16 (.03)***
Mean years of schooling	-.02 (.003)***	-.0009 (.003)	.02 (.003)***	.01 (.002)***	-.005 (.0008)***
Log mean previous wages	-.08 (.009)***	-.02 (.02)	.42 (.01)***	.06 (.007)***	-.07 (.004)***
Share female workers	-.07 (.03)**	.06 (.03)*	-.04 (.01)***	-.005 (.02)	.02 (.007)***
<i>Prior employer size controls</i>					
Unobserved	.23 (.06)***	.07 (.12)	.02 (.01)*	.11 (.04)**	.02 (.02)
Medium (10 to 100)	.01 (.02)	-.002 (.03)	-.02 (.003)***	.09 (.01)***	.004 (.005)
Large (100 to 1,000)	.11 (.02)***	.04 (.03)	-.03 (.005)***	.13 (.01)***	.006 (.006)
Very large (1,001+)	.11 (.03)***	.03 (.06)	-.02 (.01)**	.12 (.02)***	.0003 (.009)
<i>Initial size controls</i>					
5 to 10 initial empl.	-.08 (.01)***	-.03 (.02)	.07 (.006)***	-.13 (.01)***	-.001 (.004)
11 to 20 initial empl.	-.10 (.02)***	-.04 (.03)	.11 (.009)***	-.29 (.02)***	-.009 (.004)**
21 to 50 initial empl.	-.14 (.03)***	.01 (.05)	.15 (.01)***	-.44 (.03)***	-.02 (.005)***
51 to 100 initial empl.	-.21 (.05)***	-.006 (.05)	.20 (.02)***	-.61 (.05)***	-.01 (.01)
Cohort FE	Yes	Yes	Yes	Yes	Yes
Sector/ Industry FE	Yes	Yes	Yes	Yes	Yes
Obs.	303252	95113	289427	115182	115182
R^2			.39	.06	.03

Notes: Standard errors are clustered at the industry level. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table A5: Expanded results – specification *B* from table 3

	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
	(1)	(2)	(3)	(4)	(5)
Share unconnected	.16 (.08)**	-.12 (.09)	.09 (.02)***	.11 (.04)***	.02 (.01)
Share top network, rescaled	-.15 (.08)*	-.43 (.10)***	.13 (.02)***	-.004 (.06)	-.02 (.01)
<i>Human capital controls</i>					
Share w/ formal experience	-.28 (.04)***	-.11 (.06)**	.46 (.02)***	.29 (.03)***	-.06 (.009)***
Share from unemployment	.46 (.03)***	.12 (.05)**	-.12 (.01)***	-.18 (.03)***	-.005 (.01)
Share from same sector	-.29 (.04)***	-.24 (.06)***	.05 (.02)***	.05 (.02)**	-.005 (.007)
Mean employee age (/10)	.03 (.02)	.02 (.02)	.04 (.008)***	-.03 (.01)***	-.03 (.005)***
Share close to retirement	-.16 (.13)	-.31 (.24)	-.09 (.03)***	-.12 (.07)	.16 (.03)***
Mean years of schooling	-.02 (.004)***	.0004 (.004)	.02 (.003)***	.01 (.002)***	-.005 (.0007)***
Log mean previous wages	-.10 (.01)***	-.04 (.02)**	.46 (.02)***	.06 (.009)***	-.07 (.005)***
Share female workers	-.11 (.03)***	.04 (.04)	-.04 (.01)***	.003 (.03)	.02 (.007)**
<i>Prior employer size controls</i>					
Unobserved	.26 (.09)***	.05 (.23)	.02 (.02)	.10 (.06)	.01 (.03)
Medium (10 to 100)	-.004 (.02)	-.03 (.04)	-.02 (.004)***	.09 (.01)***	-.001 (.007)
Large (100 to 1,000)	.11 (.03)***	.02 (.04)	-.03 (.006)***	.13 (.02)***	-.003 (.008)
Very large (1,001+)	.11 (.04)***	.04 (.09)	-.02 (.01)*	.13 (.02)***	-.01 (.009)
<i>Initial size controls</i>					
5 to 10 initial empl.	-.06 (.02)***	-.01 (.03)	.06 (.008)***	-.09 (.01)***	.004 (.006)
11 to 20 initial empl.	-.07 (.02)***	-.02 (.03)	.10 (.01)***	-.25 (.02)***	-.002 (.006)
21 to 50 initial empl.	-.13 (.03)***	.04 (.05)	.14 (.01)***	-.40 (.03)***	-.007 (.007)
51 to 100 initial empl.	-.20 (.04)***	.02 (.05)	.19 (.02)***	-.56 (.04)***	-.002 (.01)
Cohort FE	Yes	Yes	Yes	Yes	Yes
Sector/ Industry FE	Yes	Yes	Yes	Yes	Yes
Obs.	221,893	70,883	212,579	86,043	86,043
R^2			.42	.06	.03

Notes: Standard errors are clustered at the industry level. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table A6: Expanded results – specification B' from table 3

	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
	(1)	(2)	(3)	(4)	(5)
HHI > 0	-.03 (.02)	.04 (.03)	-.003 (.005)	.02 (.01)	.008 (.006)
Share unconnected	.11 (.09)	-.03 (.11)	.08 (.02)***	.15 (.05)***	.03 (.02)*
Share top network, rescaled	-.18 (.09)**	-.40 (.11)***	.13 (.02)***	.01 (.06)	-.01 (.01)
<i>Human capital controls</i>					
Share w/ formal experience	-.27 (.04)***	-.12 (.06)**	.46 (.02)***	.29 (.03)***	-.07 (.009)***
Share from unemployment	.46 (.03)***	.12 (.05)**	-.12 (.01)***	-.18 (.03)***	-.005 (.01)
Share from same sector	-.29 (.04)***	-.24 (.06)***	.05 (.02)***	.05 (.02)**	-.005 (.007)
Mean employee age (/10)	.03 (.02)	.02 (.02)	.04 (.008)***	-.03 (.009)***	-.03 (.005)***
Share close to retirement	-.16 (.13)	-.30 (.24)	-.09 (.03)***	-.12 (.07)	.16 (.03)***
Mean years of schooling	-.02 (.004)***	.0004 (.004)	.02 (.003)***	.01 (.002)***	-.005 (.0007)***
Log mean previous wages	-.09 (.01)***	-.04 (.02)**	.46 (.02)***	.06 (.009)***	-.07 (.005)***
Share female workers	-.11 (.03)***	.04 (.04)	-.04 (.01)***	.004 (.03)	.02 (.007)**
<i>Prior employer size controls</i>					
Unobserved	.25 (.09)***	.06 (.23)	.02 (.02)	.10 (.06)	.02 (.03)
Medium (10 to 100)	-.004 (.02)	-.03 (.04)	-.02 (.004)***	.09 (.01)***	-.002 (.008)
Large (100 to 1,000)	.11 (.03)***	.02 (.04)	-.03 (.006)***	.13 (.02)***	-.003 (.008)
Very large (1,001+)	.11 (.04)***	.04 (.09)	-.02 (.01)*	.14 (.02)***	-.009 (.01)
<i>Initial size controls</i>					
5 to 10 initial empl.	-.05 (.02)***	-.01 (.03)	.06 (.008)***	-.09 (.01)***	.003 (.006)
11 to 20 initial empl.	-.07 (.02)***	-.03 (.03)	.10 (.01)***	-.25 (.02)***	-.004 (.006)
21 to 50 initial empl.	-.12 (.03)***	.02 (.05)	.14 (.01)***	-.40 (.03)***	-.009 (.007)
51 to 100 initial empl.	-.19 (.04)***	.002 (.05)	.19 (.02)***	-.56 (.04)***	-.004 (.01)
Cohort FE	Yes	Yes	Yes	Yes	Yes
Sector/ Industry FE	Yes	Yes	Yes	Yes	Yes
Obs.	221,893	70,883	212,579	86,043	86,043
R^2			.42	.06	.03

Notes: Standard errors are clustered at the industry level. Significance indicated is at 10%(*), 5%(**), and 1%(***).

A.2 Correlates of network concentration

While the main purpose of this paper is to study the role of networks in entrepreneurial outcomes, understanding the factors that influence network concentration is helpful in contextualizing our main findings.

