

ONLINE APPENDIX

Appendix A: Database Construction Procedures, Summary Statistics, Correlation Matrix, PSM Tests, and Control Variables

Database Construction Procedures

To obtain unique identifiers for entrepreneurs/firms in the Chinese Private Entrepreneurs Survey (CPES) Data from the Chinese Academy of Social Science (CASS), one approach is to use initial conditions and other fixed characteristics. Same firms have identical values for these variables, and this facilitates identifying the common firms over time. These identifying variables include founding conditions such as initial sources of funding, initial financial indicators (e.g., total assets, sales, registered capital, taxes, and fees), and initial employee makeup (e.g., number of technicians, managers, and workers), all of which are reported in different survey waves. This approach has been confirmed by data providers as an accurate way to identify firms and has been used by other researchers. These are the basic steps of this procedure:

1. Assuming a researcher is working from the individual raw data files for each year, s/he would first need to clean these files, particularly making variable names consistent. Generating a codebook to indicate the common names and available years of each variable is a helpful first step. The researcher needs to attend to this closely as order of questions, naming conventions, and other details in the survey questionnaires were not consistent over time. Generate the line number of each firm in the original dataset, and this will be the year-ID of that firm in that year. After this, the researcher can pool the original datasets together to get a full sample encompassing all the available years.
2. To get the dictionary linking year, year-IDs and the unique-IDs, merge pairs of datasets from survey waves by relying on the codebook generated from the step above, which provides availability of these identifying variables. For instance, the researcher can start by matching 1993 and 1995 data with the available identifying variables, e.g., initial investments variables. Then 1993 with 1997, 2000, 2002, and so on; 1995 with 1997, 2000, 2002, and so on; 1997 with 2000, 2002, and so on; and so on. For each pair of two waves, select the generic identifying variables that are available—such as initial sources of funding—first, and if there are too many missing values and/or inconsistencies (e.g., in terms of rounding and other reporting errors), then use other identifying variables to match the two waves. Due to different sources of inconsistency such as input or reporting errors, frequently manual inspection and human judgment is necessary to determine whether the two observations indeed match. This step results in pairs of matches of year-IDs, e.g., IDs in 1993 to be linked to IDs in 1995.

3. Regarding the unique-IDs, the researcher can set 1993 as the benchmark and begin with that year, then unique-IDs would be equal to the year-IDs or the line number in that year—1993 in this example. Find out which lines can be matched with 1993's lines in the 1995 dataset and add that corresponding unique-ID to the year-ID in 1995. Continue in this fashion, add unique-IDs for the 1997 to 2012 data. Then continue year by year and assign unique-IDs to the unmatched data. For instance, if the first unmatched line in 1995 is line (year-ID) 4, continue numbering that line based on the next number in the unique-ID sequence. For example, if the largest unique-ID number for 1993 is 1000, then assign 1001 to the line 4 of 1995 and continue until every unmatched line has a unique-ID. Then use the unmatched 1995 lines with the newly assigned unique-IDs to match 1997, 2000, and so on, continuing in this fashion to get a year-ID-line dictionary. Note that the researcher would need to cross-check the match over time after the series of dyadic matches, i.e., if line 5 in 1995 and line 9 in 1997 are both matched to line 3 in 1993, then the first two should be identical in other identifying variables—if available. Again, this process involves manual inspection and human judgment in case of different sources of inconsistency such as rounding issues of decimal places (sometimes different waves report different decimal levels), missing values (and thus need other identifying variables), and potential input or reporting errors.
4. Finally, generate a dataset containing year-ID-line for all years by pooling them together. Then merge the pooled year-ID-line with the entire sample (from step 1) to create an overall database that includes unique identifiers.

Table A1. Summary Statistics and Correlation Matrix (N = 19,729)*

Variable	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Inward internationalization (0/1)	.15	.36															
2. Inward internationalization (foreign investment/total)	.03	.06	.99														
3. Outward internationalization (0/1)	.12	.32	.23	.26													
4. Outward internationalization (overseas assets/total)	.02	.06	.26	.30	.95												
5. Ideological imprint	.25	.43	-.08	-.08	-.06	-.09											
6. Political involvement	.42	.49	.04	.06	.08	.10	-.03										
7. Government appropriation	.14	.09	.05	.06	.04	.05	.01	-.01									
8. Social network	.71	.46	.11	.11	.09	.09	.23	.18	-.01								
9. Post 2001	.69	.46	.12	.13	.17	.17	.13	-.32	.02	-.34							
10. Regional FDI intensity	.56	.40	-.01	.00	-.01	.01	.00	.04	.00	.06	-.18						
11. Age	45.35	8.84	-.02	-.01	.01	.01	-.01	.02	.01	-.08	.16	-.05					
12. Educational attainment	.28	.45	.00	.00	-.01	-.01	-.02	.11	-.01	-.05	-.13	-.01	-.03				
13. Current communist ideology	.44	.50	.24	.24	.25	.27	.43	-.21	.01	.14	.55	-.09	.08	-.09			
14. Foreign experience	.29	.45	.11	.12	.09	.10	.23	.13	.00	.20	.04	.00	-.02	-.01	.17		
15. Government work experience	.39	.49	.03	.03	.03	.04	.03	.04	.00	-.27	.19	-.04	.03	.07	.05	.23	
16. Firm size	3.68	1.65	.03	.04	.05	.05	.00	.17	.01	-.05	.05	-.01	.11	.02	.01	-.03	.02
17. Financial leverage	.10	.20	.09	.09	.08	.08	.05	.03	-.01	.13	.01	.01	.04	-.01	.14	.13	-.04
18. Firm age	7.52	5.68	.09	.09	.13	.12	-.08	-.24	.01	-.36	.65	-.17	.21	-.04	.33	-.14	.20
19. Industry average (inward internationalization 0/1)	.15	.07	.19	.18	.14	.14	.33	-.26	.01	.31	.31	.00	-.03	-.10	.56	.23	-.04
20. Industry average (inward internationalization, continuous)	.03	.01	.18	.18	.14	.14	.34	-.24	.01	.31	.31	-.01	-.03	-.10	.55	.27	-.02
21. Industry average (outward internationalization 0/1)	.14	.05	.18	.19	.23	.22	.25	-.26	.01	.04	.73	-.18	.10	-.08	.65	.10	-.05
22. Industry average (outward internationalization, continuous)	.02	.01	.17	.17	.15	.15	.40	-.18	.01	.38	.34	-.03	-.04	-.10	.56	.33	-.03
23. Manufacturing firm	.47	.50	.00	-.01	-.01	-.01	.02	-.11	.00	.05	-.02	.03	.01	-.07	.06	-.30	-.04
24. Poor Internet coverage	.11	.02	-.01	-.01	-.03	-.03	.00	.03	.00	-.06	.00	.04	.01	.02	-.05	.03	.02
25. Density of CPC membership	.06	.01	.00	.00	.01	.02	.01	-.04	.01	.01	.17	-.01	.03	-.07	.06	-.05	-.01
26. GDP per capita (logged)	9.65	.87	.05	.05	.11	.11	-.06	-.29	.01	-.36	.64	.14	.16	-.05	.25	-.16	.19
27. Population growth	4.54	3.65	-.02	-.02	-.08	-.07	.09	.16	.00	.39	-.43	.04	-.12	-.01	-.10	.12	-.16
28. Institutional development	6.96	2.45	.08	.08	.12	.12	.01	-.28	.01	-.22	.60	.26	.13	-.09	.31	-.09	.09
29. R&D investment (0/1)	.57	.50	.15	.15	.18	.20	.26	-.25	.01	-.05	.67	-.11	.03	-.12	.63	.01	.22
30. Firm performance	.14	.28	-.07	-.07	-.09	-.08	-.12	.07	.00	-.04	-.22	.09	-.06	.01	-.21	-.19	-.15
31. Industry competitiveness	.91	.22	.08	.09	.11	.10	-.04	-.14	.01	-.15	.46	-.06	.11	-.02	.29	-.30	-.24
32. Industry growth	.00	.05	.00	.00	.00	.00	.00	.01	.00	.00	.00	-.01	.00	.00	.00	-.01	.01
Variable	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
17. Financial leverage	.05																
18. Firm age	.08	.00															
19. Industry average (inward internationalization 0/1)	-.08	.15	.08														
20. Industry average (inward internationalization, continuous)	-.08	.16	.07	.99													
21. Industry average (outward internationalization 0/1)	.02	.12	.51	.63	.64												
22. Industry average (outward internationalization, continuous)	-.07	.13	.04	.92	.93	.66											
23. Manufacturing firm	.03	-.07	.05	.02	-.02	-.02	-.08										
24. Poor Internet coverage	-.01	-.02	-.05	-.03	-.03	-.06	-.03	-.03									
25. Density of CPC membership	.02	-.03	.07	.02	.02	.08	.04	.03	-.05								
26. GDP per capita (logged)	.03	-.06	.62	.04	.03	.43	.02	.09	-.06	.05							
27. Population growth	-.04	.08	-.47	.08	.09	-.27	.10	-.04	-.01	-.06	-.68						
28. Institutional development	.02	-.03	.50	.16	.15	.42	.14	.12	-.04	.03	.66	-.51					
29. R&D investment (0/1)	.01	.07	.45	.47	.47	.67	.51	.06	-.05	.12	.39	-.19	.39				
30. Firm performance	-.06	.10	-.21	-.22	-.22	-.31	-.23	-.04	.01	-.01	-.19	.16	-.15	-.20			
31. Industry competitiveness	.04	.10	.35	.18	.17	.43	.17	-.19	-.03	.06	.26	-.17	.26	.35	.07		
32. Industry growth	.01	-.01	.00	-.02	-.02	.00	-.01	-.01	-.01	.01	.01	.00	.00	.00	.00	-.03	

* Pearson correlation tests are used if both variables are continuous, and Spearman rank tests are employed otherwise. Coefficients of correlations over .015 are significant at the 1% level.

Table A2. Probit Regression Results, Pre- and Post-match Differences, and Percentage Bias Reduction

Variable	Pre-match	Pre- versus Post-match	Bias (%)	
	1 Results	2 <i>p</i> -value of differences	3 Reduction	4 Post-match
Age	.000 (.001)	.106 versus .398	38.6	1.7
Educational attainment	.129** (.026)	.001 versus .012	10.5	4.9
Current communist ideology	1.803** (.043)	.000 versus .027	96.4	4.1
Foreign experience	.219** (.033)	.000 versus .269	95.5	-2.4
Government work experience	-.056* (.027)	.000 versus .000	-70.4	11.7
Firm size	.015* (.007)	.813 versus .267	-490.6	2.3
Financial leverage	-.180** (.056)	.000 versus .250	79.2	-2.5
Firm age	-.085** (.003)	.000 versus .087	84.7	2.9
Manufacturing firm	.067* (.029)	.024 versus .001	-88.1	-7
Poor Internet coverage	-.010 (.733)	.727 versus .985	92.9	0
Density of CPC membership	-1.695 (1.724)	.207 versus .996	99.5	0
GDP per capita (logged)	-.291** (.035)	.000 versus .260	86.6	2
Population growth	-.008+ (.004)	.000 versus .112	85.6	-3
Institutional quality	.030** (.010)	.087 versus .756	79.4	.6
R&D investment (0/1)	.350** (.041)	.000 versus .001	90.3	6.3
Firm performance (return on assets)	-.199** (.050)	.000 versus .722	98.1	.6
Industry competitiveness	-.870** (.068)	.000 versus .008	-19.3	9.2
Industry growth	.027 (.223)	.940 versus .434	-1007.4	1.4
Number of observations	19,730	Mean bias (%)	19.9	3.5
Goodness of fit (pseudo R ²)	.280	Pseudo R ²	.280	.007

+ $p < .10$; * $p < .05$; ** $p < .01$; intercept not reported to save space.

