

Supplementary Materials

Specification, Heterogeneity and Functional Form in the Determinants of Small-Firm Profitability. Angelo Leogrande, Mauro di Molfetta, Valeria Notarnicola, Maria Giovanna Trotta and Antonio Volpe Plantamura.

This file contains the tables, figures and supporting analyses moved from the main text to keep the article concise. Tables and figures are numbered S1, S2 and so on in order of appearance. References to tables and figures without the prefix S, such as Table 4 or Figure 2, and to Appendices A and B refer to the main text.

Section S1. Studies Reviewed, by Theme

Table S1 lists the two hundred studies retrieved from Scopus, each assigned to one of eight themes. The themes are ordered as in Section 2, where they are grouped into five strands; Table 1 in the main text reports, for each strand, the assumption left untested and the test applied to it in this paper.

Table S1. Studies reviewed, by theme.

Theme (Section of the review; studies)	Studies
Capital structure and profitability in small firms (Section 2.1; n = 28)	Benkraiem (2016); Bhawna & Sahay (2025); Ceylan (2021); Das et al. (2022); Demiraj et al. (2025); Garcia-Martinez et al. (2023); Ghasemi & Ab Razak (2017); Gonçalves et al. (2024); Govindan et al. (2023); Gutiérrez-Ponce (2024); Hussain et al. (2024); Kalash (2023); Karim et al. (2021); Khan et al. (2025); S. Kumar & Rao (2016); S. Kumar & Verma (2025); Matias & Serrasqueiro (2017); Nkosi et al. (2026); Quoc Trung (2021); Rodríguez Valencia (2025); Rokhayati et al. (2019); Shchyrba et al. (2026); Soni et al. (2023); Tran Minh Hung et al. (2026); Vu Thi & Phung (2021); Yousaf (2025); Youssef et al. (2022); Youssef et al. (2023)
Cost structure, efficiency and the operating model (Section 2.2; n = 17)	Albaity et al. (2019); Alhassan & Ohene-Asare (2016); Bhattu-Babajee & Seetanah (2022); Campisi et al. (2019); Chadha et al. (2023); Chino (2016); Grau & Reig (2021); Lan et al. (2026); Li & Tong (2023); Maside-Sanfiz et al. (2024); Pant et al. (2026); Prakash & Nauriyal (2020); Prasad & Mondal (2025); Shyu (2013); Somya & Saripalle (2023); Spicka et al. (2017); Tripathi et al. (2024)
Endogeneity, simultaneity and identification (Section 2.3; n = 46)	Abdalla et al. (2025); Al-Khazaleh et al. (2025); Aliakbari et al. (2025); Amarhyouz & Azegagh (2025); Artica et al. (2019); Bolarinwa et al. (2026); Boshnak (2024); Canarella & Miller (2018); Dalci (2018); Damira et al. (2022); Dsouza et al. (2025); Ernesto Martínez Avella & Andrés Hernández Salazar (2026); Goh et al. (2018); Habib (2026); Hin et al. (2026); Jolly Cyril & Singla (2020); Le et al. (2025); Li et al. (2025); Lou et al. (2025); Mansour et al. (2023); Mantzari et al. (2024); Morais Francisco (2026); Neves et al. (2026); Nguyen & Huynh (2023); Nguyen & Nguyen (2026); Novaković et al. (2025); Nyeadi et al. (2018); Pham & Ngo (2025); Pham & Nguyen (2020); Rahman & Harun (2025); Rastogi & Kumar (2024); Saiz-Sepulveda et al. (2026); Samal & Yadav (2025); Sari et al. (2026); Shakri et al. (2025); Sigalas & Gerakoudi (2025); Singh et al. (2025); Sultana & Zeya (2026); Surjanto & Ariefianto (2025); Talamas Marcos (2025); Tran & Nguyen (2025); Vengesai & Kwenda