In this section we examine the factors that predict the level of connectedness in a startup. First, in table A7 we report a correlation matrix of the network measures and the main controls employed. Some interesting patterns emerge: average (firm-level) schooling is negatively correlated with network strength. This is perhaps explained by the strong negative correlation between education and age: since this runs counter to the mechanical positive link (i.e. people employed at younger ages are on average less educated), it bears witness to a large increase in public schooling in Brazil in the couple of decades that predate the sample. As expected, prior wages increase with age and education. The share of female workers is negatively associated with age and (very) positively associated with education, indicating recent entry of women into the labor force, in particular in white collar occupations.

Finally, the number of initial workers increases with prior wage, but decreases with education and the share of women – which suggests again that men have stronger employment networks, but are on average less educated. Again, this is consistent with a recent expansion of public schooling, as well as a pronounced increase in female labor force participation. These dramatic recent trends, as well as the direct link between networks and the length of formal sector participation, make it difficult to determine the direct link between demographic variables (age, education, gender) and the propensity to collaborate with former co-workers. To distinguish between the different channels, we make use of multivariate regression.

Specifically, we regress HHI and the other network measures on the secondary controls from table 4 and report results in table A8: in column 1 we find that by far the strongest predictors of new firm connectedness are the share of workers with a prior formal job (positive effect) and the share of workers who were unemployed during the previous year (negative effect). An older

Table A7: Correlation matrix over the estimation sample

	HHI	HHI>0	share uncon.	top network	formal exper	from unemp	same sector	age	near retire	yrs sch	prev wage	share fem	num ini
HHI	1.00												
HHI>0	0.64	1.00											
Share unconnected	-0.90	-0.88	1.00										
Share top network	0.97	0.78	-0.96	1.00									
Share w/ formal exp	0.23	0.27	-0.27	0.26	1.00								
Share from unemp	-0.18	-0.12	0.17	-0.18	0.30	1.00							
Share same sector	0.27	0.23	-0.28	0.29	0.11	-0.12	1.00						
Mean age (/10)	0.13	0.16	-0.16	0.15	0.35	0.22	0.08	1.00					
Near retirement	0.06	0.04	-0.05	0.06	0.05	0.02	0.04	0.37	1.00				
Mean yrs school (/10)	-0.03	-0.09	0.06	-0.04	-0.02	-0.07	-0.05	-0.21	-0.07	1.00			
Log mean prev wage	-0.01	0.08	-0.03	0.01	0.18	0.01	-0.06	0.18	0.02	0.14	1.00		
Share female workers	-0.04	-0.11	0.08	-0.06	-0.15	-0.09	0.00	-0.14	0.00	0.32	-0.11	1.00	
Num initial empl	0.04	0.43	-0.24	0.13	-0.01	-0.04	0.04	0.05	-0.01	-0.09	0.10	-0.07	1.00

workforce is also associated with more connected new firms, but most of that effect comes from being within 5 years of retirement.

Interestingly, gender composition has no measurable effect, and the coefficient on schooling is quite small in magnitude: if workers have on average 5 more years of education, this predicts a 1 percent lower HHI on average.

In column 2 we introduce more human-capital controls: the share of workers with prior experience in the same sector as the startup, average prior wages, as well as prior firm size, averaged over initial workers. There is little measurable effect on the coefficients of age, proximity to retirement, and gender composition, although education once again becomes insignificant.

In the remaining columns we consider the alternate network strength measured: in column 3 we regress the $\text{HHI} > 0$ indicator and find consistent results – some of the coefficients are now larger or gain significance, but this is difficult to interpret, considering the strong correlation of the dependent variable with firm size, and the fact that both education and gender composition are also linked to the number of initial firm employees. In columns 4 and 5 we regress the share of workers with no (observed) prior links, and the size of the top network as a share of the initial firm size. Coefficients are consistent in sign with the pattern implied by column 2.

Interestingly, despite the potential for more connections at medium and large firms, it appears that coming from small firm prior employment is more predictive of start-up concentration.

Next, we take the specification from column 2 in table A8 to analyze the determinants of HHI for each of the eight largest sectors.¹ Excluding real estate, there is little to no link between average education level and connectedness, once sector and other demographics are controlled for. There is a small but robust negative relationship with wage. The share of employees with formal sector experience predicts a dramatically higher HHI level, but this is hardly surprising, given that for employees *without* such experience we fail to observe any potential links.

The next two most predictive variables both in terms of statistical significance and size are the

¹See appendix section A.7 for summary statistics and main analysis results by sector.

Table A8: Explaining firm-level network concentration measures

	HHI		HHI>0	Share unconnected	Share top network
	(1)	(2)	(3)	(4)	(5)
<i>Human capital controls</i>					
Share w/ formal experience	.32 (.009)***	.33 (.009)***	.66 (.01)***	-.47 (.01)***	.40 (.01)***
Mean employee age (/10)	.03 (.003)***	.03 (.002)***	.03 (.003)***	-.03 (.002)***	.03 (.002)***
Share close to retirement	.09 (.03)***	.06 (.02)**	.04 (.02)*	-.05 (.02)**	.06 (.02)**
Mean years schooling (/10)	-.02 (.007)**	-.008 (.005)	-.02 (.006)***	.01 (.005)***	-.009 (.005)
Share female workers	.0002 (.003)	-.004 (.005)	-.02 (.006)***	.008 (.006)	-.005 (.005)
Share from unemployment	-.28 (.02)***	-.26 (.01)***	-.34 (.01)***	.31 (.01)***	-.29 (.01)***
Share from same sector		.15 (.01)***	.20 (.01)***	-.18 (.01)***	.17 (.01)***
Log mean previous wages		-.03 (.003)***	-.03 (.004)***	.03 (.004)***	-.03 (.004)***
<i>Prior employer size controls</i>					
Unobserved		.12 (.006)***	.15 (.007)***	-.13 (.007)***	.13 (.007)***
Medium (10 to 100)		-.05 (.003)***	-.05 (.004)***	.05 (.003)***	-.05 (.003)***
Large (100 to 1,000)		-.05 (.007)***	-.06 (.01)***	.05 (.008)***	-.05 (.008)***
Very large (1,001+)		.05 (.03)*	.05 (.02)**	-.06 (.02)**	.05 (.03)**
<i>Initial size controls</i>					
3 to 4 initial empl.	.04 (.002)***	.05 (.002)***	.15 (.005)***	-.10 (.003)***	.08 (.003)***
5 to 10 initial empl.	.05 (.004)***	.08 (.003)***	.36 (.006)***	-.19 (.003)***	.13 (.003)***
11 to 20 initial empl.	.05 (.005)***	.08 (.003)***	.62 (.008)***	-.26 (.005)***	.17 (.004)***
21 to 50 initial empl.	.04 (.006)***	.08 (.005)***	.82 (.005)***	-.32 (.008)***	.19 (.006)***
51 to 100 initial empl.	.05 (.009)***	.08 (.007)***	.91 (.006)***	-.39 (.01)***	.20 (.01)***
Cohorts	Yes	Yes	Yes	Yes	Yes
Sectors (detailed)	Yes	Yes	Yes	Yes	Yes
Obs.	552559	552559	552559	552559	552559
R ²	.15	.22	.42	.32	.27

Notes: Regressions of network measures on human capital and other controls. Standard errors are clustered at the industry level. Significance indicated is at 10%(*), 5%(**), and 1%(***).

^a We say an employee is coming in “from unemployment” if he or she did not have a job during the year before the new firm enters (but *did* have a job at some time prior).

share of employees who were unemployed and the share of employees who worked in the same sector as the start-up in their immediately prior job.

The most dramatic variation of HHI with age is registered for real estate. In this sector, new firms are also more likely to be connected when employees have low levels of education, lower previous wages, and are majority male.

In the hospitality industry, connectedness only depends on age insofar as founders are within 5 years of retirement age. Education does *not* determine strong networks in most industries, with the exception of hotels and restaurants and real estate (negative effect) and professional services (positive). Across all sectors, startups are more tightly connected if founding members were in relatively low paying jobs (conditional on age and education). This suggests that there is possibly negative selection into starting a venture with former colleagues.

To summarize, firms comprising workers with same-industry formal sector experience and contiguous employment tend to be more concentrated, as do firms with workers very close to retirement. Once we control for these aspects, however, other demographic characteristics play only a small role in networks, or none at all, and this pattern holds across industries.

We interpret this as saying that selection into working with former co-workers is not strongly tied to personal characteristics other than labor market experience and opportunity.