Figure A1. Graphical illustration of PSM matching quality.

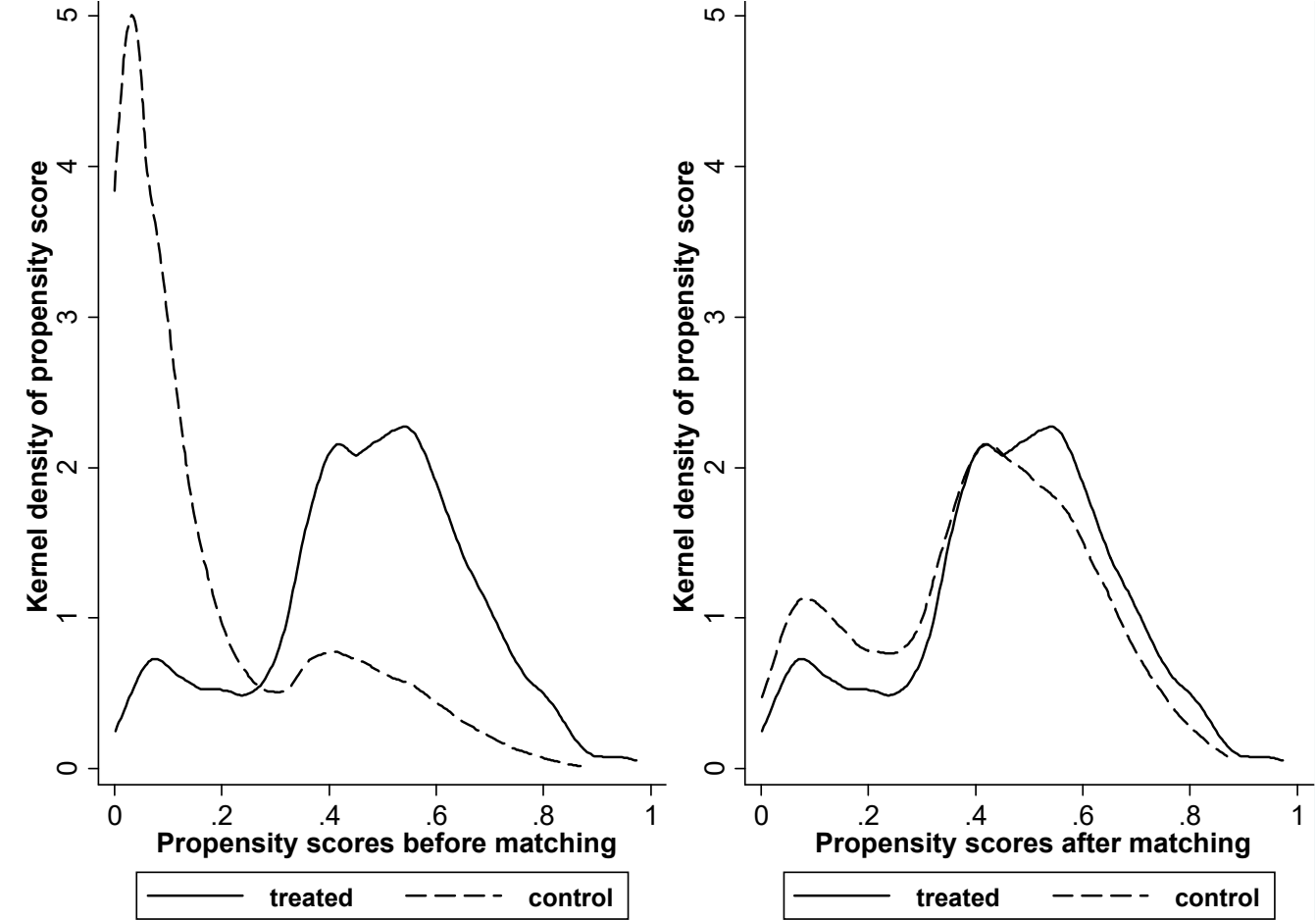


Table A3a. Results of Control Variables in Table 2*

Variable	1	2	3	4	5	6	7	8
Panel A: First-stage results of Heckman (random effects probit) after PSM								
Age	-.009** (.001)	-.010** (.001)	-.010** (.001)	-.010** (.001)	-.010** (.001)	-.010** (.001)	-.010** (.001)	-.010** (.001)
Educational attainment	.073** (.024)	.066** (.025)	.069** (.025)	.065** (.025)	.072** (.025)	.063* (.025)	.066** (.025)	.077** (.025)
Current communist ideology	.522** (.032)	.551** (.034)	.531** (.034)	.550** (.034)	.583** (.034)	.536** (.034)	.551** (.034)	.537** (.035)
Government work experience	.298** (.030)	.248** (.034)	.240** (.034)	.248** (.034)	.235** (.034)	.261** (.034)	.247** (.034)	.232** (.035)
Foreign experience	.107** (.026)	.184** (.027)	.170** (.027)	.182** (.027)	.184** (.027)	.196** (.027)	.183** (.027)	.171** (.028)
Firm size	.026** (.006)	.015* (.007)	.013* (.007)	.014* (.007)	.015* (.007)	.016* (.007)	.015* (.007)	.013+ (.007)
Financial leverage	.424** (.050)	.360** (.051)	.353** (.051)	.359** (.051)	.351** (.051)	.358** (.051)	.359** (.051)	.327** (.052)
Firm age	.015** (.003)	.025** (.003)	.027** (.003)	.026** (.003)	.021** (.003)	.026** (.003)	.025** (.003)	.025** (.003)
Manufacturing firm	.178** (.027)	.179** (.028)	.184** (.028)	.175** (.028)	.179** (.028)	.178** (.028)	.179** (.028)	.182** (.028)
Industry average (for corresponding dependent variable)	2.074** (.227)	2.082** (.241)	2.241** (.242)	2.058** (.242)	2.022** (.242)	2.207** (.243)	2.080** (.241)	2.460** (.248)
Poor Internet coverage	-.446 (.713)	-.208 (.729)	-.107 (.732)	-.239 (.730)	-.205 (.731)	-.129 (.730)	-.174 (.729)	.104 (.737)
Density of CPC membership	-3.423* (1.612)	-4.518** (1.651)	-4.749** (1.657)	-4.373** (1.654)	-4.467** (1.653)	-4.170* (1.653)	-4.603** (1.653)	-4.283* (1.669)
GDP per capita (logged)	-.077* (.033)	-.006 (.035)	-.008 (.035)	-.012 (.035)	-.023 (.035)	.009 (.035)	-.003 (.035)	-.007 (.035)
Population growth	-.005 (.004)	-.011* (.004)	-.011* (.004)	-.011** (.004)	-.012** (.004)	-.010* (.004)	-.010* (.004)	-.011* (.004)
Institutional development	.006 (.009)	-.007 (.010)	-.007 (.010)	-.006 (.010)	-.009 (.010)	-.008 (.010)	-.009 (.010)	-.013 (.010)
R&D investment (0/1)	-.152** (.036)	-.024 (.047)	-.055 (.047)	-.019 (.047)	-.016 (.046)	-.040 (.047)	-.024 (.047)	-.081+ (.046)
Firm performance (return on assets)	-.171** (.050)	-.144** (.052)	-.146** (.051)	-.145** (.051)	-.154** (.052)	-.152** (.052)	-.142** (.051)	-.170** (.052)
Industry competitiveness	.504** (.078)	.588** (.084)	.608** (.084)	.587** (.084)	.582** (.084)	.499** (.084)	.590** (.084)	.473** (.085)
Industry growth	.054 (.269)	.030 (.273)	.031 (.278)	-.024 (.282)	.033 (.275)	.009 (.276)	.033 (.273)	-.072 (.300)

Panel B: Second-stage results of Heckman (random effects) after PSM								
Age	.001**	.001**	.001**	.001**	.001**	.001**	.001**	-.000**
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Educational attainment	-.007**	-.008**	-.008**	-.008**	-.008**	-.008**	-.008**	.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Current communist ideology	-.036**	-.049**	-.048**	-.047**	-.047**	-.049**	-.047**	.003**
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Government work experience	-.014**	-.014**	-.014**	-.014**	-.014**	-.014**	-.014**	.000
	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Foreign experience	-.003**	-.011**	-.011**	-.010**	-.011**	-.011**	-.011**	.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Firm size	-.001**	-.001**	-.001**	-.001**	-.001**	-.001**	-.001**	.000**
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Financial leverage	-.019**	-.027**	-.026**	-.025**	-.026**	-.027**	-.026**	-.000
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Firm age	-.001**	-.002**	-.002**	-.002**	-.002**	-.002**	-.002**	.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Manufacturing firm	-.011**	-.012**	-.012**	-.012**	-.011**	-.012**	-.012**	.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Industry average (for corresponding dependent variable)	-.896**	-1.239**	-1.230**	-1.207**	-1.198**	-1.240**	-1.205**	-.026
	(.025)	(.022)	(.022)	(.022)	(.022)	(.022)	(.021)	(.020)
Poor Internet coverage	-.017	-.034**	-.034**	-.035**	-.034**	-.034**	-.025**	-.004
	(.013)	(.009)	(.009)	(.009)	(.009)	(.009)	(.009)	(.005)
Density of CPC membership	.279**	.366**	.361**	.359**	.359**	.366**	.353**	-.009
	(.029)	(.021)	(.021)	(.020)	(.020)	(.021)	(.020)	(.013)
GDP per capita (logged)	.005**	.001**	.001**	.001*	.001**	.001**	.001**	-.001*
	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Population growth	.001**	.001**	.001**	.001**	.001**	.001**	.001**	.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Institutional development	.001**	.001**	.001**	.001**	.001**	.001**	.001**	.000**
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)

⁺ $p < .10$; * $p < .05$; ** $p < .01$; intercept not reported to save space.

* 17,934/2,785 out of 19,729/3,011 observations are retained for the PSM analysis.