	(2018); Vieira (2017); Vijayakumaran & Vijayakumaran (2019); Waheed & Bagh (2025); Xie & Yuan (2025)
Profit persistence and dynamic adjustment (Section 2.3; n = 16)	Abdeljawad & Farhood (2025); Al Barakat & Ali (2026); Al-Homaidi et al. (2020); Choi et al. (2024); Doyran & Santamaria (2019); Hirsch & Hartmann (2014); Hirsch et al. (2021); Hoang et al. (2026); Jaisinghani (2016); Jouda (2018); Killins (2020); Opstad et al. (2022); Rios-Vazquez & Portela-Maseda (2026); Sinha & Vodwal (2022); Yadav et al. (2022); Yang et al. (2015)
Unsupervised classification and firm taxonomies (Section 2.4; n = 23)	Bai & Harith (2023); Barlatier et al. (2023); Bayaraa et al. (2019); Camuffo & Poletto (2024); Cantabene & Grassi (2024); Costa et al. (2017); Davcik & Grigoriou (2020); Godley et al. (2021); Juntunen et al. (2022); S. S. Kumar et al. (2024); Lee & Hallak (2020); Linares-Mustarós et al. (2018); Ljungkvist & Andersén (2021); Meena et al. (2023); Mladenova et al. (2025); Nazarova et al. (2021); Pagaddut (2021); Risberg et al. (2023); Salina et al. (2020); Salles-Filho et al. (2023); Thomä & Bizer (2013); Williams et al. (2025); Xiao et al. (2025)
Machine learning for distress, default and fraud prediction (Section 2.4; n = 18)	Antar & Tayachi (2025); Ariza-Garzón et al. (2024); Balzano & Magrini (2026); Bijoy et al. (2026); Cheraghali & Molnár (2026); Coronell et al. (2026); Gavurova et al. (2022); Martínez et al. (2021); Mousa et al. (2025); Papíková & Papík (2024); Phothong et al. (2026); Sehgal et al. (2021); Sehgal et al. (2023); Shetty et al. (2022); Sodnomdavaa & Lkhagvadorj (2026); Wei et al. (2024); Zhang et al. (2025); Zhu et al. (2022)
Machine learning applied to firm performance (Section 2.4; n = 32)	Abbate et al. (2023); Ali (2026); Anton et al. (2021); Banna & Alam (2025); Bednarek et al. (2025); Deepak Kumar et al. (2024); Feng et al. (2025); Forradellas et al. (2026); Gao (2022); Garg et al. (2025); Giudici et al. (2023); Gotti et al. (2025); Gupta et al. (2025); Houssaini & Miloud (2025); Kahlen et al. (2024); Kanyepe et al. (2025); Khayer et al. (2020); Kohtamäki et al. (2024); Kristóf & Virág (2022); Mahmood et al. (2025); Mikram et al. (2025); Neha et al. (2025); Pap et al. (2022); Phuensane et al. (2026); Sengar et al. (2026); Sultana et al. (2026); Sura & Ventura (2026); Vajjhala & Strang (2024); Vašaničová et al. (2025); Xu & Sam (2025); Yang et al. (2023); Zhan et al. (2022)
Innovative start-ups and certified SMEs (Section 2.5; n = 20)	Aiello et al. (2024); Al-Eitan et al. (2022); Amoa-Gyarteng & Dhliwayo (2023); Anderloni & Harasheh (2025); Angilella et al. (2023); Antenzio et al. (2026); Charity et al. (2019); Di Berardino & Antenzio (2026); Domma & Errico (2023); Kayani et al. (2025); Manaresi et al. (2021); Migliaccio & Pavone (2021); Minola et al. (2026); Scandurra et al. (2025); Schifilliti & La Rocca (2024); Tong & Serrasqueiro (2020); Tons & Serrasqueiro (2020); Tran & Santarelli (2014); Vannoni (2019); Zheng et al. (2016)

Note. Two hundred studies retrieved from Scopus across eighteen queries and assigned to a single theme each, so the counts sum to the corpus. Full details are in the reference list.

Section S2. Panel Estimation: Additional Figures

The two figures below plot the weighting sensitivity and the stability of the coefficients across estimators discussed in Appendix A of the main text.