Table A9: Explaining firm-level HHI, by sector

dep var = HHI	ALL (1)	Manuf (2)	Car services (3)	Whole- sale (4)	Retail (5)	Hotels, restaur (6)	Prof. services (7)	Cons- truction (8)	Real Estate (9)
<i>Human capital controls</i>									
Mean employee age (/10)	.03 (.002)***	.02 (.002)***	.03 (.003)***	.03 (.003)***	.02 (.004)***	.005 (.004)	.04 (.004)***	.01 (.002)***	.04 (.002)***
Share close to retirement	.06 (.02)**	.04 (.02)**	.006 (.04)	-.009 (.03)	.02 (.02)	.04 (.02)***	.05 (.03)	.04 (.02)*	.07 (.01)***
Mean years of school (/10)	-.008 (.005)	.002 (.005)	-.004 (.01)	-.01 (.007)	-.009 (.008)	-.02 (.008)*	.008 (.005)	.0001 (.003)	-.02 (.008)**
Share female workers	-.004 (.005)	-.006 (.008)	-.02 (.01)**	-.02 (.005)***	-.009 (.007)	.03 (.006)***	-.002 (.007)	.004 (.008)	-.07 (.02)***
Share w/ formal exper.	.33 (.009)***	.34 (.007)***	.34 (.01)***	.29 (.01)***	.31 (.009)***	.27 (.008)***	.34 (.02)***	.28 (.01)***	.41 (.007)***
Share from unemployment	-.26 (.01)***	-.26 (.005)***	-.24 (.01)***	-.24 (.01)***	-.22 (.005)***	-.23 (.02)***	-.29 (.01)***	-.24 (.02)***	-.40 (.01)***
Share from same sector	.15 (.01)***	.14 (.01)***	.11 (.01)***	.21 (.01)***	.12 (.01)***	.11 (.02)***	.20 (.04)***	.14 (.02)***	.20 (.01)***
Log mean previous wages	-.03 (.003)***	-.02 (.002)***	-.02 (.003)***	-.02 (.003)***	-.02 (.002)***	-.03 (.003)***	-.02 (.006)***	-.02 (.006)***	-.07 (.008)***
<i>Prior employer size controls</i>									
Unobserved	.12 (.006)***	.15 (.009)***	.11 (.008)***	.13 (.01)***	.09 (.006)***	.07 (.01)***	.13 (.008)***	.11 (.03)***	.21 (.02)***
Medium (10 to 100)	-.05 (.003)***	-.05 (.003)***	-.05 (.001)***	-.05 (.005)***	-.05 (.004)***	-.05 (.008)***	-.06 (.005)***	-.06 (.006)***	-.05 (.004)***
Large (100 to 1,000)	-.05 (.007)***	-.06 (.006)***	-.05 (.005)***	-.07 (.006)***	-.05 (.006)***	-.05 (.007)***	-.08 (.006)***	-.08 (.009)***	.03 (.02)*
Very large (1,001+)	.05 (.03)*	.02 (.006)***	.007 (.004)*	-.03 (.006)***	.005 (.005)	.01 (.003)***	.01 (.007)	.01 (.02)	.27 (.03)***
Initial size controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector (detailed) FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	552559	81017	22154	30012	175132	48047	42573	31982	23155
R ²	.22	.21	.19	.2	.18	.17	.25	.19	.46

Notes: Regressions of the HHI measure of network concentration on human capital and other controls, by broad sector. Standard errors are clustered at the industry level. Significance indicated is at 10%(*), 5%(**), and 1%(***).

A.3 Observable and unobservable variables

Table A10: Main explanatory variable, with and without human capital controls

HHI coefficients	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
controls included:	(1)	(2)	(3)	(4)	(5)
[only cohort+industry FE]	-.42 (.02)***	-.30 (.03)***	.14 (.01)***	-.06 (.01)***	-.04 (.004)***
share tracked	-.41 (.02)***	-.29 (.03)***	.04 (.009)***	-.12 (.02)***	-.02 (.004)***
share tracked + ini size	-.35 (.02)***	-.28 (.03)***	.02 (.01)*	-.08 (.007)***	-.01 (.004)***
ini size	-.38 (.02)***	-.29 (.03)***	.13 (.01)***	-.03 (.007)***	-.04 (.004)***
ini size + prior employer size	-.36 (.02)***	-.29 (.03)***	.14 (.01)***	-.02 (.008)*	-.04 (.005)***
prior employer size	-.42 (.02)***	-.30 (.03)***	.16 (.01)***	-.07 (.01)***	-.05 (.005)***
tracked + ini size + prior empl (table 5)	-.34 (.02)***	-.27 (.03)***	.02 (.009)***	-.06 (.007)***	-.01 (.004)***
tracked + ini size + prior empl + human capital controls (table 4 = baseline)	-.20 (.02)***	-.20 (.03)***	.02 (.008)**	-.10 (.01)***	-.01 (.005)**

Notes: All specifications include cohort and industry fixed effects. The sample is held fixed. As usual, standard errors are clustered at the industry level and significance indicated is at the 10%(*), 5%(**), and 1%(***) levels.

To investigate any potential bias due to unobserved characteristics, table A10 reports the main coefficients when different sets of controls are included. Survival results are remarkably stable, further solidifying our main result. There is more variation in the coefficients on wages and employment growth, but the patterns are entirely consistent with intuition: for instance, failing to control for share of employees who are tracked to a prior formal sector job (i.e. with formal sector experience) yields a much higher positive “effect” of close networks on initial wages, which makes sense given that we can only identify employee links for tracked employees, and we can expect there to be a significant premium in wages to previous formal sector experience.

In table A11 we reprise the analysis reported in the main regression table, but omit the concentration measure (HHI), in order to discern how much explanatory power it takes from the other controls.

Comparing columns 1 and 2 with the parallel coefficients in table 4, we find that the introduction of networks attenuated the apparent effect of the first two control variables (share of employees with formal sector experience and share previously unemployed), but had a much more subtle (or zero) influence on the remaining variables. Coefficients in the 3rd regression (average initial wage) are essentially unaffected by the introduction of the network control.

In column 4, the coefficients on formal labor and sector-specific experience, as well as unemployment, are attenuated in the absence of the network measure, which is consistent with the fact that prior links have a *negative* effect on employment growth. For instance, the second coefficient in column 4 of table A11 suggests that a startup whose employees were all unemployed immediately before will grow 12% slower than an otherwise similar startup with no workers from unemployment. However, once we account for the fact that tightly linked firms tend to grow slower (table 4), the effect of coming from unemployment is more pronounced: -15%. There are no noticeable differences in coefficients in the last column (growth in wages).

Overall, therefore, the only significant changes were on the coefficients on the control variables most directly linked to our measure of network strengths – those measuring the extent of employees' prior experience in the labor market.

Table A11: Survival and performance analysis

	Cox hazard		initial ln(wage)	growth t to t+3	
	t to t+3	t+3 to t+6		empl	wage
	(1)	(2)	(3)	(4)	(5)
HHI	-	-	-	-	-
<i>Human capital controls</i>					
Share w/ formal experience	-.27 (.03)***	-.14 (.03)***	.35 (.01)***	.19 (.02)***	-.08 (.005)***
Share from unemployment	.35 (.02)***	.16 (.03)***	-.11 (.01)***	-.12 (.02)***	.02 (.006)***
Share from same sector	-.22 (.03)***	-.19 (.04)***	.05 (.01)***	.01 (.01)	-.01 (.004)***
Mean employee age (/10)	.004 (.009)	.01 (.01)	.06 (.006)***	-.02 (.006)***	-.05 (.004)***
Share close to retirement	-.09 (.06)	-.01 (.13)	-.06 (.03)**	-.12 (.04)***	.13 (.02)***
Mean years of schooling	-.01 (.003)***	-.001 (.003)	.02 (.003)***	.008 (.002)***	-.006 (.0006)***
Log mean previous wages	-.04 (.007)***	-.02 (.01)	.32 (.01)***	.05 (.005)***	-.05 (.003)***
Share female workers	.02 (.02)	.08 (.03)**	-.06 (.008)***	-.02 (.02)	.03 (.006)***
<i>Prior employer size controls</i>					
Unobserved	-.004 (.02)	.06 (.04)	.07 (.005)***	.04 (.01)***	-.01 (.008)*
Medium (10 to 100)	.01 (.01)	.002 (.02)	-.02 (.003)***	.08 (.007)***	.01 (.003)***
Large (100 to 1,000)	.08 (.01)***	.04 (.02)**	-.03 (.004)***	.10 (.01)***	.01 (.004)***
Very large (1,001+)	.05 (.02)***	.01 (.04)	-.04 (.007)***	.06 (.01)***	.009 (.004)**
<i>Initial size controls</i>					
3 to 4 initial empl.	-.13 (.008)***	-.03 (.02)**	.05 (.003)***	-.14 (.007)***	-.004 (.003)
5 to 10 initial empl.	-.24 (.01)***	-.06 (.02)**	.10 (.005)***	-.30 (.01)***	-.01 (.004)**
11 to 20 initial empl.	-.28 (.03)***	-.08 (.03)***	.14 (.007)***	-.47 (.02)***	-.02 (.004)***
21 to 50 initial empl.	-.34 (.04)***	-.02 (.04)	.18 (.01)***	-.63 (.03)***	-.03 (.005)***
51 to 100 initial empl.	-.42 (.05)***	-.04 (.05)	.24 (.02)***	-.79 (.05)***	-.02 (.01)**
Cohort FE	Yes	Yes	Yes	Yes	Yes
Sector/ Industry FE	Yes	Yes	Yes	Yes	Yes
Obs.	552559	163575	520602	195986	195986
R^2			.34	.07	.03