Table A3b. Results of Control Variables in Table 3*

Variable	1	2	3	4	5	6	7	8
Panel A: First-stage results of Heckman (random effects probit) after PSM								
Age	-.005** (.001)	-.006** (.001)	-.006** (.001)	-.006** (.001)	-.006** (.001)	-.006** (.001)	-.006** (.001)	-.006** (.001)
Educational attainment	.072** (.027)	.052+ (.028)	.062* (.028)	.052+ (.028)	.055* (.028)	.054* (.028)	.051+ (.028)	.079** (.029)
Current communist ideology	.517** (.034)	.609** (.036)	.526** (.037)	.609** (.037)	.645** (.037)	.598** (.037)	.600** (.036)	.526** (.037)
Government work experience	.292** (.033)	.213** (.038)	.200** (.039)	.214** (.038)	.202** (.039)	.219** (.038)	.212** (.038)	.192** (.040)
Foreign experience	.094** (.028)	.175** (.030)	.155** (.031)	.172** (.030)	.175** (.030)	.172** (.030)	.165** (.031)	.120** (.032)
Firm size	.032** (.007)	.015* (.007)	.009 (.007)	.014+ (.007)	.015* (.007)	.016* (.007)	.014+ (.007)	.009 (.008)
Financial leverage	.253** (.056)	.150** (.058)	.135* (.059)	.148* (.058)	.143* (.058)	.144* (.058)	.147* (.058)	.094 (.060)
Firm age	-.000 (.003)	.016** (.003)	.020** (.003)	.017** (.003)	.012** (.003)	.017** (.003)	.017** (.003)	.018** (.003)
Manufacturing firm	.140** (.029)	.162** (.032)	.194** (.032)	.157** (.032)	.159** (.032)	.166** (.032)	.165** (.032)	.211** (.033)
Industry average (for corresponding dependent variable)	1.150** (.359)	1.682** (.387)	2.517** (.396)	1.654** (.388)	1.481** (.390)	1.705** (.388)	1.687** (.388)	2.561** (.407)
Poor Internet coverage	-.945 (.773)	-.374 (.807)	.124 (.824)	-.392 (.808)	-.399 (.810)	-.361 (.807)	-.233 (.811)	.429 (.839)
Density of CPC membership	-3.046+ (1.706)	-4.059* (1.786)	-4.637* (1.819)	-3.874* (1.793)	-3.850* (1.790)	-3.981* (1.787)	-4.824** (1.804)	-4.832** (1.860)
GDP per capita (logged)	-.035 (.037)	.108** (.039)	.098* (.039)	.101** (.039)	.086* (.039)	.111** (.039)	.128** (.039)	.086* (.040)
Population growth	-.014** (.005)	-.020** (.005)	-.021** (.005)	-.021** (.005)	-.021** (.005)	-.020** (.005)	-.018** (.005)	-.020** (.005)
Institutional development	.005 (.010)	-.007 (.011)	-.002 (.011)	-.005 (.011)	-.007 (.011)	-.007 (.011)	-.013 (.011)	-.008 (.012)
R&D investment (0/1)	-.093* (.040)	.190** (.052)	.071 (.051)	.197** (.052)	.193** (.051)	.189** (.052)	.200** (.052)	.072 (.051)
Firm performance (return on assets)	-.309** (.071)	-.190* (.076)	-.226** (.077)	-.196** (.076)	-.197** (.076)	-.188* (.076)	-.179* (.076)	-.234** (.079)
Industry competitiveness	.236 (.154)	1.279** (.201)	1.493** (.212)	1.282** (.202)	1.250** (.199)	1.282** (.204)	1.267** (.200)	1.535** (.224)
Industry growth	-.286 (.325)	-.221 (.325)	-.238 (.341)	-.273 (.327)	-.227 (.327)	-.223 (.325)	-.223 (.326)	-.352 (.346)
Inward internationalization [†]	2.220** (.649)	.448 (.693)	.513 (.709)	.439 (.697)	.180 (.699)	.461 (.693)	.492 (.704)	.196 (.734)

Panel B: Second-stage results of Heckman (random effects) after PSM								
Age	-.000** (.000)	-.000+ (.000)	-.000* (.000)	-.000* (.000)	-.000+ (.000)	-.000* (.000)	-.000* (.000)	-.001** (.000)
Educational attainment	-.011** (.002)	-.009** (.002)	-.009** (.002)	-.009** (.002)	-.009** (.002)	-.009** (.002)	-.009** (.002)	-.001 (.002)
Current communist ideology	.022** (.003)	.011** (.003)	.016** (.003)	.012** (.003)	.012** (.003)	.011** (.003)	.012** (.003)	.067** (.005)
Government work experience	-.016** (.002)	-.024** (.003)	-.022** (.003)	-.024** (.003)	-.023** (.003)	-.024** (.003)	-.024** (.003)	-.006* (.003)
Foreign experience	.002 (.002)	-.011** (.002)	-.010** (.002)	-.011** (.002)	-.011** (.002)	-.011** (.002)	-.011** (.002)	-.001 (.002)
Firm size	-.000 (.000)	-.001 (.000)	-.000 (.000)	-.001 (.000)	-.001 (.000)	-.001 (.000)	-.001 (.000)	.001 (.000)
Financial leverage	-.002 (.004)	-.002 (.003)	-.002 (.003)	-.002 (.003)	-.002 (.003)	-.002 (.003)	-.002 (.003)	.004 (.003)
Firm age	-.000 (.000)	-.001** (.000)	-.001** (.000)	-.001** (.000)	-.002** (.000)	-.001** (.000)	-.001** (.000)	.000 (.000)
Manufacturing firm	-.011** (.002)	-.013** (.002)	-.012** (.002)	-.013** (.002)	-.013** (.002)	-.013** (.002)	-.013** (.002)	-.000 (.002)
Industry average (for corresponding dependent variable)	-1.336** (.117)	-1.725** (.134)	-1.597** (.139)	-1.714** (.134)	-1.712** (.134)	-1.716** (.134)	-1.710** (.134)	-.381* (.157)
Poor Internet coverage	-.048 (.052)	-.038 (.050)	-.034 (.050)	-.038 (.050)	-.043 (.050)	-.039 (.050)	-.036 (.050)	-.024 (.048)
Density of CPC membership	.408** (.113)	.533** (.109)	.481** (.110)	.531** (.109)	.534** (.109)	.532** (.109)	.522** (.110)	.017 (.111)
GDP per capita (logged)	.007** (.003)	-.005+ (.003)	-.004 (.003)	-.005+ (.003)	-.005+ (.003)	-.005+ (.003)	-.004+ (.003)	.003 (.003)
Population growth	.001* (.000)	.001** (.000)	.001** (.000)	.001** (.000)	.001** (.000)	.001** (.000)	.001** (.000)	-.001** (.000)
Institutional development	-.001* (.001)	-.002** (.001)	-.002** (.001)	-.002** (.001)	-.002** (.001)	-.002** (.001)	-.002** (.001)	-.002** (.001)
Inward internationalization [†]	-.017 (.049)	.058 (.046)	.063 (.046)	.059 (.046)	.034 (.047)	.059 (.046)	.063 (.046)	.029 (.046)

⁺ $p < .10$; * $p < .05$; ** $p < .01$; intercept not reported to save space.

* 14,604/2,178 out of 16,025/2,342 observations are retained for the PSM analysis.

[†] Foreign investment over total assets was instrumented by presence of foreign concession, whose test results—the first-stage F-statistic is over 29 and above the critical value of 5-percent bias (24.06)—suggested they are not weak instruments.

Appendix B: Robustness Checks and Additional Analyses

Table B1a. Results from Heckman Model: Inward Internationalization (1993–2012)

Variable	1	2	3	4	5	6	7
Panel A: First-stage results of Heckman (random effects probit); controls are the same as Panel A of table A3a and not reported							
Ideological imprint (H1)	-.170** (.030)	-.338** (.037)	-.471** (.050)	-.996** (.119)	-.778** (.090)	-.261** (.050)	-3.017** (.189)
Political involvement	.310** (.028)	.192** (.032)	.306** (.028)	.328** (.028)	.309** (.028)	.309** (.028)	.158** (.032)
Government appropriation	.813** (.125)	.796** (.125)	.179 (.150)	.821** (.125)	.825** (.125)	.814** (.125)	.169 (.149)
Social network	.367** (.038)	.395** (.038)	.368** (.038)	.255** (.041)	.348** (.038)	.368** (.038)	.240** (.041)
Post 2001	-.173* (.074)	-.157* (.073)	-.170* (.074)	-.151* (.074)	-.265** (.075)	-.168* (.074)	-.257** (.074)
Regional FDI intensity	.016 (.038)	.015 (.038)	.018 (.038)	.037 (.038)	.008 (.038)	-.022 (.042)	-.025 (.041)
Ideological imprint × Political involvement (H2a)		.413** (.053)					.667** (.059)
Ideological imprint × Government appropriation (H2b)			2.098** (.275)				2.352** (.285)
Ideological imprint × Social network (H3)				.883** (.120)			1.178** (.128)
Ideological imprint × Post 2001 (H4a)					.705** (.095)		1.133** (.103)
Ideological imprint × Regional FDI intensity (H4b)						.164* (.072)	.291** (.079)
Number of observations	19,729	19,729	19,729	19,729	19,729	19,729	19,729
Goodness of fit (chi-squared)	1,482.43	1,539.91	1,531.15	1,525.93	1,486.45	1,487.55	1,686.99
Panel B: Second-stage results of Heckman (random effects linear); controls are the same as Panel B of table A3a and not reported							
Inverse Mills ratio	-.115** (.001)	-.110** (.001)	-.112** (.001)	-.112** (.001)	-.115** (.001)	-.112** (.001)	.003+ (.002)
Ideological imprint (H1)	-.006** (.000)	-.009** (.001)	-.011** (.001)	-.024** (.002)	-.007** (.002)	-.014** (.001)	-.315** (.004)
Political involvement	-.014** (.000)	-.015** (.000)	-.013** (.000)	-.013** (.000)	-.014** (.000)	-.013** (.000)	.000 (.000)
Government appropriation	-.006** (.002)	-.002 (.002)	-.014** (.002)	-.003+ (.002)	-.006** (.002)	-.004* (.002)	.001 (.001)
Social network	-.022** (.001)	-.020** (.001)	-.021** (.001)	-.022** (.001)	-.022** (.001)	-.021** (.001)	.000 (.000)
Post 2001	.018** (.001)	.018** (.001)	.018** (.001)	.018** (.001)	.018** (.001)	.019** (.001)	.001 (.001)
Regional FDI intensity	.006** (.001)	.006** (.001)	.006** (.001)	.006** (.001)	.006** (.001)	.002** (.001)	.000 (.000)
Ideological imprint × Political involvement (H2a)		.007** (.001)					.058** (.001)
Ideological imprint × Government appropriation (H2b)			.031** (.004)				.220** (.003)
Ideological imprint × Social network (H3)				.019** (.002)			.119** (.002)
Ideological imprint × Post 2001 (H4a)					.001 (.002)		.116** (.002)
Ideological imprint × Regional FDI intensity (H4b)						.014** (.001)	.038** (.001)
Number of observations	3,011	3,011	3,011	3,011	3,011	3,011	3,011
Goodness of fit (chi-squared)	21,642.43	22,264.73	22,177.16	22,233.30	21,637.40	23,280.26	75,320.29

+ $p < .10$; * $p < .05$; ** $p < .01$; intercept not reported to save space.