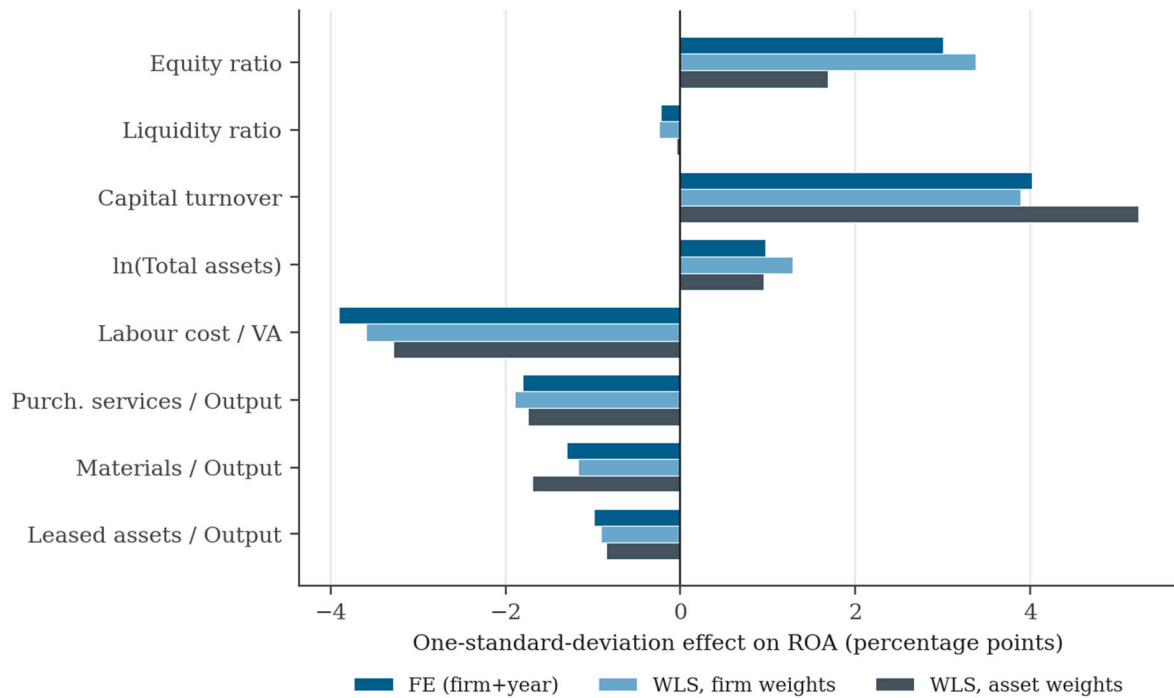


Figure S1. Weighting sensitivity, rescaled to one-standard-deviation effects.

Figure S1 shows that the weighting scheme is not a technicality. Weighting each firm equally rather than each observation raises the capitalisation coefficient from 0.177 to 0.199 and size from 0.646 to 0.855, so the relationship is somewhat stronger among firms observed briefly. Weighting by total assets, which recovers the coefficient relevant to the economic aggregate rather than to the median firm, halves capitalisation to 0.099 while raising capital turnover from 5.29 to 6.89 and materials from -0.114 to -0.149 : in the firms that carry the aggregate the operating model matters more and capital structure less. Neither weighting is correct in the abstract, since the two answer different questions. Figure S2 turns from weighting to the choice of estimator.