Notes: All controls from table 4 are included, except for the network measure. Standard errors are clustered at the industry level. Significance indicated is at 10%(*), 5%(**), and 1%(***).

A.4 The role of initial firm size

There are two meaningful ways in which the initial scale of the firm may be important for our analysis. First, we expect that - holding network strength constant - new firms of different sizes may have different survival and growth patterns on average. Second, we want to test the possibility that the strength of networks varies across firm size.

Table 4 in the main text addresses the first question by reporting the coefficients on initial size bins. Unsurprisingly, firms which start larger have better odds of survival, although the effect is smaller and less prevalent past the first 3 years. Bigger startups also offer significantly higher initial wages, and while their wage growth is slower, the initial advantage persists through the first few years. However, larger firms have slower employment growth.

These results are consistent with small firms being more capital constrained, which limits their initial productivity and survival rates, but leaves room for growth conditional on survival, as they are able to make up for the initial disadvantage. Small firms are also more likely to rely on alternate compensation schemes—such as equity shares— which would contribute to a lower observed initial wage, and may have fewer managerial employees who earn higher pay.

Next, we address the second question, namely whether employee networks are more or less predictive of success for different firm sizes. We run the baseline survival and performance regressions separately for the six bins of initial firm size we considered.

Table A12 suggests that the effect of prior connections on the hazard of closure within the first 3 years is large and significantly negative for all firm sizes up to around 50 initial employees. Furthermore, the effect increases with size (until around 20 employees), which is consistent with the increased difficulty of coordinating larger teams.

Survival for the subsequent 3 years (column 2) is predicted by network concentration in a very similar fashion: tightly linked firms are less likely to exit, and the prior connections have a relatively larger impact for larger firms (up to around 50 employees).

There is no measurable correlation between network links and the survival of firms with over

Table A12: Survival and performance by initial size

HHI coefficients	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
	(1)	(2)	(3)	(4)	(5)
2 initial empl	-.08 (.02)***	-.12 (.05)**	.02 (.01)	-.08 (.01)***	-.004 (.007)
3 to 4	-.21 (.02)***	-.10 (.05)**	.03 (.01)***	-.09 (.01)***	-.008 (.007)
5 to 10	-.26 (.04)***	-.36 (.06)***	.005 (.009)	-.11 (.03)***	-.02 (.01)*
11 to 20	-.41 (.06)***	-.50 (.13)***	.04 (.01)***	-.20 (.04)***	-.04 (.02)**
21 to 50	-.26 (.09)***	-.52 (.15)***	.12 (.02)***	-.16 (.08)**	-.03 (.03)
51 to 100	-.04 (.19)	-.28 (.37)	.20 (.04)***	-.05 (.17)	-.02 (.05)

Notes: This is a summary of baseline survival regressions run separately for each initial size category. Only the coefficient on the HHI measure is reported. Standard errors are clustered at the industry level.

50 initial employees. This is unsurprising, since the large upfront investment needed to start a firm on that scale means that these entrants are likely different in fundamental ways from *de novo* start-ups of interest.

For most starting sizes, new firms with tight employment links tend to pay higher initial wages (column 3) but increase pay at a slower rate (col 5). Again, very large startups (over 50) are outliers, and the effect is smaller or zero for startups with just 2 employees – perhaps because these are more likely to be linked in ways we cannot observe (such as school friends or relative). Employment growth is consistently slower for connected firms, although again we lack precise estimates for the top 2 categories.

On the whole, table A12 is reassuring in its consistency with baseline results.

Table A13: Survival and performance by initial size

	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
	(1)	(2)	(3)	(4)	(5)
3 to 4					
Share unconnected	.22 (.16)	.25 (.28)	-.02 (.04)	-.01 (.13)	.003 (.05)
Share top parent	.04 (.20)	.42 (.34)	.05 (.04)	-.16 (.14)	-.05 (.06)
5 to 10					
Share unconnected	.24 (.08)***	-.20 (.11)*	.01 (.02)	.09 (.04)**	.02 (.02)
Share top parent	-.04 (.09)	-.58 (.13)***	.04 (.02)**	-.02 (.06)	-.01 (.03)
11 to 20					
Share unconnected	.16 (.10)	-.08 (.15)	.14 (.02)***	.12 (.09)	.04 (.03)
Share top parent	-.25 (.11)**	-.52 (.18)***	.16 (.02)***	-.04 (.11)	.004 (.03)
21 to 50					
Share unconnected	.23 (.13)*	-.19 (.17)	.14 (.02)***	.05 (.10)	.01 (.04)
Share top parent	-.14 (.14)	-.59 (.16)***	.20 (.03)***	-.09 (.11)	-.02 (.04)
51 to 100					
Share unconnected	.04 (.22)	.60 (.32)*	.10 (.04)**	.36 (.20)*	.03 (.06)
Share top parent	-.11 (.22)	.26 (.37)	.21 (.04)***	.22 (.20)	.003 (.06)

Notes: This is a summary of baseline survival regressions run separately for each initial size category. Only the coefficient on the HHI measure is reported. Standard errors are clustered at the industry level.

A.5 The role of prior employer size

Another dimension naturally relevant for employee networks is the size of the prior employer. We separate the sample by the log average size of prior employers, taken over all employees in the start-up, and run the main regressions separately for each category, again using HHI to measure networks. Table A14 reports only the main coefficient from each regression.

Table A14: Survival and performance analysis, separate estimations by prior employer size

HHI coefficients	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
	(1)	(2)	(3)	(4)	(5)
Unobserved	-.03 (.07)	-.17 (.15)	-.003 (.01)	-.04 (.05)	.04 (.02)
Small (< 10)	-.10 (.03)***	-.14 (.05)***	.003 (.007)	-.11 (.01)***	-.002 (.007)
Medium-sized (10 to 100)	-.25 (.03)***	-.18 (.05)***	.04 (.007)***	-.08 (.01)***	-.02 (.007)***
Large (100 to 1,000)	-.28 (.03)***	-.23 (.06)***	.04 (.009)***	-.07 (.02)***	-.03 (.009)***
Very large (over 1,000)	-.20 (.05)***	-.29 (.10)***	.06 (.03)**	-.10 (.03)***	-.02 (.01)

Notes: This is a summary of baseline survival regressions run separately for categories of (log) average prior employer size, measured at the end of the previous year. Only the coefficient on the HHI measure is reported. Standard errors are clustered at the industry level.

For medium, large, and very large prior employers—which collectively account for 70% of the sample—results are remarkably robust across for all target outcomes.

When it comes to small prior employer size, the effect of networks on survival maintains its direction, albeit with a lower magnitude. The effect on employment growth is similar or greater in magnitude. Interestingly, however, neither initial wages nor wage growth differ by network strength for small prior employers - possibly because small firms are better able to gauge true productivity (and we are controlling for prior wage).

The first row of the table refers to the “unobserved prior employer size” category: these are the

cases in which former employers are absent in the data at the end of the year before the start-up's birth, indicating likely closure.² This category also includes new firms wherein very few employees can be traced back to *any* previous job, or can only be traced back to a much earlier job, making it difficult to accurately assess their prior work connections, perhaps explaining why these estimates are too noisy to make meaningful comparisons.