Table B1b. Results from Heckman Model: Outward Internationalization (2000–2012)

Variable	1	2	3	4	5	6	7
Panel A: First-stage results of Heckman (random effects probit); controls are the same as Panel A of table A3b and not reported							
Ideological imprint (H1)	-.274** (.033)	-.864** (.046)	-.647** (.056)	-1.353** (.161)	-.876** (.177)	-.763** (.062)	-5.564** (.304)
Political involvement	.572** (.033)	.194** (.038)	.569** (.033)	.588** (.033)	.572** (.033)	.573** (.033)	.169** (.038)
Government appropriation	.644** (.140)	.603** (.143)	-.143 (.169)	.658** (.140)	.648** (.140)	.652** (.140)	-.172 (.167)
Social network	.452** (.042)	.534** (.042)	.454** (.042)	.333** (.044)	.448** (.042)	.463** (.042)	.370** (.045)
Post 2001	-.444** (.085)	-.427** (.085)	-.440** (.086)	-.414** (.085)	-.491** (.086)	-.439** (.086)	-.529** (.087)
Regional FDI intensity	.089+ (.053)	.080 (.054)	.090+ (.053)	.128* (.053)	.087 (.053)	-.167** (.060)	-.138* (.061)
Ideological imprint × Political involvement (H2a)		1.316** (.063)					1.686** (.070)
Ideological imprint × Government appropriation (H2b)			2.592** (.305)				3.510** (.346)
Ideological imprint × Social network (H3)				1.142** (.163)			1.826** (.178)
Ideological imprint × Post 2001 (H4a)					.634** (.180)		1.762** (.197)
Ideological imprint × Regional FDI intensity (H4b)						.885** (.092)	1.176** (.105)
Number of observations	16,025	16,025	16,025	16,025	16,025	16,025	16,025
Goodness of fit (chi-squared)	1,262.74	1,690.39	1,323.28	1,301.50	1,258.02	1,335.89	1,862.58
Panel B: Second-stage results of Heckman (random effects linear); controls are the same as Panel B of Table A3b and not reported							
Inverse Mills ratio	-.086** (.003)	-.075** (.004)	-.084** (.003)	-.085** (.003)	-.084** (.003)	-.083** (.003)	-.048** (.010)
Ideological imprint (H1)	-.005* (.002)	-.017** (.005)	-.011** (.004)	-.022 (.016)	-.052** (.017)	-.013** (.005)	-.643** (.047)
Political involvement	-.007** (.002)	-.008** (.002)	-.007** (.003)	-.007** (.003)	-.007** (.002)	-.006* (.003)	.005+ (.003)
Government appropriation	.014 (.009)	.018+ (.009)	.002 (.011)	.014 (.009)	.013 (.009)	.016+ (.009)	-.015 (.011)
Social network	-.036** (.003)	-.033** (.003)	-.036** (.003)	-.037** (.003)	-.036** (.003)	-.035** (.003)	-.002 (.004)
Post 2001	.035** (.006)	.032** (.006)	.035** (.006)	.035** (.006)	.033** (.006)	.035** (.006)	-.002 (.006)
Regional FDI intensity	.019** (.004)	.019** (.004)	.019** (.004)	.019** (.004)	.019** (.004)	.015** (.004)	-.002 (.004)
Ideological imprint × Political involvement (H2a)		.021** (.007)					.166** (.012)
Ideological imprint × Government appropriation (H2b)			.045* (.021)				.311** (.028)
Ideological imprint × Social network (H3)				.018 (.016)			.209** (.021)
Ideological imprint × Post 2001 (H4a)					.048** (.017)		.227** (.021)
Ideological imprint × Regional FDI intensity (H4b)						.014* (.007)	.114** (.010)
Number of observations	2,342	2,342	2,342	2,342	2,342	2,342	2,342
Goodness of fit (chi-squared)	2,426.15	2,443.66	2,434.62	2,427.50	2,441.72	2,433.55	2,836.13

+ $p < .10$; * $p < .05$; ** $p < .01$; intercept not reported to save space.

A Counterfactual Analysis for Understanding Effects of Profitability from Internationalization (1993–2012)

To better triangulate the imprinting processes we theorize, we also examined whether entrepreneurs with a communist ideological imprint tend to ignore profitable business opportunities from internationalization, showing whether the imprint dominates self-interest in considering foreign cooperation, i.e., whether the information filter by the communist ideological imprint motivates cognition. We measured profitable internationalization opportunities by calculating profitability differentials between internationalized firms and their non-internationalized/domestic counterparts of focal firms' institutional equivalents, i.e., those in the same geographical location (province) and industry as their counterfactuals, respectively (Marquis and Tilcsik, 2016). Then we interacted these two proxies of information of profitable internationalization with the communist ideological imprint to predict the tendency (hazard rate) of internationalization of the focal firm with a Cox proportional hazard model.

Table B2 shows that the interaction terms of all three variables with ideological imprint are negative and significant ($p < .05$), suggesting that firm leaders filtered information about profitable opportunities via internationalization with a communist ideological imprint. These results lend support to our theorizing that the communist ideological imprint acts as an information filter that motivates cognition—entrepreneurs eschew cooperation with foreign capitalists even when it is in their economic self-interest.

Table B2. Cox Proportional Hazard Model for Inward Internationalization with Counterfactual Profits (1993–2012)*

	1	2	3	4
	Single event model		Repeated events model	
With PSM	No	Yes	No	Yes
Ideological imprint	–128.647* (56.105)	–122.496* (53.377)	–120.455* (49.376)	–134.744** (46.859)
Profit premium by geography	–2.326 (11.521)	–6.958 (8.910)	–4.509 (10.447)	–7.466 (8.532)
Profit premium by industry	–5.519 (11.656)	–9.589 (9.105)	–9.648 (10.562)	–11.349 (8.647)
Profit premium by geography × Profit premium by industry	–5.676 (11.646)	–10.354 (8.970)	–8.189 (10.556)	–10.967 (8.595)
Ideological imprint × Profit premium by geography	–128.264* (58.216)	–121.319* (55.240)	–118.871* (51.368)	–132.605** (48.632)
Ideological imprint × Profit premium by industry	–125.707* (56.749)	–118.955* (53.981)	–117.696* (49.939)	–131.864** (47.380)
Ideological imprint × Profit premium by geography × Profit premium by industry	–125.118* (58.904)	–117.463* (55.883)	–115.850* (51.971)	–129.384** (49.190)
Number of observations	15,936	18,427	19,730	22,969
Goodness of fit (chi-squared)	687.46	732.23	808.36	831.83

⁺ $p < .10$; ^{*} $p < .05$; ^{**} $p < .01$; intercept not reported to save space.

* The sample size for the single event history analysis varies because the right-censored firms are dropped whenever they internationalized their firms, while propensity score matching also affects the number of observations used. The repeated events model retains all observations even if the focal firm/individual has already adopted the action, and therefore numbers of observations are similar to those in tables A3a and A3b for corresponding dependent variables (but still vary due to missing values). We winsorized top and bottom 1% profit premiums to avoid undue influence from outliers and considered mainly inward internationalization because it may affect its outward counterpart. Controls and moderators are the same as table A3a, respectively (not reported); interaction terms are excluded.

Diff-in-diff and Related Estimation of Internationalization

We provide a finer-grained analysis of communist ideological imprinting and also rule out cohort effects as an alternative explanation—it is not only those born before 1959 (and thus were already 18 years old and qualified to join the CPC) who are antagonistic to foreign cooperation but it is the importance of the socialization of joining the CPC—based on a diff-in-diff analysis. We distinguished those born before and after 1959, in addition to communist ideology. Therefore we analyzed four groups of entrepreneurs: pre- and post-reform communists and non-communists. Unlike traditional diff-in-diff estimation, our diff-in-diff analysis is not about the event of imprinting but about whether the entrepreneur had the ideological imprint (treatment) or not (control), and the “event” was whether opening up in 1978 invalidates the negative impact of current ideology on internationalization such that the post-reform government encourages capital and foreign cooperation more fully and that entrepreneurs who joined the CPC after 1978 were less antagonistic to internationalizing their ventures. Our approach resembles the diff-in-diff estimation for repeated cross sections, which utilizes between-differences of individuals rather than within ones, i.e., different individuals before and after the events (see Abadie, 2005: 9, for a summary of exemplar studies).

We found that pre-reform communists are least prone to internationalization; we present results in Panel A of table B3. We also combined the diff-in-diff analysis with PSM, as going through the imprinting process might be self-selected and thus endogeneity could be an issue. The PSM approach helps generate a random sample in terms of entrepreneurs with or without a communist ideological imprint based on observable variables. The results are shown in Panel B of table B3 and are consistent with the main analyses. Other birth year cutoffs indicating entrepreneurs who were 18 years old already in 1978—from birth in 1949, when the communist regime in China was established, to that in 1958—yielded similar results.

Table B3. Diff-in-diff and Related Estimation of Internationalization*

Panel A: Diff-in-diff estimation of internationalization		
Variable	Inward internationalization (0/1)	Outward internationalization (0/1)
Column	1	2
CPC membership	-.211 [*] (.100)	-.296 ⁺ (.179)
Born before 1959	-.213 ^{**} (.048)	-.208 ^{**} (.056)
CPC membership × Born before 1959	-.736 ^{**} (.050)	-.447 ^{**} (.056)
Number of observations	19,730	16,026
Goodness of fit (chi-squared)	1682.98	1042.99
Panel B: Diff-in-diff estimation of internationalization combined with propensity score matching		
Variable	Inward internationalization (0/1)	Outward internationalization (0/1)
CPC membership	-.245 [*] (.096)	-.291 ⁺ (.160)
Born before 1959	-.158 ^{**} (.045)	-.201 ^{**} (.052)
CPC membership × Born before 1959	-.772 ^{**} (.046)	-.469 ^{**} (.051)
Number of observations	17,934	14,604
Goodness of fit (chi-squared)	1,957.23	1,234.71

⁺ $p < .10$; ^{*} $p < .05$; ^{**} $p < .01$; intercept not reported to save space.
^{*} Numbers of observations for the analyses vary because of missing values and matching. Controls and moderators are the same as tables A3a and A3b, respectively (not reported); interaction terms are excluded.

Other Endogeneity Concerns and Moderators Unrelated to Information Filtering

We tried a number of additional approaches to address a variety of endogeneity concerns. For instance, we use the variation of entrepreneurs' parents' government work experience, which highly correlates with CPC membership (Bian, Shu, and Logan, 2001), as an instrumental variable. Parents' (communist) government work experience is exogenous and not affected by entrepreneurs' later new ventures. Studies have shown intergenerational transmissions of ideology (Jost, Federico, and Napier, 2009), and CPC membership is required for governmental work (Wu and Treiman, 2007). Lastly, parents' work experience is at best weakly correlated with firms' internationalization, given that nearly all entrepreneurs' parents—assuming they gave birth to their children in their twenties—are in their late eighties. Therefore the instrumental variable is valid conceptually. We still ran a first-stage F-test and obtained supporting evidence that our instrumental variable is valid. Results are reported in column 1 of table B4 and are similar to our main results.

Endogeneity issues may also arise because of omitted control variables. Therefore we also controlled for (1) state-owned-enterprise work experience and political rank to indicate whether the focal entrepreneur ever worked as a government official, and (2) the exclusion restrictions, i.e., technological resource, firm performance, industry competitiveness, and growth (Villalonga and McGahan, 2005). We report results in columns 2 to 7 of table B4, which are similar in terms of sign, significance, and magnitude on the communist ideological imprint.