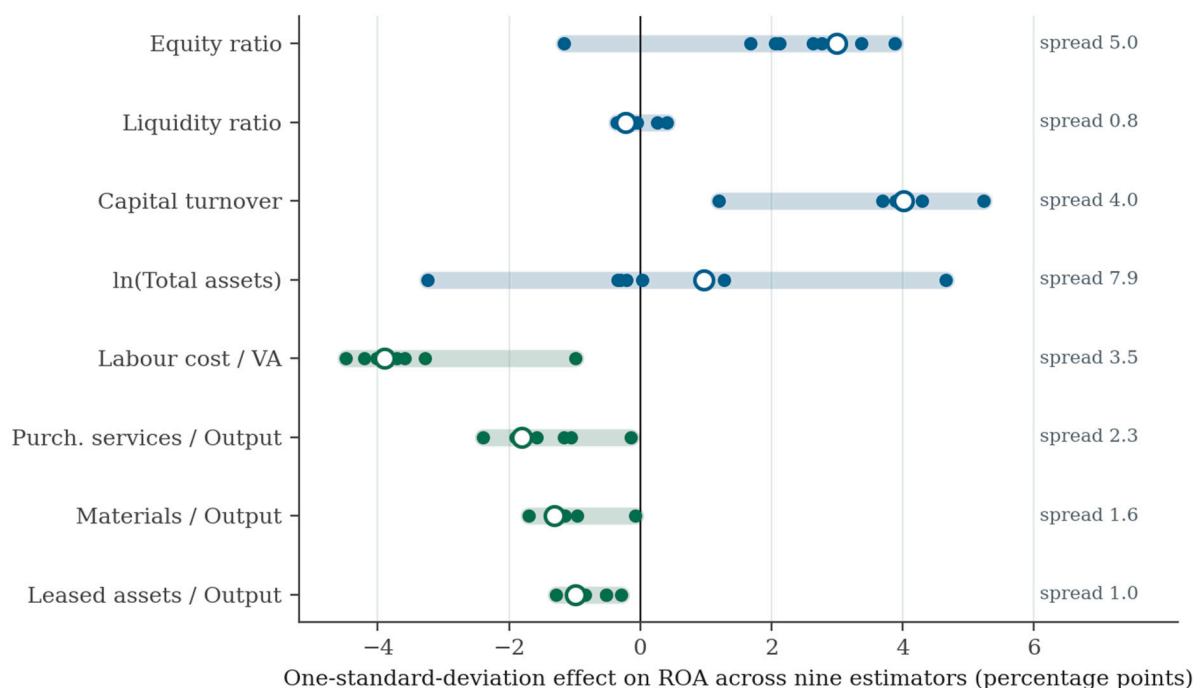


Figure S2. Coefficient stability across nine estimators.**Section S3. Instrumental Variables: Attempt and Diagnostics**

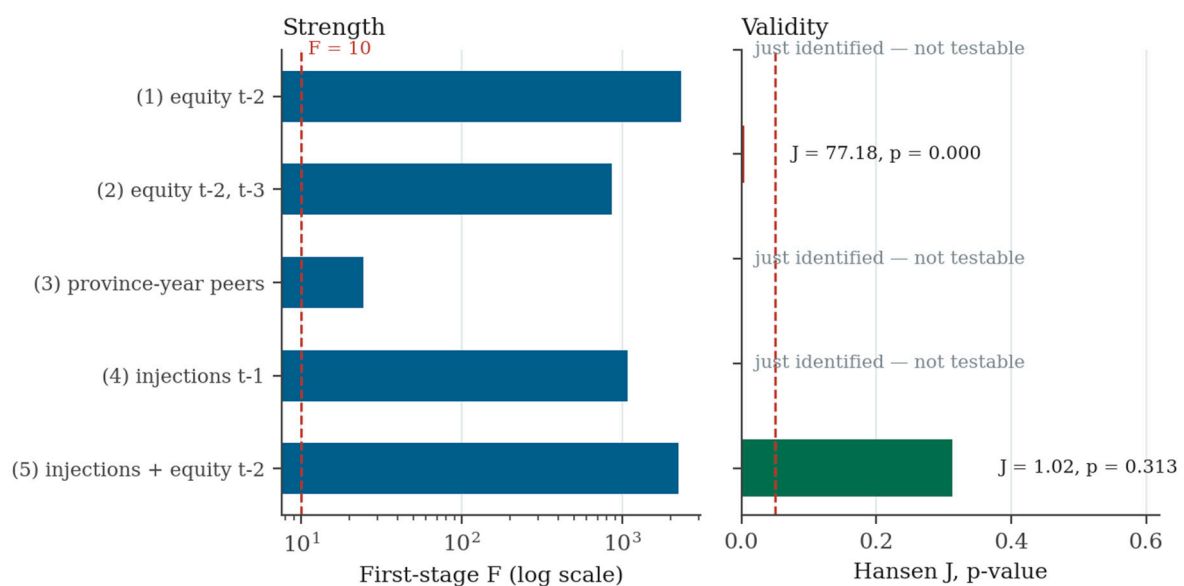
The equity ratio is the regressor most exposed to simultaneity, and it is the one whose sign is unstable across the specifications of Table 6. Five instrument sets were therefore tested within the two-way fixed-effects framework, with the equity ratio treated as endogenous and the remaining seven regressors as exogenous. Three come from inside the panel — the second lag of the equity ratio, the second and third lags jointly, and the second lag combined with an external instrument — and two are constructed: a leave-one-out province-year peer mean of the equity ratio, and net external equity injections, reconstructed as the change in net worth less the current year's profit, which is the one quantity that enters net worth without passing through current earnings.