One take-away from table A14 is the fact that strong networks predict higher survival and slower employment growth across *all* known average prior employer sizes, which eliminates the possibility that our baseline results are driven by divestitures or are an artifact of the way we measure networks. In the section that follows we discuss more potential interpretations.

²While it is tempting to think of these cases as exogenous setups for examining networks, several factors jeopardize this interpretation, including lower productivity and/or unfavorable market conditions within the workers' sector.

A.6 Occupation compositional effects

There is some concern – in particular when it comes to our analysis of average wages – that much of the variation could in fact be driven by the compositional mix of occupations in the firm. If for instance workers in white collar occupations are more likely to start firms together, we might mistakenly attribute the higher initial average wage to network links, rather than to worker composition. Similarly, if tightly linked firms tend to bring blue collar workers on board in greater numbers, we may observe slower wage growth and mis-attribute that as well.

We address these concerns by including detailed sector fixed effects, as well as including average worker education and previous wages. Still, it is possible that even holding these parameters fixed, the occupational mix within a firm is not completely determined, and there may be systematic differences by network connectedness.

We can address this concern when it comes to initial wage, by employing individual-level data for the firm's founding employees. Thus, as a robustness check, we regressed logged hourly wage on occupation fixed effects (350+ categories), then calculated wage residuals for each initial employee. We then ran the firm-level initial wage regression (column 3 of table 4), but substituting average wage *residuals* on the left-hand side. With this approach, the coefficient on HHI was estimated at 0.016 (+/- 0.007),³ essentially identical to the baseline coefficient of 0.019.

Ideally, we would like to run a similar robustness exercise for wage *growth*, but – for practical reasons related to data access – we are unable to do so.⁴ It seems reasonable however to argue that if occupational mix is not sufficiently different in tightly linked firms at formation to account for the observed correlation between networks and wages, this is unlikely to change within the firm's first 3 years.

³Full regression results are not currently reported, but are available upon request.

⁴We only have individual-level data for workers employed during the firm's first year, so we cannot explicitly include occupational controls in our wage growth analysis.

A.7 Sectoral differences

In table A15 we report the count and characteristics of new firms by sector, as grouped into broad categories by the Brazilian labor ministry into the *CNAE* classification system (*Classificação Nacional de Atividade Econômica*). The largest sectors in our sample are retail sale (31.7%), manufacturing (14.7%), hospitality services (hotels and restaurants - 8.7%) and professional services (legal, accounting, marketing, management, etc - 7.7%). There is variation in the degree of interconnectedness of employees: new hotels and restaurants have the lowest average HHI at 0.072, while medical practices have a concentration index almost twice as high (0.124), and the real estate services sector is an outlier with average HHI=0.236. Nearly 50% of new real estate firms have employees with prior work connections, compared to 30% for the economy overall.

Table A15 also lists average survival rates by sector, measured at 3 and 6 years after entry. Once again, real estate and medical services firms lead the pack, with 85% and 69% odds of surviving the first six years, respectively.⁵ Most other sectors hover around 42 to 46 percent likelihood of lasting six years. The sectors with lowest survival odds are wholesale and construction services, at 37-38%. Appendix figure 1 shows the distribution of initial firm size by sector. The median starting size ranges between 3 and 5 across most sectors, with retail establishments starting out the smallest and construction firms starting out the largest (median of 7 initial employees).

As figure 1 shows, firms across all sectors are relatively small at startup, with the median number of employees ranging between 3 and 5 in our sample (once we've excluded 1-person firms, divestitures, public corporations, etc.), with only construction sector firms starting on a larger scale (median initial size of 7). The strength of employee networks varies across sectors, as table A15 reported, with new hotels, restaurants, and retail sales establishments being the least connected, and real estate firms having the most concentrated networks, followed by medical and transportation services. This is consistent with the expectation that, the more easily observable work output is,

⁵Law firms are included within the *Professional services* category, as they are too few to analyze separately. However, their survival rate at 6 years after launch is 69%, the same as medical firms.

Table A15: Network concentration and survival by sector

Sector	Num firms	Median ini size	HHI			Share not connected	Share top network	% survived	
			Mean	SD	> 0			t+3	t+6
Manufacturing ^a	81,017	5	0.117	0.249	37.7%	0.783	0.165	67.8	45.8
Car sales, repair	22,154	3	0.103	0.252	26.4%	0.835	0.135	69.0	46.3
Wholesale	30,012	4	0.103	0.244	29.9%	0.825	0.140	61.9	37.2
Retail sale	175,132	3	0.088	0.240	21.6%	0.864	0.113	65.6	42.0
Hotels, restaurants	48,047	5	0.072	0.205	25.5%	0.868	0.102	61.1	42.2
Medical activities	11,025	3	0.124	0.283	28.0%	0.815	0.155	81.4	68.7
Prof. services ^b	42,573	4	0.111	0.251	35.6%	0.802	0.152	65.3	45.9
Construction	31,982	7	0.089	0.204	43.9%	0.791	0.142	55.6	38.2
Real estate serv.	23,155	5	0.236	0.352	49.5%	0.664	0.292	91.6	85.4
Transportation	14,637	3	0.107	0.248	30.8%	0.816	0.145	68.1	43.3
Computer services	6,306	4	0.096	0.244	23.8%	0.852	0.124	64.9	41.6
Other ^c	66,519	4	0.103	0.243	31.3%	0.818	0.141	71.4	52.4
Total	552,559	4	0.103	0.247	30.1%	0.824	0.139	67.1	46.9

Notes: Median initial employment is computed after removing firms that start with 1 or over 100 workers.

^a Within manufacturing, we list separately the industries with at least 5,000 new firms in our regression sample. See appendix A.8 for a more complete list.

^b “Professional services” include legal activities, accounting, marketing, management, architecture, photography, etc.

^c “Other” category includes agriculture, mining, travel agencies, gas stations, education, and other services.

the less need there is to rely on personal networks.

In figures 2 and 3 we illustrate the density and survival rates of firms by sector, size, and network concentration. Some interesting patterns emerge: for instance, while initial size is predictive of higher survival odds at t+3, it is only relevant longer term for *some* sectors (e.g. hospitality, real estate) and much less so for others (e.g. manufacturing, wholesale, professional services). Network concentration appears to be correlated with increased survival across most – if not all – sectors and firm sizes.

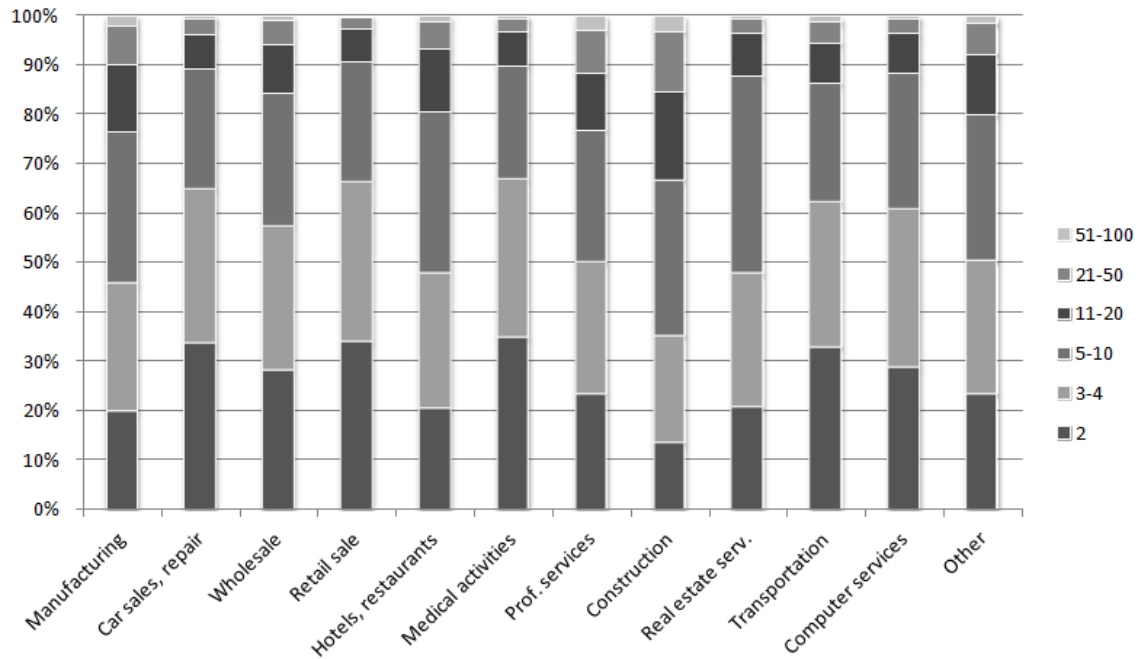


Figure 1: Distribution of new firm size by sector in the baseline regression sample