In addition, we interacted firm age, entrepreneurs' age, and time trend with their ideological imprint. If the interaction terms are positively significant and the magnitudes are commensurate to interactions of our moderators, then our arguments that the availability and credibility of contradictory information are two key conditions for imprint decay would be undermined; the imprint decays automatically over time. The last six models (models 8 to 13) of both panels in table B4 suggest this was not the case; the effect sizes of the interaction terms are less than 10 percent of other moderating effects. Therefore the imprinting effect does not decay with an increase in firm age, entrepreneurs' age, or time but remains persistent.

Table B4. Results from Heckman Model of Internationalization (1993–2012): Other Endogeneity Concerns, Additional Control Variables, and Interactive Effects*

	1 (B) [†]	2 (B)	3 (C)	4 (B)	5 (C)	6 (C)	7 (C)	8 (C)	9 (C)	10 (C)	11 (C)	12 (C)	13 (C)
Approach	Additional control variables					New moderators: Interactive effects[§]							
Internationalization	Inward[‡]	Outward		Inward	Outward	Firm age		Founder's age		Time trend			
Ideological imprint (H1)	−3.865 [*] (1.638)	−.165 ^{**} (.027)	−.006 ^{**} (.000)	−.272 ^{**} (.029)	−.003 (.002)	−.006 ^{**} (.000)	−.005 [*] (.002)	−.010 ^{**} (.001)	−.014 ^{**} (.005)	−.005 ^{**} (.002)	−.041 ^{**} (.010)	−1.631 ^{**} (.280)	−15.501 ^{**} (1.771)
State-owned-enterprise work experience		.047 (.223)	.004 ⁺ (.003)	−.521 ⁺ (.275)	.007 (.020)								
Political rank as a government official		−.141 ^{**} (.031)	−.001 ^{**} (.000)	−.147 ^{**} (.034)	.011 ^{**} (.002)								
Tech resources						.009 ^{**} (.000)	−.003 (.004)						
Firm performance						.017 ^{**} (.001)	.018 ^{**} (.005)						
Industry competitiveness						−.047 ^{**} (.001)	−.153 ^{**} (.014)						
Industry growth						−.003 (.003)	.043 ⁺ (.026)						
Firm age								−.002 ^{**} (.000)	−.002 ^{**} (.000)				
Ideological imprint × Firm age								.001 ^{**} (.000)	.001 [*] (.001)				
Entrepreneur's age										.001 ^{**} (.000)	−.000 ^{**} (.000)		
Ideological imprint × Entrepreneur's age										−.000 (.000)	.001 ^{**} (.000)		
Time trend												−.000 (.000)	−.002 ^{**} (.001)
Ideological imprint × Time trend												.001 ^{**} (.000)	.008 ^{**} (.001)
Number of observations	17,926	17,926	2,785	14,598	2,178	2,785	2,178	2,785	2,178	2,785	2,178	2,785	2,178
Goodness of fit (chi-squared)	1,681.57	1,727.22	7,872.98	1,495.25	2,031.47	10,064.12	2,139.42	7,885.84	2,007.20	7,853.20	2,016.49	7,886.89	2,078.09

⁺ $p < .10$; ^{*} $p < .05$; ^{**} $p < .01$; intercept not reported to save space.

* The number of observations changes due to data availability.

[†] B: binary measure of dependent variable, C: continuous measure of dependent variable.

[‡] Instrumental variable is parent's government work experience. First-stage F test statistic of 8.10—above the critical value of 20% instrumental variable bias (6.66)—thus passed the test.

[§] We considered inward internationalization mainly.

Effects of Heterogeneity of Communist Ideological Imprint on Internationalization

The communist ideological imprint may exhibit some heterogeneity, the substantial magnitude of which may threaten the validity of our moderators as decaying factors of the communist ideological imprint. To address this issue, we considered regional variation—coastal region, geographic proximity to special economic zones, and local communist density in 1956 (Liu, Buck, and Shu, 2005; Luo, Xue, and Han, 2010; Kung and Chen, 2011)—that may also influence the intensity of the imprinting effect. Likewise, age groups of entrepreneurs may also affect the intensity of communist ideological indoctrination. During the Cultural Revolution (1966–1976), communist ideology was strengthened (Lu, 2004; Wang, Du and Marquis, 2018). Therefore the group born and brought up during the Cultural Revolution might exhibit a stronger imprinting effect and thus more heightened information filtering.

The results are reported in table B5, showing that most of these effects do not change the magnitude or significance of the communist ideological imprint. Furthermore, all five contingencies are unstable: they are either insignificant or contradict each other on measures of internationalization. Specifically, some significant results suggest that locating in more ideologically westernized areas reduces the negative effect of the ideological imprint, consistent with our results showing the moderating effect by regional FDI intensity. But the magnitudes are on the average less than 10% of our theorized moderating effects. The last three contingencies indicate the effects of communist ideology at the regional level during the imprinting period or birth, and their effects are unstable—both positive and negative coefficients are found—and the results are contrary to the conjecture that Cultural Revolution experience strengthens ideological indoctrination and thus heightens the information-filtering mechanism on outward internationalization. The results suggest that the heterogeneity of the ideological imprint does not pose a substantive threat to our analysis.

Table B5. Results from Heckman Model of Internationalization Considering Heterogeneity of Communist Ideological Imprint (1993–2012)*

Model	1 (B) [†]	2 (C)	3 (B)	4 (C)	5 (B)	6 (C)	7 (B)	8 (C)	9 (B)	10 (C)
Contingency	Coastal region		Contiguous to or contain special economic zone		CPC density in 1956 (before 1978)		Cultural Revolution (1966–1976)		Cultural Revolution until Nixon visited China (1966–1972)	
Panel A: Dependent variable—inward internationalization										
Controls and other moderators are the same as table A3a and not reported and other interaction terms are excluded										
Ideological imprint (H1)	–.267** (.039)	–.011** (.001)	–.181** (.029)	–.007** (.000)	–.158** (.039)	–.003** (.000)	–.169** (.029)	–.005** (.000)	–.171** (.029)	–.005** (.000)
Contingency	–.002 (.037)	–.003** (.000)	.033 (.043)	–.002** (.001)	–.000 (.003)	.000** (.000)	–.427** (.038)	–.004** (.001)	–.350** (.039)	–.003** (.001)
Ideological imprint × Contingency	.166** (.049)	.008** (.001)	.087 (.069)	.007** (.001)	.003 (.005)	–.000** (.000)	–.062 (.065)	–.004** (.001)	–.035 (.069)	–.003** (.001)
Number of observations	17,926	2,785	17,926	2,785	16,130	2,541	17,926	2,785	17,926	2,785
Goodness of fit (chi-squared)	1,716.41	8,043.48	1,709.80	7,918.17	1,533.19	7,331.02	1,832.78	7,989.48	1,786.81	7,954.30
Panel B: Dependent variable—outward internationalization										
Controls and other moderators are the same as table A3b and other interaction terms are excluded										
Ideological imprint (H1)	–.348** (.032)	–.003 (.002)	–.574** (.047)	–.009** (.003)	–.132** (.042)	–.005* (.003)	–.337** (.032)	–.005** (.002)	–.334** (.032)	–.005** (.002)
Contingency	–.099* (.048)	.006* (.003)	–.146** (.042)	–.004 (.003)	.018** (.003)	–.001** (.000)	–.430** (.041)	–.048** (.003)	–.357** (.042)	–.042** (.003)
Ideological imprint × Contingency	.518** (.075)	–.003 (.004)	.477** (.056)	.008* (.004)	–.024** (.005)	.000 (.000)	.318** (.067)	.031** (.004)	.364** (.071)	.030** (.004)
Number of observations	14,598	2,178	14,598	2,178	13,496	2,018	14,598	2,178	14,598	2,178
Goodness of fit (chi-squared)	1,511.37	2,005.24	1,526.76	2,006.44	1,371.50	1,845.44	1,557.86	2,304.06	1,526.86	2,216.11

⁺ $p < .10$; * $p < .05$; ** $p < .01$; intercept not reported to save space.
* The number of observations varies because of missing values of certain variables.
[†] B: binary measure of dependent variable, C: continuous measure of dependent variable.

Pseudo-panel Data Analysis

While we were unable to trace all firms in the sample consistently across years, the whole sample can be analyzed as a pseudo-panel dataset because individual firms in each cohort were randomly chosen within geographies. We thus employed pseudo-panel data estimation to investigate the hypotheses with our whole sample. Using this econometric approach, we can still make inferences about individual entrepreneurs (Deaton, 1985; Baltagi, 2008). At the same time, this approach further avoids possible bias resulting from attrition, i.e., untraced private firms that would be dropped if we used genuine panel data.

In pseudo-panel analyses, individual values are aggregated to statistically representative cohorts, and true panel data analyses are conducted on the cohort averages (Inoue, 2008). We defined cohorts according to the six-digit postal code per the original sampling process. According to Baltagi (2008), consistent estimates can be obtained by pseudo-panel data if one uses true panel data methods—random effects or fixed effects at the cohort level—by averaging the individual characteristics, i.e., collapsing the data at the cohort level. More importantly, our inference of an individual firm’s strategy can still be drawn using the pseudo-panel dataset estimation technique (Deaton, 1985). We conducted (1) the Breusch and Pagan test, which indicated that panel data approaches were more appropriate (i.e., fixed and random effects estimations are preferred to pooled OLS; Breusch and Pagan, 1979); and (2) the Hausman test, which indicated that unobservable fixed effects were not correlated with variables on the right-hand side of the model and thus a random effects model was most appropriate (Hausman, 1978). Therefore we reported results based on the random effects model described above.

Table B6a and B6b display the results, which are largely consistent with our reported results using the genuine panel data. This is consistent with econometric research on pseudo-panel models, which shows that the pseudo-panel data analysis produces similar results as those using genuine panel data and related approaches (Deaton, 1985; Baltagi, 2008; Inoue, 2008).