Table S2. Two-stage least squares with firm and year fixed effects. Coefficient on the equity ratio.

Instrument set	β (equity)	s.e.	First-stage F	Hansen J	p(J)
FE, no instruments (benchmark)	0.1766***	(0.0049)	—	—	—
(1) Equity ratio (t-2)	-0.2769***	(0.0307)	2,326.5	—	—
(2) Equity ratio (t-2, t-3)	-0.2644***	(0.0344)	862.3	77.18	0.000
(3) Province-year peer mean	-0.1227	(0.2482)	24.4	—	—
(4) Net equity injections (t-1)	-0.2346***	(0.0411)	1,073.9	—	—
(5) Injections (t-1) + equity (t-2)	-0.2579***	(0.0225)	2,224.0	1.02	0.313

Note. Firm-clustered standard errors. All specifications include firm and year fixed effects and the seven remaining regressors as exogenous controls. Hansen J is reported only for the overidentified specifications; sets (1), (3) and (4) are just identified and the restriction cannot be tested. Observations: 64,557 for (1) and (5), 53,386 for (2), 91,534 for (3), 64,656 for (4).

The table reports strength and validity in adjacent columns, which invites the conflation the two diagnostics are meant to prevent. Figure S3 separates them.

**Figure S3.** Instrument strength and overidentification.

On the left, four of the five instrument sets clear the conventional weak-instrument threshold by two orders of magnitude and the province peer mean clears it by a factor of

two. On the right, the specification using the second and third lags fails the overidentification test decisively; the specification combining injections with the second lag passes it. Strength and validity are independent properties, and these instruments have the first without reliably having the second. Figure S4 shows what the instruments do to the estimate that matters.

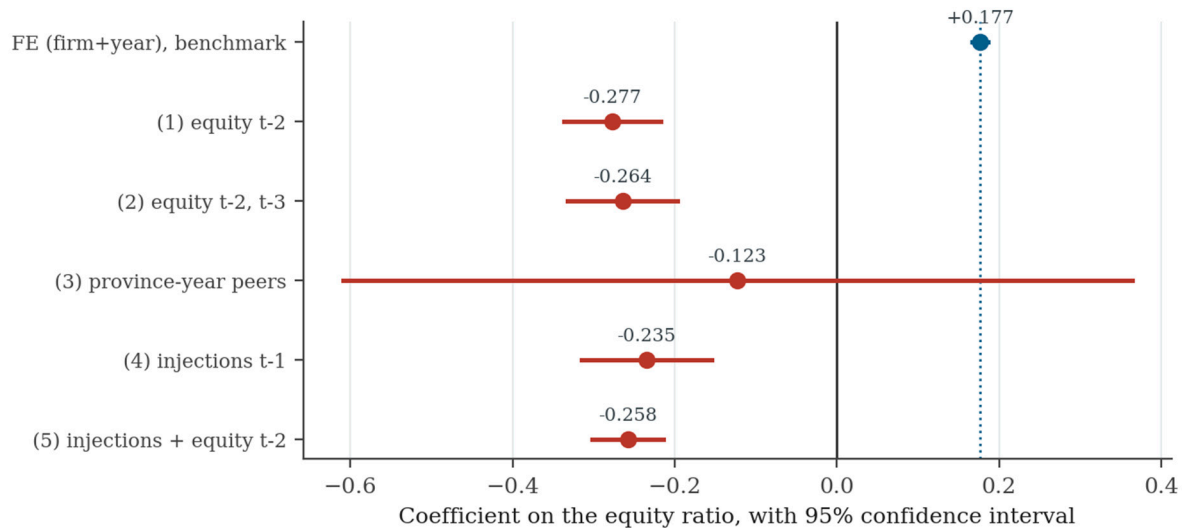


Figure S4. The equity coefficient under five instrument sets.