Survival at t+3 and t+6 - by sector, initial size, and HHI range

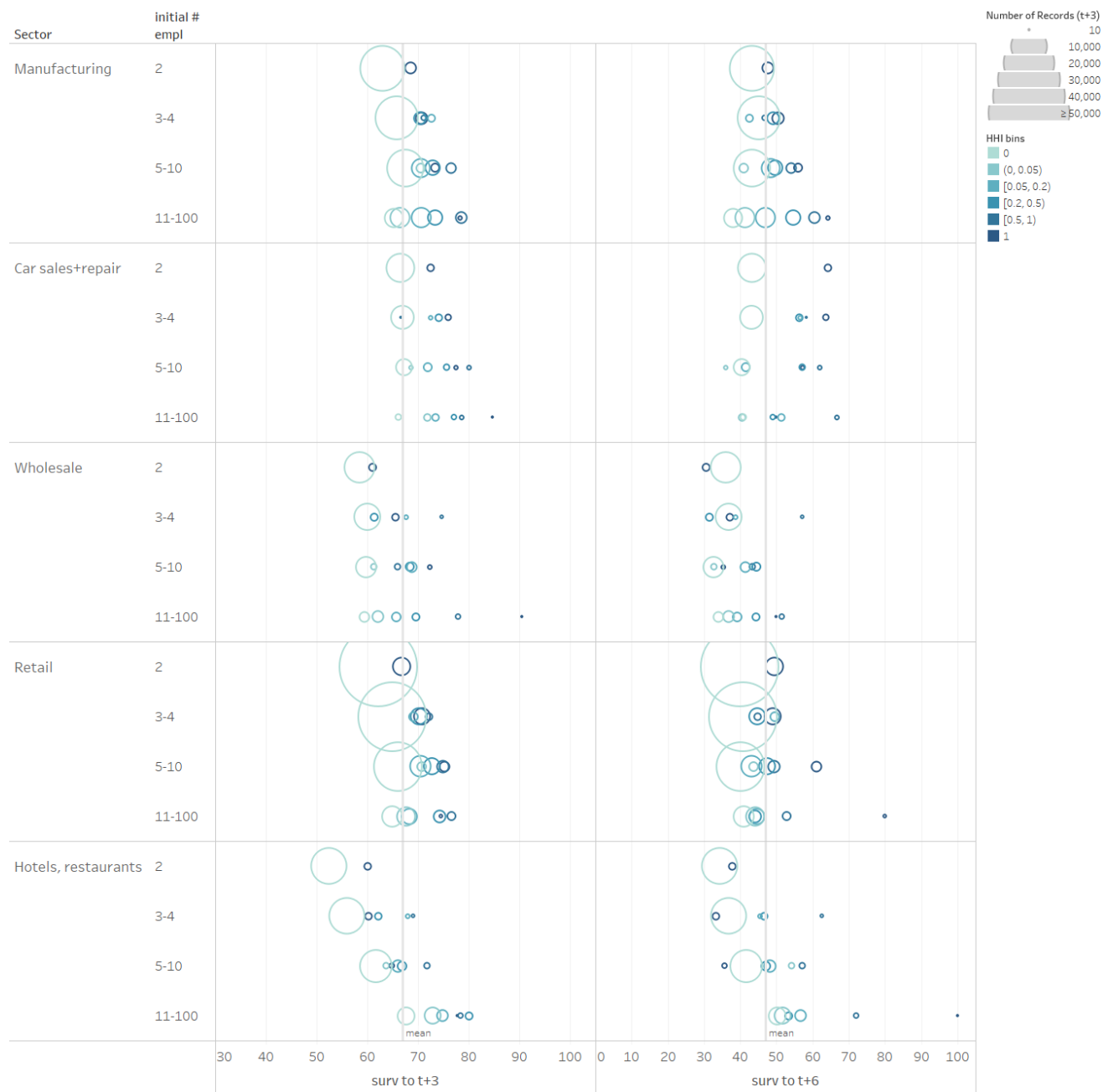


Figure 2: Density and survival rates for the main sectors, by initial size and concentration

Survival at t+3 and t+6 - by sector, initial size, and HHI range, CONT'D

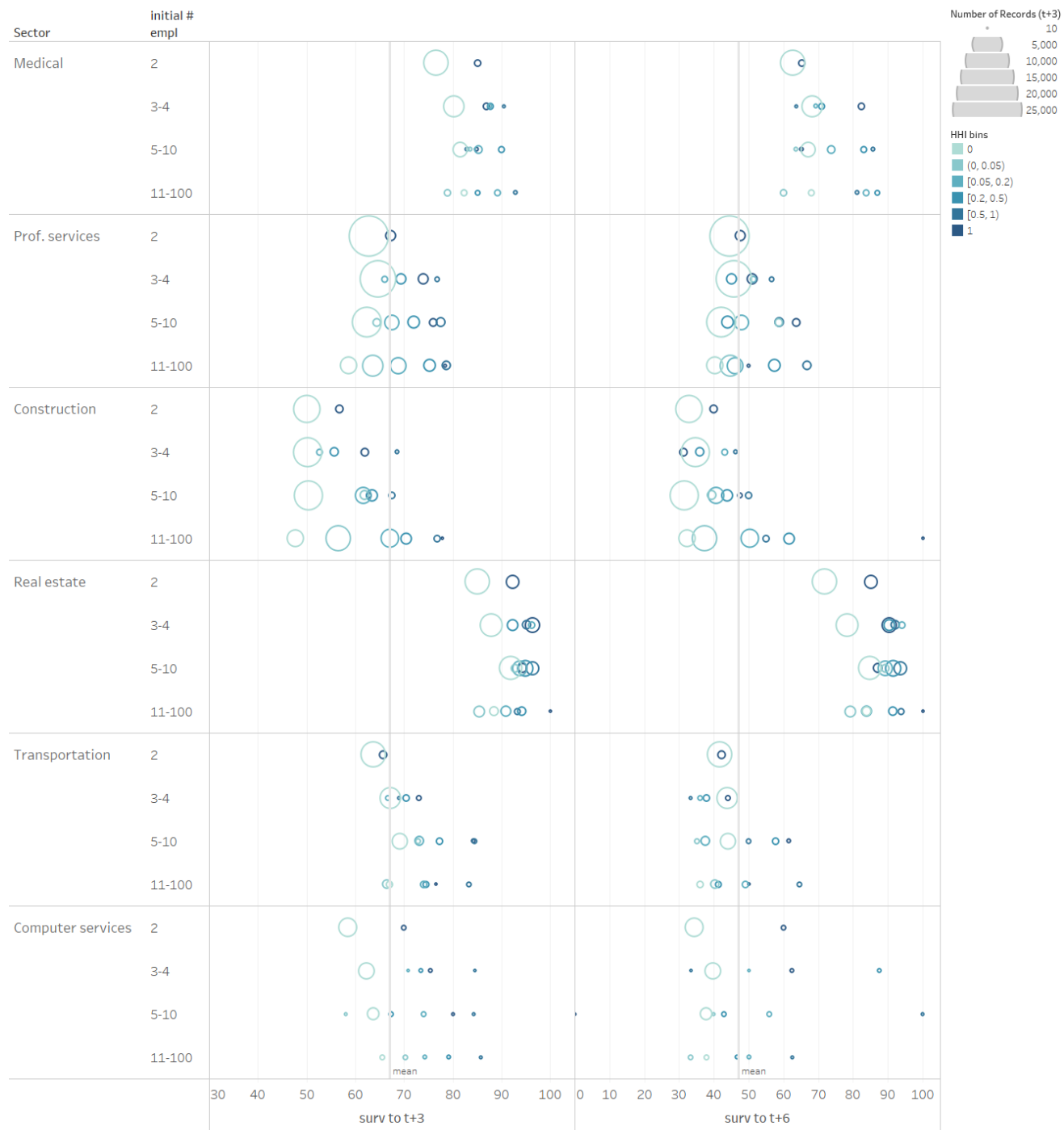


Figure 3: Density and survival rates for the main sectors, by initial size and concentration

Table A16 shows regression results (only the main coefficient) for survival and performance, run separately for each sector. Columns 1 and 2 display Cox hazard model estimation results. Networks decrease the hazard of closure within the first 3 years for every sector except real estate.⁶ One possible reason for the different result in real estate is that individual employees in this sector tend to have direct and exclusive contact with clients, and are thus less likely to be held liable for each other's mistakes. For the subsequent 3 years (column 2), estimates are far less precise, as there is less data to work with. Nonetheless, the coefficient on HHI is uniformly negative, albeit not always significant.