Table B6a. Results from Pseudo-panel Estimation: Inward Internationalization (1993–2012)

Variable	1	2	3	4	5	6	7
Panel A: Likelihood of internationalization with random effects linear model; controls are the same as Panel A of table A3a and not reported							
Ideological imprint (H1)	-.029** (.010)	-.077** (.013)	-.077** (.013)	-.137** (.020)	-.115** (.019)	-.057** (.016)	-.513** (.036)
Political involvement	.033** (.008)	.009 (.009)	.033** (.008)	.039** (.008)	.031** (.008)	.033** (.008)	-.004 (.009)
Government appropriation	.299** (.038)	.297** (.038)	.165** (.044)	.287** (.038)	.306** (.038)	.300** (.038)	.143** (.044)
Social network	.060** (.008)	.066** (.008)	.058** (.008)	.035** (.009)	.058** (.008)	.060** (.008)	.031** (.009)
Post 2001	.032** (.012)	.035** (.012)	.027* (.012)	.031* (.012)	.010 (.013)	.032** (.012)	-.011 (.013)
Regional FDI intensity	.017+ (.009)	.018+ (.009)	.019* (.009)	.022* (.009)	.012 (.009)	.007 (.010)	.006 (.010)
Ideological imprint × Political involvement (H2a)		.098** (.016)					.169** (.017)
Ideological imprint × Government appropriation (H2b)			.482** (.082)				.538** (.082)
Ideological imprint × Social network (H3)				.129** (.021)			.180** (.022)
Ideological imprint × Post 2001 (H4a)					.114** (.022)		.217** (.023)
Ideological imprint × Regional FDI intensity (H4b)						.050* (.022)	.062** (.022)
Number of observations	7,809	7,809	7,809	7,809	7,809	7,809	7,809
Goodness of fit (chi-squared)	1,096.65	1,137.52	1,135.10	1,139.71	1,126.82	1,102.39	1,339.88
Panel B: Amount of internationalization with random effects linear model; controls are the same as Panel B of table A3a and not reported							
Ideological imprint (H1)	-.006* (.003)	-.021** (.004)	-.020** (.004)	-.039** (.006)	-.026** (.006)	-.013** (.005)	-.142** (.012)
Political involvement	.008** (.002)	.001 (.003)	.008** (.002)	.010** (.002)	.008** (.002)	.008** (.002)	-.003 (.003)
Government appropriation	.092** (.012)	.092** (.012)	.054** (.014)	.088** (.012)	.094** (.012)	.092** (.012)	.047** (.014)
Social network	.012** (.003)	.014** (.003)	.011** (.003)	.004 (.003)	.012** (.003)	.012** (.003)	.003 (.003)
Post 2001	.023** (.004)	.024** (.004)	.022** (.004)	.023** (.004)	.018** (.004)	.023** (.004)	.012** (.004)
Regional FDI intensity	.002 (.003)	.003 (.003)	.003 (.003)	.004 (.003)	.001 (.003)	.000 (.003)	.000 (.003)
Ideological imprint × Political involvement (H2a)		.031** (.005)					.050** (.005)
Ideological imprint × Government appropriation (H2b)			.138** (.026)				.152** (.026)
Ideological imprint × Social network (H3)				.040** (.007)			.054** (.007)
Ideological imprint × Post 2001 (H4a)					.027** (.007)		.057** (.007)
Ideological imprint × Regional FDI intensity (H4b)						.012+ (.007)	.015* (.007)
Number of observations	7,809	7,809	7,809	7,809	7,809	7,809	7,809
Goodness of fit (chi-squared)	677.24	714.54	707.39	715.82	692.94	680.36	864.53

+ $p < .10$; * $p < .05$; ** $p < .01$; intercept not reported to save space.

Table B6b. Results from Pseudo-panel Estimation: Outward Internationalization (2000–2012)

Variable	1	2	3	4	5	6	7
Panel A: Likelihood of internationalization with random effects linear model; controls are the same as Panel A of table A3b and not reported							
Ideological imprint (H1)	-.042** (.010)	-.144** (.013)	-.084** (.013)	-.144** (.020)	-.078** (.030)	-.124** (.017)	-.652** (.044)
Political involvement	.062** (.008)	.002 (.010)	.062** (.008)	.069** (.009)	.062** (.008)	.062** (.008)	-.006 (.009)
Government appropriation	.377** (.041)	.373** (.040)	.248** (.048)	.365** (.041)	.378** (.041)	.378** (.041)	.210** (.047)
Social network	.069** (.008)	.083** (.008)	.068** (.008)	.044** (.010)	.069** (.008)	.071** (.008)	.045** (.009)
Post 2001	.008 (.014)	.007 (.014)	.003 (.014)	.006 (.014)	.003 (.015)	.009 (.014)	-.028* (.014)
Regional FDI intensity	.019+ (.012)	.018* (.012)	.020* (.012)	.026* (.012)	.018 (.012)	-.017 (.013)	-.012 (.013)
Ideological imprint × Political involvement (H2a)		.226** (.017)					.292** (.018)
Ideological imprint × Government appropriation (H2b)			.442** (.086)				.530** (.085)
Ideological imprint × Social network (H3)				.124** (.021)			.194** (.021)
Ideological imprint × Post 2001 (H4a)					.041 (.032)		.209** (.032)
Ideological imprint × Regional FDI intensity (H4b)						.150** (.025)	.155** (.025)
Number of observations	6,768	6,768	6,768	6,768	6,768	6,768	6,768
Goodness of fit (chi-squared)	887.37	1,089.37	916.82	925.88	889.12	927.60	1,309.90
Panel B: Amount of internationalization with random effects linear model; controls are the same as Panel B of table A3b and not reported							
Ideological imprint (H1)	-.008* (.003)	-.039** (.004)	-.026** (.004)	-.050** (.007)	-.016* (.010)	-.036** (.006)	-.214** (.014)
Political involvement	.015** (.003)	-.003 (.003)	.015** (.003)	.018** (.003)	.015** (.003)	.015** (.003)	-.006* (.003)
Government appropriation	.114** (.014)	.114** (.013)	.060** (.016)	.110** (.014)	.115** (.014)	.115** (.014)	.048** (.016)
Social network	.018** (.003)	.022** (.003)	.017** (.003)	.007* (.003)	.018** (.003)	.019** (.003)	.008* (.003)
Post 2001	.016** (.005)	.016** (.005)	.014** (.005)	.015** (.005)	.015** (.005)	.017** (.005)	.004 (.005)
Regional FDI intensity	.001 (.004)	.001 (.004)	.002 (.004)	.004 (.004)	.001 (.004)	-.011* (.004)	-.009* (.004)
Ideological imprint × Political involvement (H2a)		.069** (.006)					.092** (.006)
Ideological imprint × Government appropriation (H2b)			.185** (.029)				.210** (.028)
Ideological imprint × Social network (H3)				.051** (.007)			.072** (.007)
Ideological imprint × Post 2001 (H4a)					.009 (.011)		.064** (.011)
Ideological imprint × Regional FDI intensity (H4b)						.052** (.008)	.053** (.008)
Number of observations	6,768	6,768	6,768	6,768	6,768	6,768	6,768
Goodness of fit (chi-squared)	600.39	761.23	647.24	657.25	601.16	642.66	1,014.55

+ $p < .10$; * $p < .05$; ** $p < .01$; intercept not reported to save space.

REFERENCES

Abadie, A.

2005 “Semiparametric difference-in-differences estimators.” *Review of Economic Studies*, 72: 1–19.

Baltagi, B.

2008 *Econometric Analysis of Panel Data*. Hoboken, NJ: Wiley.

Bian, Y., X. Shu, and J. R. Logan

2001 “Communist Party membership and regime dynamics in China.” *Social Forces*, 79: 805–841.

Breusch, T. S., and A. R. Pagan

1979 “A simple test for heteroscedasticity and random coefficient variation.” *Econometrica*, 47: 1287–1294.

Deaton, A.

1985 “Panel data from time series of cross-sections.” *Journal of Econometrics*, 30: 109–126.

Hausman, J. A.

1978 “Specification tests in econometrics.” *Econometrica*, 46: 1251–1271.

Inoue, A.

2008 “Efficient estimation and inference in linear pseudo-panel data models.” *Journal of Econometrics*, 142: 449–466.

Jost, J. T., C. M. Federico, and J. L. Napier

2009 “Political ideology: Its structure, functions, and elective affinities.” *Annual Review of Psychology*, 60: 307–337.

Kung, J. K.-s., and S. Chen

2011 “The tragedy of the nomenklatura: Career incentives and political radicalism during China's Great Leap famine.” *American Political Science Review*, 105: 27–45.

Liu, X., T. Buck, and C. Shu

2005 “Chinese economic development, the next stage: Outward FDI?” *International Business Review*, 14: 97–115.

Lu, X.

2004 *Rhetoric of the Chinese Cultural Revolution: The Impact on Chinese Thought, Culture, and Communication*. Columbia, SC: University of South Carolina Press.

Luo, Y., Q. Xue, and B. Han

2010 “How emerging market governments promote outward FDI: Experience from China.” *Journal of World Business*, 45: 68–79.

Marquis, C., and A. Tilcsik

2016 “Institutional equivalence: How industry and community peers influence corporate philanthropy.” *Organization Science*, 27: 1325–1341.

Villalonga, B., and A. M. McGahan

2005 “The choice among acquisitions, alliances, and divestitures.” *Strategic Management Journal*, 26: 1183–1208.

Wang, D., F. Du, and C. Marquis

2018 “Defending Mao’s dream: Politicians’ ideological imprinting and firms’ political appointment in China.” *Academy of Management Journal* (forthcoming), DOI: <https://doi.org/10.5465/amj.2016.1198>.

Wu, X., and D. J. Treiman

2007 “Inequality and equality under Chinese socialism: The hukou system and intergenerational occupational mobility.” *American Journal of Sociology*, 113: 415–445.

Appendix C: Clarifying Data and Analyses

Potential Sources of Error

Like many other surveys, the Chinese Private Enterprise Survey (CPES) data contain errors. We are unable to enumerate all sources of errors, but we discuss below that survey design, implementation, and careless responses or coding errors could lead to errors in the data (Lynn, 2009). Many other surveys that are implemented longitudinally, such as the U.S. Current Population Survey (CPS), have also been shown to exhibit these types of errors (e.g., Peracchi and Welch, 1995; Madrian and Lefgren, 2000; Feng, 2008; Kambourov and Manovskii, 2013). The organizers and providers of the CPES data have acknowledged these three sources of error in a paper summarizing the history of the survey (Chen et al., 2018).

Survey design. Chen et al. (2018) suggested that the “survey design was flawed . . . especially in early years” (p. 19) and “some questionnaire items were changed over time” (p. 25). We also found that the survey questionnaires contained some ambiguous and inconsistent questions. For example, in 1995 and from 2004 to 2010, respondents were asked to report their *main* career experience—the type of their employment and organization, their job title, etc.—while from 1997 to 2002, they were asked to report *all* their job experiences, although the section name was still “main experience.” And for each question item, the framing of the question (main job versus experience) and values for these items indicating types of employer also changed; some years distinguished collective-owned firms based on urban and rural residence and/or included the military.

Data collection. Survey implementation is another potential source of error according to Chen et al. (2018: 19), particularly in the early years. Reporting errors are

common in surveys, and data-collecting agencies might not check the consistency of these responses across different waves (e.g., CPS; Peracchi and Welch, 1995; Madrian and Lefgren, 2000). Furthermore, wrong and/or different persons might have filled in the questionnaire over time, i.e., the survey team might have asked different co-owners, managers, or the current owners who were not the original entrepreneurs due to leadership changes and other ownership arrangements. Therefore, entrepreneurs' characteristics could suffer from error, e.g., within the same firm, reporting errors and potential change of surveyed individuals can make time-invariant characteristics change over time. Surveyors' handbooks, questionnaires in later years that aimed to correct this problem, and our private communications with data providers all indicated that this was an issue.

Careless response or coding error. Several indicators suggest the necessity of treating some values as careless responses or a coding error. We found that the data suffered from two issues—negative values where there should not be and erroneously large values—and these outliers are likely the result of careless responses or coding errors (e.g., Meade and Craig, 2012; Aguinis, Gottfredson, and Joo, 2013).