Every instrument set moves the coefficient from about +0.18 to between -0.12 and -0.28 , and the four precisely estimated sets agree closely with one another while disagreeing sharply with the benchmark. That pattern is the opposite of reassuring. Instruments that identify the same parameter through genuinely different exclusion restrictions should agree with one another only if all the restrictions hold; here the internal instruments share a single failure mode, which is why they agree, and the one instrument with an independent restriction is too weak to arbitrate.

Section S3.1. Why the Instruments Fail

First, overidentification is rejected where it can be tested. Specification (2) returns a Hansen J of 77.18 with p below 0.001. Specification (5) does not reject at 1.02, but its point estimate of -0.258 lies inside the range spanned by the specifications that do reject, so passing the test does not distinguish it substantively; it is a weaker test on a smaller set of restrictions, not independent evidence of validity.

Second, the mechanism of failure is identifiable rather than merely suspected. The autoregressive coefficient of ROA is 0.271, so profitability at t is substantially predicted by profitability at $t-2$. Since the equity ratio at $t-2$ mechanically embeds the profit of $t-2$, lagged capitalisation is correlated with the current error through persistence alone. This is a direct violation of the exclusion restriction rather than a subtle one, and no lag length escapes it while profitability remains persistent: lengthening the lag weakens the first stage without breaking the chain that runs from past profit to present profit.

Third, the one instrument whose exclusion restriction is defensible on economic grounds is too weakly related to the endogenous regressor to be usable. Net external equity injections enter net worth without passing through current profit, which is exactly what the restriction requires, but their within-firm correlation with the equity ratio is 0.103. Because the asymptotic bias of two-stage least squares stands to that of least squares roughly as the correlation between instrument and error stands to the correlation between instrument and regressor, a correlation of 0.103 amplifies any violation of exclusion by a

factor of about ten. The first-stage F of 1,074 reflects the size of the sample rather than instrument strength in the sense that matters here, which is a distinction the F statistic is not designed to make.

Table S3. Assessment of each instrument set against the two requirements.

Instrument set	Exclusion restriction rests on	Strength	Validity	Usable
(1) Equity (t-2)	Absence of profit persistence	Very strong	Violated by construction	No
(2) Equity (t-2, t-3)	Absence of profit persistence	Very strong	Rejected, J = 77.18	No
(3) Province-year peers	No local demand shocks	Weak in substance	Not testable	No
(4) Equity injections (t-1)	Injections independent of current profit	Nominal F only	Defensible	No
(5) Injections + equity (t-2)	Both of the above jointly	Very strong	Not rejected, J = 1.02	Not decisive

Note. Strength is assessed on the within-firm correlation between instrument and regressor rather than on the first-stage F alone, since with more than sixty thousand observations the F statistic is large for correlations far too low to control bias amplification.

Section S3.2. Placement of the Instrumental-Variable Results

Whether these results belong in the body of the paper or in an appendix is a judgement, and the case for each is worth stating. The argument for the body is that the instrumented estimates reverse the sign of the leading coefficient, and a result of that magnitude ordinarily belongs where the reader will see it. The argument for the appendix is that the instruments do not identify the parameter: two of the five fail a test they can be given, a third is weak in substance, and the mechanism of failure is common to all the internal instruments. Estimates from a rejected identification strategy reported alongside estimates from an accepted one invite the reader to average them, which would be the wrong operation.

The compromise adopted here places the summary in the body and the machinery in the appendix. Table 6 in Section 4 carries every instrumental-variable point estimate, so the reader sees the reversal and its magnitude without being asked to accept it; one paragraph states that the instruments fail and why; and the tables, diagnostics and the reasoning behind the verdict sit here. What is not done, and what the literature surveyed in Section 2 does frequently, is to report a single instrumented column as the paper's preferred specification on the strength of a first-stage F alone. On this sample that column exists — specification (5), which passes its overidentification test with a coefficient of -0.258 — and reporting it as the answer would be the most misleading single choice available.