Columns 3 and 5 illustrate the connection between network concentration and average wages. The real estate sector is once again an outlier, with average wages being 10% higher for fully connected over unconnected firms, although it is worth noting that wage growth is also slowed down more severely for connected firms in this sector. Professional services and wholesale firms also pay higher initial wages if they have strong network ties. On the other hand, the estimate is zero or inconclusive for the remaining sectors.

Next, we turn our attention to employment growth. As shown in column 4, growth is slower for connected firms across almost all sectors – with the exception of construction, real estate, and computer services, where the effect is absent or too small to distinguish accurately from zero.

Overall, this analysis tells a consistent story and supports our baseline results regarding survival and employment growth, by showing that they are not driven by a single or even a minority of sectors. The link between networks and compensation continues to be less clearly delineated.

⁶This sector behaves differently along a number of other dimensions, aside from the summary statistics already discussed: for instance the benefit of experience within the same industry is much higher (about 5 times higher than average), and firms survive longer if employees are older and have *less* education. This also illustrates the limitations of looking simply at correlations, as in figure 3.

Table A16: Survival and performance by sector

	Cox survival		log initial ave wage	growth from t to t+3	
	at t+3	at t+6		empl	ave wage
HHI coefficients	(1)	(2)	(3)	(4)	
Manufacturing	-.23 (.03)***	-.12 (.07)*	.001 (.007)	-.12 (.02)***	-.03 (.01)***
Car sales, repair	-.34 (.07)***	-.46 (.15)***	.02 (.01)	-.11 (.03)***	-.01 (.02)
Wholesale	-.23 (.05)***	-.03 (.10)	.03 (.01)**	-.11 (.04)***	.009 (.02)
Retail	-.14 (.02)***	-.17 (.05)***	-.006 (.005)	-.08 (.01)***	-.001 (.007)
Hotels, restaurants	-.29 (.04)***	-.32 (.11)***	-.01 (.01)	-.12 (.03)***	.02 (.02)
Medical activities	-.21 (.11)*	-.36 (.20)*	.01 (.02)	-.14 (.04)***	.008 (.02)
Professional services	-.22 (.04)***	-.13 (.09)	.05 (.01)***	-.21 (.03)***	-.02 (.02)
Construction	-.30 (.05)***	-.17 (.10)*	.02 (.01)	.003 (.06)	-.05 (.02)**
Real estate	.33 (.09)***	-.20 (.17)	.10 (.01)***	-.005 (.02)	-.04 (.01)***
Transportation	-.23 (.07)***	-.33 (.16)**	-.009 (.02)	-.15 (.05)***	.03 (.02)
Computer services	-.41 (.13)***	-.72 (.29)**	.02 (.03)	-.05 (.08)	-.10 (.05)**
Other	-.07 (.04)*	-.14 (.08)*	.004 (.009)	-.09 (.02)***	-.008 (.01)

Notes: Only the main coefficient is reported for brevity, but all other controls from table 4 are included. Robust standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

A.8 Top manufacturing industries

In this section we report summary statistics and empirical estimates across different sub-industries within the manufacturing sector.

Table A17 lists counts, network concentration, and survival statistics for all manufacturing industries (these statistics were presented for a subset of industries in table A15). There is significant variation in network concentration, with Machinery and Transport equipment industries having the highest HHI levels (ignoring Tobacco and Processing of fuels since they saw very little new firm creation). The food and recycling industries have the least tight employee networks at inception.

It is interesting to compare survival statistics with the statistics presented for US firms in Audretsch (1991). Overall survival in the manufacturing sector is remarkably similar, although Brazilian firms have somewhat lower survival rates at $t+4$. Even more striking is the difference in distribution of new firms by manufacturing industry: American industries with highest rates of entry are printing and machinery, whereas in Brazil most entry occurs into relatively low-tech sectors such as apparel, food, stone/clay/glass, and furniture.

Table A18 reports results from survival and performance analysis for the manufacturing industries with at least 5,000 firms in the regression sample. Remarkably – given the reduced sample and the still extensive set of controls – we continue to measure a large and statistically significant effect of networks on early survival (column 1): for all manufacturing industries considered, network ties significantly reduce the hazard of closure during the first 3 years. In the remaining columns, estimate precision is reduced due to the smaller sample size.

Table A17: Network concentration and survival by manufacturing industry

Sector	Num firms	Median ini size	HHI			Share not connected	Share top network	% survived	
			Mean	SD	> 0			t+3	t+6
Food	12,841	4	0.106	0.249	29.2%	0.823	0.141	66.4	47.7
Tobacco	32	13	0.166	0.268	56.3%	0.647	0.251	72.2	20.0
Textiles	3,069	5	0.123	0.255	38.9%	0.772	0.172	66.9	39.0
Apparel	14,423	5	0.108	0.237	36.3%	0.792	0.154	63.9	37.9
Leather	4,918	6	0.121	0.237	48.3%	0.732	0.181	61.1	35.2
Lumber	5,963	7	0.126	0.244	46.3%	0.739	0.186	67.1	46.4
Paper	1,006	7	0.138	0.256	48.4%	0.733	0.202	68.9	47.4
Printing	4,194	4	0.111	0.255	29.0%	0.820	0.146	71.5	50.0
Processing of fuels	24	7	0.171	0.318	45.8%	0.698	0.227	64.7	25.0
Chemicals	2,063	5	0.114	0.242	36.5%	0.790	0.161	70.2	51.9
Rubber and plast	3,375	7	0.127	0.246	45.8%	0.754	0.187	75.6	54.7
Stone, clay, glass	6,255	5	0.127	0.256	39.4%	0.766	0.176	70.4	53.7
Primary metals	1,639	5	0.128	0.255	42.2%	0.762	0.181	75.1	53.6
Fabricated metal	6,894	4	0.110	0.244	34.8%	0.801	0.155	70.4	49.7
Machinery (non-elec)	2,638	5	0.148	0.277	44.5%	0.739	0.204	74.0	50.8
Electrical equipment	1,663	5	0.132	0.258	42.8%	0.760	0.186	71.9	52.4
Instruments	460	4	0.113	0.237	35.0%	0.793	0.161	79.1	58.6
Transport eqpmt	1,415	5	0.142	0.268	44.8%	0.744	0.199	69.8	52.4
Furniture	5,586	5	0.122	0.252	38.2%	0.778	0.172	66.3	42.7
Recycling	455	7	0.074	0.200	35.2%	0.837	0.116	61.9	42.3
Miscellaneous	2,104	5	0.114	0.257	31.2%	0.811	0.152	68.7	50.0
Total	81,017	5	0.117	0.249	37.7%	0.783	0.165	67.8	45.8

Table A18: Survival and performance for the largest manufacturing industries

	Cox survival reg		initial ln(ave wage)	growth from t to t+3	
	t to t+3	t+3 to t+6		ln(empl)	ln(ave wage)
HHI coefficients	(1)	(2)	(3)	(4)	(5)
Food	-.15 (.07)**	-.04 (.15)	-.007 (.01)	-.10 (.06)*	-.03 (.02)
Apparel	-.21 (.09)**	-.05 (.20)	.003 (.02)	-.05 (.08)	-.04 (.03)
Lumber	-.38 (.12)***	-.39 (.29)	-.03 (.02)	-.02 (.08)	-.02 (.04)
Cement, clay, stone	-.27 (.11)**	-.01 (.25)	-.03 (.02)	-.09 (.07)	.001 (.03)
Fabricated metal	-.33 (.13)**	-.29 (.26)	.03 (.02)	-.05 (.08)	-.06 (.04)
Furniture	-.35 (.12)***	-.31 (.28)	-.05 (.02)**	-.31 (.09)***	.03 (.03)

Notes: We analyze manufacturing industries with at least 5,000 new firms in our largest estimation sample. Only the main coefficient is reported for brevity, but all other controls from table 4 are in the regressions (including indicators for finer industry classification, where available). Robust standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

A.9 Additional robustness checks

Tables reported here are in support of the discussion in the main text.