Data Cleaning

We took a number of measures to address these issues. For the potential errors resulting from survey design, e.g., entrepreneurs' characteristics, because we are unable to know in which wave the correct person answered the questions, we left those values as they were. Furthermore, we did not use variables suffering these potential errors, including founding year, for matching. But we performed sensitivity tests detailed below that suggest these errors are noise and do not affect the results. We also treated the careless responses in ways appropriate to the specific variable. We regarded negative values that are out of the meaningful range—e.g., initial investments,

taxes, fees, and initial industry category—as missing. For binary variables such as membership in the communist party, association of private entrepreneurs, and government work experience, we considered negative values as suggesting “no” in generating corresponding variables. Negative values for some continuous variables, such as profits, could be meaningful. We thus treated negative values for continuous variables on a case-by-case basis. For instance, we maintained negative values of profit measures, while for government appropriations, since some entrepreneurs understood this question as how much was subtracted from their profits—and so they reported negative values—we used absolute values. For variables that were coded with impossibly large values—e.g., over 9 (Ph.D. level) for educational attainment and magnitude over 10^{15} —we replaced them with the largest possible value. Our results are robust to whether we cleaned these variables or not.

Sensitivity Analyses

In addition to mitigating errors in variables during cleaning and computing processes, we also ran regressions (1) with subsamples whose values of the variables are less subject to errors and (2) with or without variables containing potential errors (Aguinis, Gottfredson, and Joo, 2013) to ensure that these errors did not affect our results. In particular, because entrepreneurs’ characteristics contain errors, it is especially important to determine whether these errors affected our key variable of ideological imprint. (3) We further show that the results are insensitive to matching, as the potential data errors could result in false matches.

We first checked the robustness of our results by keeping observations with consistent values of ideological imprint, which we defined in different ways: (1) Firms that never experienced change in this variable, which was the case for 29.74 percent of the enterprises in the data, and those with only a one-time deviation, e.g., all ones but one

zero and vice versa (also excluding firms with two years of observations with a one and zero). Firms with no more than a one-time deviation are less subject to potential matching errors, e.g., for all nine observations only one is different from others. Using this definition, more than 68% of the 19,729 observations remained. (2) Observations whose value of ideological imprint equals the sample period mode; e.g., if a firm has three ones and two zeroes, the mode is one and we dropped the observations with zeroes. Values equal to the mode are likely less subject to errors. Using this definition, 75.8% of observations remained. (3) Firms with a supermajority (no less than two-thirds) of zeroes or ones, as values in a supermajority are more likely the consistent value. Using this definition, 81.3% of observations remained. These ratios are consistent with the changing values of time-invariant variables or inconsistencies of deterministic (age) variables in other well-known longitudinal surveys, e.g., see Black, Sanders, and Taylor (2003: 547–549) on the CPS and Kambourov and Manovskii (2013: 175, 179) on both the CPS and Panel Studies of Income Dynamics (PSID). More importantly, results from all these subsample analyses (available upon request) were similar to those reported in our article and supportive of our overall conclusions. Given the robustness of our results, we conclude that we can treat these errors as noise (Aguinis, Gottfredson, and Joo, 2013). We did not use any of these subsamples for the main analyses because sample attrition from dropping cases may lead to bias, and measurement errors in surveys do not necessarily bias the results (Peracchi and Welch, 1995; Lynn, 2009).

Second, we altered the model specifications by dropping control variables that could be subject to high errors because of survey implementation, e.g., entrepreneur's characteristics. We controlled for them in the main analysis following existing imprinting studies (Azoulay, Liu, and Stuart, 2017), but we then also excluded these variables and obtained results (available upon

request) that were similar to those reported and supported our hypotheses.

Third, we used the full sample without matching to show that our results are insensitive to potential matching errors. Direct cross-sectional estimation with this sample may suffer from biases due to individual heterogeneity. However, recent econometrics literature suggests that PSM can deal with individual heterogeneity effectively (Arkhangelsky and Imbens, 2018), and thus we combined cross-sectional analyses with PSM and report the results in table C1 (those without using PSM are available upon request). The results still support our conclusions, although support for hypothesis 4b is weaker. This is not surprising because imprinting mainly focuses on how cross-sectional variation in the past affects future behaviors. These results, combined with those from pseudo panel estimation that does not require matching (in Online Appendix B), support that our results hold regardless of any potential matching errors.

Table C1. Results from Heckman Model and Cross-sectional Analyses with PSM (1993–2012)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Dependent variable: Inward internationalization								Dependent variable: Outward internationalization						
Panel A: First-stage results of Heckman (random effects probit); controls are the same as Panel A of tables A3a and A3b and not reported														
Ideological imprint (H1)	-0.219** (0.027)	-0.351** (0.036)	-0.461** (0.043)	-1.122** (0.098)	-0.695** (0.099)	-0.195** (0.049)	-2.578** (0.173)	-0.310** (0.030)	-0.802** (0.045)	-0.593** (0.047)	-1.452** (0.122)	-0.559** (0.186)	-0.218** (0.058)	-3.695** (0.253)
Political involvement	0.307** (0.031)	0.169** (0.040)	0.304** (0.031)	0.320** (0.031)	0.301** (0.031)	0.307** (0.031)	0.063 (0.040)	0.749** (0.036)	0.336** (0.044)	0.752** (0.036)	0.764** (0.036)	0.749** (0.036)	0.749** (0.036)	0.233** (0.044)
Government appropriation	1.649** (0.136)	1.647** (0.136)	0.793** (0.178)	1.625** (0.136)	1.654** (0.136)	1.650** (0.136)	0.757** (0.177)	1.583** (0.149)	1.600** (0.151)	0.599** (0.194)	1.549** (0.150)	1.585** (0.149)	1.585** (0.149)	0.432** (0.193)
Social network	0.554** (0.045)	0.558** (0.045)	0.558** (0.045)	0.239** (0.053)	0.551** (0.045)	0.554** (0.045)	0.204** (0.053)	0.603** (0.048)	0.615** (0.048)	0.611** (0.048)	0.259** (0.056)	0.603** (0.048)	0.600** (0.048)	0.171** (0.056)
Post 2001	0.677** (0.111)	0.667** (0.111)	0.697** (0.111)	0.698** (0.111)	0.543** (0.113)	0.677** (0.111)	0.481** (0.113)	0.241+ (0.140)	0.241+ (0.142)	0.265+ (0.140)	0.248** (0.140)	0.187 (0.145)	0.241+ (0.140)	0.045 (0.145)
Regional FDI intensity	0.137** (0.047)	0.139** (0.047)	0.142** (0.047)	0.171** (0.048)	0.131** (0.048)	0.155** (0.057)	0.178** (0.056)	0.223** (0.070)	0.220** (0.071)	0.240** (0.070)	0.263** (0.072)	0.223** (0.070)	0.296** (0.080)	0.368** (0.080)
Ideological imprint × Political involvement (H2a)		0.309** (0.055)					0.593** (0.059)		1.003** (0.063)					1.389** (0.068)
Ideological imprint × Government appropriation (H2b)			1.972** (0.266)				2.111** (0.275)			2.322** (0.291)				3.050** (0.316)
Ideological imprint × Social network (H3)				1.011** (0.102)			1.181** (0.107)				1.274** (0.126)			1.645** (0.132)
Ideological imprint × Post 2001 (H4a)					0.521** (0.104)		0.869** (0.109)					0.257 (0.190)		1.025** (0.193)
Ideological imprint × Regional FDI intensity (H4b)						-0.045 (0.077)	-0.008 (0.080)						-0.174** (0.094)	-0.247** (0.102)
Number of observations	10,986	10,986	10,986	10,986	10,986	10,986	10,986	9,669	9,669	9,669	9,669	9,669	9,669	9,669
Panel B: Second-stage results of Heckman (random effects linear); controls are the same as Panel B of tables A3a and A3b and not reported														
Inverse Mills ratio	-0.456** (0.010)	-0.420** (0.012)	-0.443** (0.011)	-0.433** (0.011)	-0.428** (0.010)	-0.456** (0.010)	-0.052** (0.017)	-0.484** (0.018)	-0.482** (0.031)	-0.480** (0.020)	-0.469** (0.019)	-0.477** (0.018)	-0.487** (0.018)	-0.292** (0.056)
Ideological imprint (H1)	-0.008** (0.001)	-0.012** (0.001)	-0.011** (0.001)	-0.034** (0.004)	-0.066** (0.004)	-0.007** (0.001)	-0.245** (0.008)	-0.023** (0.002)	-0.023** (0.005)	-0.024** (0.004)	-0.067** (0.014)	-0.090** (0.018)	-0.032** (0.004)	-0.224** (0.033)
Political involvement	-0.005** (0.001)	-0.007** (0.001)	-0.004** (0.001)	-0.004** (0.001)	-0.003** (0.001)	-0.005** (0.001)	-0.002** (0.001)	-0.009** (0.003)	-0.009** (0.004)	-0.009** (0.004)	-0.008** (0.003)	-0.008** (0.003)	-0.010** (0.003)	-0.002 (0.004)
Government appropriation	-0.028** (0.005)	-0.018** (0.005)	-0.035** (0.005)	-0.022** (0.005)	-0.020** (0.005)	-0.028** (0.005)	0.018** (0.005)	0.009 (0.012)	0.009 (0.014)	0.006 (0.015)	0.013 (0.012)	0.009 (0.012)	0.008 (0.012)	0.021 (0.015)
Social network	-0.010** (0.002)	-0.008** (0.002)	-0.009** (0.002)	-0.012** (0.002)	-0.008** (0.001)	-0.010** (0.002)	-0.001 (0.001)	-0.027** (0.004)	-0.027** (0.004)	-0.027** (0.004)	-0.030** (0.004)	-0.027** (0.004)	-0.026** (0.004)	-0.025** (0.004)
Post 2001	0.002 (0.003)	0.002 (0.002)	0.003 (0.003)	0.003 (0.002)	-0.006** (0.002)	0.002 (0.003)	-0.001 (0.002)	0.028** (0.009)	0.028** (0.009)	0.028** (0.009)	0.027** (0.009)	0.020** (0.009)	0.028** (0.009)	0.018** (0.009)
Regional FDI intensity	-0.015** (0.001)	-0.014** (0.001)	-0.014** (0.002)	-0.013** (0.001)	-0.014** (0.001)	-0.014** (0.002)	-0.002 (0.002)	-0.039** (0.005)	-0.039** (0.005)	-0.039** (0.005)	-0.038** (0.005)	-0.039** (0.005)	-0.046** (0.006)	-0.034** (0.007)
Ideological imprint × Political involvement (H2a)		0.010** (0.002)					0.052** (0.002)		0.001 (0.008)					0.042** (0.013)
Ideological imprint × Government appropriation (H2b)			0.023** (0.008)				0.168** (0.009)			0.011 (0.024)				0.091** (0.035)
Ideological imprint × Social network (H3)				0.027** (0.004)			0.090** (0.004)				0.047** (0.014)			0.077** (0.016)
Ideological imprint × Post 2001 (H4a)					0.061** (0.004)		0.108** (0.004)					0.069** (0.018)		0.092** (0.019)
Ideological imprint × Regional FDI intensity (H4b)						-0.001 (0.002)	-0.002 (0.002)						0.018** (0.007)	0.010 (0.007)
Number of observations	2,594	2,594	2,594	2,594	2,594	2,594	2,594	2,041	2,041	2,041	2,041	2,041	2,041	2,041
+ p < .10; * p < .05; ** p < .01; intercept not reported to save space.														

* $p < .10$; ** $p < .05$; *** $p < .01$; intercept not reported to save space.

Variants of PSM in Panel Settings

PSM deals with endogeneity issues by generating weights for regression, dropping observations not on the common support (assigning a zero or missing weight). In panel settings, STATA uses a cross-sectional logistic or Probit regression (`psmatch2`), which treats observations of the same firm in different periods as independent of each other, and thus there could be variations of weights for the same firm over the sample period. However, panel data approaches require a constant weight over the sample period for the same firm, and there is no consensus for PSM implementation in the panel setting as to how we should weight different observations. STATA's default command to generate weights ("`bys firm_id: egen`") uses the mean value of the weight over the sample period for each firm. For example, if in years 1 and 3 a firm's weight is 3 while that in year 2 is missing, then STATA assigns a mean weight of $3 = (3+3)/2$ to all observations of the firm, including the observation with missing value. But including firms for which the calculated weight is missing would include observations that should have been dropped if run in a cross-sectional setting. Thus, to be more conservative we ran additional analyses: (1) We dropped observations with missing weight first, assigned the mean value of weights to observations with non-missing weight for each firm, and then used panel data methods; (2) We used weights from the first step of PSM directly and then the cross-sectional method that ignores the longitudinal nature of the data. An issue with these two approaches is that they dropped more than 60% of observations, and a known pitfall of the PSM approach is that it can delete too many observations (Caliendo and Kopeinig, 2008). Results for alternative approach (1), which are reported in table C2, and for approach (2), which are available upon request, were consistent with those reported in the article and support our conclusions.

Table C2. Results from Heckman Model and the PSM Implementation by Dropping Observations with Missing Values First (1993–2012)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Dependent variable: Inward internationalization								Dependent variable: Outward internationalization						
Panel A: First-stage results of Heckman (random effects probit); controls are the same as Panel A of tables A3a and A3b and not reported														
Ideological imprint (H1)	-0.171** (0.032)	-0.375** (0.041)	-0.529** (0.058)	-1.202** (0.130)	-0.740** (0.103)	-0.247** (0.058)	-3.302** (0.210)	-0.370** (0.035)	-0.862** (0.047)	-0.658** (0.063)	-1.596** (0.158)	-0.682** (0.206)	-1.014** (0.068)	-5.344** (0.313)
Political involvement	0.401** (0.039)	0.097 (0.053)	0.396** (0.039)	0.416** (0.039)	0.405** (0.039)	0.400** (0.039)	0.008 (0.053)	0.987** (0.044)	0.378** (0.057)	0.988** (0.044)	1.002** (0.044)	0.986** (0.044)	0.999** (0.044)	0.296** (0.058)
Government appropriation	1.328** (0.168)	1.301** (0.168)	-0.044 (0.248)	1.325** (0.168)	1.355** (0.169)	1.327** (0.168)	-0.064 (0.243)	1.647** (0.179)	1.673** (0.183)	0.617* (0.258)	1.672** (0.180)	1.654** (0.179)	1.654** (0.181)	0.550* (0.254)
Social network	0.624** (0.067)	0.630** (0.067)	0.626** (0.067)	0.177+ (0.082)	0.610** (0.067)	0.625** (0.067)	0.063 (0.081)	0.840** (0.072)	0.840** (0.072)	0.847** (0.072)	0.405** (0.086)	0.839** (0.072)	0.847** (0.073)	0.230* (0.086)
Post 2001	0.701** (0.168)	0.757** (0.167)	0.723** (0.167)	0.704** (0.166)	0.428* (0.173)	0.696** (0.168)	0.319 (0.170)	0.014 (0.201)	0.076 (0.208)	0.032 (0.201)	0.010 (0.199)	-0.072 (0.207)	-0.003 (0.203)	-0.289 (0.204)
Regional FDI intensity	0.178* (0.060)	0.169* (0.060)	0.177* (0.060)	0.232** (0.060)	0.159* (0.061)	0.105 (0.076)	0.068 (0.074)	0.136 (0.077)	0.076 (0.078)	0.129 (0.077)	0.217* (0.078)	0.136 (0.077)	-0.543** (0.099)	-0.537** (0.100)
Ideological imprint × Political involvement (H2a)		0.558** (0.066)					0.820** (0.071)		1.226** (0.073)					1.562** (0.078)
Ideological imprint × Government appropriation (H2b)			2.523** (0.339)				2.705** (0.342)			1.991** (0.359)				2.702** (0.380)
Ideological imprint × Social network (H3)				1.114** (0.134)			1.417** (0.140)				1.313** (0.162)			1.933** (0.173)
Ideological imprint × Post 2001 (H4a)					0.639** (0.108)		1.105** (0.115)					0.322 (0.209)		1.430** (0.221)
Ideological imprint × Regional FDI intensity (H4b)						0.143 (0.090)	0.281* (0.093)						1.229** (0.110)	1.462** (0.117)
Number of observations	7,767	7,767	7,767	7,767	7,767	7,767	7,767	6,472	6,472	6,472	6,472	6,472	6,472	6,472
Panel B: Second-stage results of Heckman (random effects linear); controls are the same as Panel B of tables A3a and A3b and not reported														
Inverse Mills ratio	-0.107** (0.001)	-0.111** (0.001)	-0.105** (0.001)	-0.108** (0.001)	-0.103** (0.001)	-0.105** (0.001)	0.006+ (0.003)	-0.094** (0.003)	-0.083** (0.005)	-0.092** (0.003)	-0.094** (0.003)	-0.093** (0.003)	-0.097** (0.003)	0.047** (0.014)
Ideological imprint (H1)	-0.010** (0.000)	-0.008** (0.001)	-0.013** (0.001)	-0.009** (0.002)	-0.034** (0.002)	-0.018** (0.001)	-0.323** (0.007)	-0.002 (0.002)	-0.014* (0.005)	-0.010+ (0.004)	-0.008 (0.015)	-0.060** (0.018)	0.005 (0.005)	-0.629** (0.059)
Political involvement	-0.004** (0.001)	-0.003** (0.001)	-0.004** (0.001)	-0.004** (0.001)	-0.003** (0.001)	-0.004** (0.001)	-0.001 (0.000)	-0.020** (0.004)	-0.021** (0.004)	-0.019** (0.004)	-0.020** (0.004)	-0.019** (0.004)	-0.022** (0.004)	0.007 (0.004)
Government appropriation	0.012** (0.002)	0.009** (0.002)	0.002 (0.003)	0.012** (0.002)	0.017** (0.002)	0.014** (0.002)	0.000 (0.002)	-0.009 (0.011)	0.002 (0.012)	-0.030+ (0.015)	-0.008 (0.012)	-0.008 (0.011)	-0.012 (0.012)	0.032* (0.016)
Social network	-0.004** (0.001)	-0.005** (0.001)	-0.003* (0.001)	-0.004** (0.001)	-0.003* (0.001)	-0.003* (0.001)	0.001 (0.001)	-0.028** (0.005)	-0.025** (0.005)	-0.028** (0.005)	-0.029** (0.005)	-0.028** (0.005)	-0.029** (0.005)	-0.011+ (0.005)
Post 2001	0.011** (0.002)	0.010** (0.002)	0.011** (0.002)	0.011** (0.002)	0.005* (0.002)	0.010** (0.002)	0.002 (0.001)	0.064** (0.013)	0.061** (0.013)	0.064** (0.013)	0.064** (0.013)	0.059** (0.013)	0.063** (0.013)	0.029* (0.013)
Regional FDI intensity	0.004** (0.001)	0.004** (0.001)	0.005** (0.001)	0.004** (0.001)	0.004** (0.001)	-0.003** (0.001)	-0.001 (0.001)	0.035** (0.005)	0.035** (0.005)	0.035** (0.005)	0.035** (0.005)	0.036** (0.005)	0.040** (0.006)	-0.021+ (0.008)
Ideological imprint × Political involvement (H2a)		-0.005** (0.001)					0.060** (0.001)		0.021* (0.007)					0.164** (0.016)
Ideological imprint × Government appropriation (H2b)			0.022** (0.005)				0.230** (0.005)			0.051* (0.022)				0.248** (0.030)
Ideological imprint × Social network (H3)				-0.000 (0.002)			0.122** (0.003)				0.006 (0.015)			0.217** (0.025)
Ideological imprint × Post 2001 (H4a)					0.025** (0.002)		0.118** (0.002)					0.059** (0.018)		0.209** (0.023)
Ideological imprint × Regional FDI intensity (H4b)						0.016** (0.001)	0.039** (0.001)						-0.012 (0.008)	0.127** (0.015)
Number of observations	1,600	1,600	1,600	1,600	1,600	1,600	1,600	1,262	1,262	1,262	1,262	1,262	1,262	1,262
+ p < .10; * p < .05; ** p < .01; intercept not reported to save space.														

+ $p < .10$; * $p < .05$; ** $p < .01$; intercept not reported to save space.

REFERENCES

- Aguinis, H., R. K. Gottfredson, and H. Joo
2013 “Best-practice recommendations for defining, identifying, and handling outliers.” *Organizational Research Methods*, 16: 270–301.
- Arkhangelsky, D., and G. W. Imbens
2018 “The role of the propensity score in fixed effect models.” *National Bureau of Economic Research*.
- Azoulay, P., C. C. Liu, and T. E. Stuart
2017 “Social influence given (partially) deliberate matching: Career imprints in the creation of academic entrepreneurs.” *American Journal of Sociology*, 122: 1223–1271.
- Black, D., S. Sanders, and L. Taylor
2003 “Measurement of higher education in the census and current population survey.” *Journal of the American Statistical Association*, 98: 545–554.
- Caliendo, M., and S. Kopeinig
2008 “Some practical guidance for the implementation of propensity score matching.” *Journal of Economic Surveys*, 22: 31–72.
- Chen, G., P. Lu, Z. Lin, and N. Song
2018 “The 25 years of Chinese Private Enterprise Survey: Points and prospects.” *Nankai Business Review (in Chinese)*, 21: 17–27.
- Feng, S.
2008 “Longitudinal matching of recent current population surveys: Methods, non-matches and mismatches.” *Journal of Economic and Social Measurement*, 33: 241–252.
- Kambourov, G., and I. Manovskii
2013 “A cautionary note on using (March) Current Population Survey and Panel Study of Income Dynamics data to study worker mobility.” *Macroeconomic Dynamics*, 17: 172–194.
- Lynn, P.
2009 *Methodology of Longitudinal Surveys*. Hoboken, NJ: Wiley.
- Madrian, B. C., and L. J. Lefgren
2000 “An approach to longitudinally matching Current Population Survey (CPS) respondents.” *Journal of Economic and Social Measurement*, 26: 31–62.
- Meade, A. W., and S. B. Craig
2012 “Identifying careless responses in survey data.” *Psychological Methods*, 17: 437–455.
- Peracchi, F., and F. Welch
1995 “How representative are matched cross-sections? Evidence from the Current Population Survey.” *Journal of Econometrics*, 68: 153–179.