First, we attempt to distinguish between urban and rural areas by looking at agricultural intensity and population density. These two dimensions overlap imperfectly, each splitting the sample into two roughly equal sub-samples. In table A20 we report regression coefficients over these sub-samples: the effect of networks on survival is essentially the same, however tight links are only predictive of higher initial wages and slower wage growth in *urban* (i.e. little agriculture, high population density) areas. This last result is likely due to the different sectoral composition of firms (for instance, real estate and professional services firms are over-represented in areas with low agriculture)

In the next robustness test, we eliminate potentially absorbed firms and find results to be unchanged.

We also separate the sample by the level of financing needs associated with the sector. However, we are missing this indicator for most of the sample, and coefficients are poorly estimated. Nonetheless, we measure a significant positive impact of tight networks on survival even in the low-financing-needs subsample.

Table A19: New firm count and network concentration by agricultural intensity, etc.

Sector	Num firms	Median ini size	HHI			Share not connected	Share top network	% survived	
			Mean	SD	> 0			t+3	t+6
High agricultural intensity ^a	239,844	4	0.116	0.262	31.9%	0.804	0.154	66.6	46.7
Low agricultural intensity	257,094	4	0.094	0.234	29.4%	0.836	0.130	67.2	47.0
High population density ^b	234,244	4	0.104	0.248	30.5%	0.824	0.140	68.2	47.6
Low population density	262,694	4	0.106	0.248	30.8%	0.818	0.143	65.8	46.1
Firms not “absorbed” ^c	547,401	4	0.103	0.247	29.8%	0.826	0.138	67.5	47.6
Firms w/ high finance needs ^d	3,138	7	0.115	0.232	43.8%	0.769	0.173	74.4	56.2
Firms w/ low finance needs	22,719	5	0.114	0.244	37.0%	0.786	0.161	65.9	41.8

^a Agricultural intensity is defined at the municipality level, and it consists of the share of value added that agriculture accounts for in the local economy. The sample is split by the median over the firms in the sample.

^b Population density is available at the state level. Again, the sample is split by the median over the firms in the regression: “high density” includes Sao Paulo and 6 states, while “low density” includes the other 20 states.

^c We label a firm as absorbed if 70% or more of its employees are found employed together at a previously existing firm after the new firm’s demise.

^d High or low financial needs are classified according to Rajan and Zingales (1998). This only applies to 5 percent of our sample.

Table A20: Survival and performance regression analysis, robustness checks

	Cox survival		log initial	growth from t to t+3	
	at t+3	t+3 to t+6	ave wage	empl	ave wage
HHI coefficients	(1)	(2)	(3)	(4)	(5)
Baseline	-.20 (.02)***	-.20 (.03)***	.02 (.008)**	-.10 (.01)***	-.01 (.005)**
High agricultural intensity	-.21 (.02)***	-.16 (.04)***	-.01 (.006)**	-.07 (.01)***	-.0003 (.007)
Low agricultural intensity	-.20 (.03)***	-.18 (.04)***	.08 (.008)***	-.12 (.02)***	-.03 (.007)***
High population density	-.19 (.03)***	-.22 (.04)***	.05 (.008)***	-.11 (.01)***	-.02 (.008)***
Low population density	-.23 (.02)***	-.14 (.04)***	-.003 (.007)	-.09 (.01)***	-.005 (.006)
Excluding “absorbed” firms	-.24 (.02)***	-.22 (.03)***	.02 (.008)**	-.10 (.01)***	-.01 (.005)**
High finance needs	-.05 (.20)	-.29 (.52)	.02 (.02)	.03 (.09)	-.16 (.02)***
Low finance needs	-.17 (.07)***	-.04 (.09)	-.0005 (.01)	-.10 (.04)***	-.01 (.02)

Notes: This is a summary of baseline survival regressions run separately for the different categories of firms described in table A19. Only the HHI coefficient and the number of observations are reported. Standard errors are clustered at the industry level.

A.10 HHI and share-based concentration measures - Examples

All network measures are based solely on workers for whom we have the record of at least one prior job. We only consider new firms with at least 2 employees, but in some cases fewer than 2 can be tracked to a prior employer. In new firms where only one worker can be tracked, we assume he or she has no previous connection to the other workers (who are either coming in from the informal sector or are new to the labor market). In that case, HHI is set to 0, $share\ unconnected=1$, and $share\ top\ network = 0$. Firms wherein *no* workers can be tracked are dropped from the sample.

Here are the possible scenarios of worker connectedness for firms of 3, 4, and 5 initial employees. In the case of firms with 2 workers, HHI and share unconnected take only two possible values: 0 or 1, as does the rescaled share top network measure.

Table A21: Employee network scenarios, 3-person firm

Employee network	Prior employers	HHI value	Shares of:	
			top network	unconnected workers
No connection	X Y Z	0	0	1
Two workers are connected	(XX) Y	1/3	1/2	1/3
Three workers are connected	(XXX)	1	1	0

Table A22: Employee network scenarios, 4-person firm

Employee network	Prior employers	HHI value	Shares of:	
			top network	unconnected workers
No connection	X Y Z V	0	0	1
Two workers are connected	(XX) Y Z	1/6	1/3	1/2
Two pairs of connected workers	(XX) (YY)	1/3	1/3	0
Three workers are connected	(XXX) Y	1/2	2/3	1/4
Four workers are connected	(XXXX)	1	1	0

Table A23: Employee network scenarios, 5-person firm

Employee network	Prior employers	HHI value	Shares of:	
			top network	unconnected workers
No connection	X Y Z V W	0	0	1
Two workers are connected	(XX) Y Z V	1/10	1/4	3/5
Two pairs of connected workers	(XX) (YY) Z	1/5	1/4	1/5
Three workers are connected	(XXX) Y Z	3/10	1/2	2/5
Three and two workers connected	(XXX) (YY)	2/5	1/2	0
Four workers are connected	(XXXX) Y	3/5	3/4	1/5
Five workers are connected	(XXXXX)	1	1	0

A.11 RAIS data cleaning and sample selection

A job spell in RAIS is identified each year by the employee ID, the employer's tax ID (CNPJ), and dates of job accession and separation (if the employee has hired/left during that particular year). To avoid double-counting employees at new firms, we keep only one observation for each employer-employee pair, choosing the job with the earliest hiring date. If the employee has two jobs at the firm starting in the same month, we keep the highest paying one. The rules on tax ID assignments make it possible to identify new firms (the first eight digits of the tax ID) and new plants within firms (the last six digits of the tax ID).

However, the month of separation does not appear to be a reliable indicator of when the employee stopped working at the given firm, hence we implicitly assume the employee was working the entire year if he or she appeared

From the full universe of new businesses, we exclude branches of government, firms with state ownership, cooperatives, any type of holding company, and branches of foreign firms - as indicated by the legal form variable (*natureza juridica*) in the dataset.⁷ We also limit our analyses to new firms that are under 100 employees since it is unclear that a firm starting at a very large initial size is truly a new start-up. In our final sample, 91.6% of new firms have 10 or fewer employees while 96.4% of new firms have 20 or fewer employees.

Our data contains no information on ownership, and only a small minority of new firms in Brazil have a management employee on their payroll. Relying on the latter severely limits our sample size without reliably providing the upside of accurately identifying founders - who may not appear in the payroll altogether, especially if they simply pay themselves in equity. Another possibility is to employ wages to identify founders. However, it is unclear that the founder will always pay herself more, given that she has residual claims on the firm's profits.

⁷See table C.1 in Muendler, Rauch and Tocoian (2012) for a full list of legal forms in the data.

References

Audretsch, David B, “New-firm survival and the technological regime,” *The Review of Economics and Statistics*, 1991, pp. 441–450.

Muendler, Marc-Andreas, James E Rauch, and Oana Tocoian, “Employee spinoffs and other entrants: Stylized facts from Brazil,” *International Journal of Industrial Organization*, 2012, 30 (5), 447–458.

Rajan, Raghuram G. and Luigi Zingales, “Financial dependence and growth,” *American Economic Review*, 1998, 88, 559–586.