The wider conclusion is negative and is stated as such: these data contain no valid instrument for capital structure. The feature that makes the dynamic specification necessary, the persistence of profitability at 0.271, is the same feature that destroys the internal instruments, and the external instrument that survives the argument does not survive the arithmetic of bias amplification. Identification of the capitalisation coefficient would require exogenous variation in capital structure — a regulatory threshold, a tax change, a lending shock differentially binding across otherwise similar firms — and no such variation is present in an accounting panel of this construction. The one candidate the Italian scheme provides, the step increases in the tax deduction granted to investors in certified firms, is examined in Section S8 and does not supply it.

Section S4. Clustering: Full Validity Battery and Within-Cluster Estimates

Reading down the silhouette column of Table S4 shows why no single index can decide the question. Silhouette falls monotonically in k for every partitioning method, so on that criterion $k = 2$ always wins, and Calinski–Harabasz behaves the same way; partition R^2 and the Dunn index move in the opposite direction. The two density-based configurations attain high silhouette while assigning between four and thirty-three per cent of firms to noise, and they are excluded on that ground rather than on their index values. The table is reported in full, including the configurations that were rejected, because the argument for the composite ranking rests on the disagreement between the indices and cannot be assessed from the winning row alone.

Table S4. Validity battery across algorithms and numbers of clusters.

Algorithm	k	Silh.	CH	DB	Dunn	MaxDiam	MinSe P	γ	Entropy	HHI	R^2	BIC	Min %
k-means	2	0.210	3472	1.851	0.0186	25.66	0.478	0.305	0.668	0.525	0.189	—	38.8
Ward	2	0.167	2731	2.110	0.0175	25.66	0.449	0.232	0.690	0.503	0.155	—	46.2
Gaussian mix.	2	0.115	1631	2.716	0.0135	25.66	0.345	0.154	0.683	0.510	0.099	316,325	43.0
Fuzzy c-means	2	0.198	3377	1.902	0.0134	25.66	0.344	0.281	0.690	0.504	0.184	—	45.7
k-means	3	0.172	2880	1.836	0.0161	21.65	0.348	0.281	1.091	0.338	0.278	—	28.4
Ward	3	0.146	2149	2.145	0.0180	21.58	0.389	0.285	1.043	0.369	0.224	—	19.6
Gaussian mix.	3	0.120	1460	3.324	0.0122	25.66	0.314	0.243	1.089	0.340	0.164	282,504	27.9
Fuzzy c-means	3	0.146	2503	2.217	0.0147	23.15	0.340	0.277	1.088	0.340	0.251	—	26.7
k-means	4	0.172	2601	1.784	0.0159	21.55	0.342	0.316	1.347	0.269	0.343	—	14.9
Ward	4	0.154	1965	1.804	0.0180	21.58	0.389	0.313	1.130	0.358	0.283	—	3.2
Gaussian mix.	4	0.077	1227	3.300	0.0097	25.66	0.249	0.272	1.331	0.275	0.198	273,297	13.2
Fuzzy c-means	4	0.108	1909	2.841	0.0118	22.44	0.264	0.271	1.285	0.292	0.277	—	8.3
k-means	5	0.177	2518	1.654	0.0162	21.55	0.348	0.366	1.460	0.247	0.403	—	3.5
Ward	5	0.148	1935	1.921	0.0180	21.58	0.389	0.344	1.439	0.257	0.342	—	3.2
Gaussian mix.	5	0.065	1284	3.335	0.0101	25.66	0.260	0.256	1.562	0.217	0.256	262,697	9.9
Fuzzy c-means	5	0.077	1557	5.071	0.0118	22.01	0.260	0.265	1.441	0.258	0.294	—	5.1
k-means	6	0.179	2453	1.533	0.0162	21.55	0.348	0.403	1.572	0.229	0.451	—	3.4
Ward	6	0.141	1855	1.691	0.0180	21.58	0.389	0.358	1.553	0.243	0.383	—	3.2
Gaussian mix.	6	0.059	1101	3.129	0.0102	25.54	0.260	0.246	1.749	0.180	0.269	257,268	8.2
Fuzzy c-means	6	0.053	1225	10.733	0.0118	22.01	0.260	0.262	1.430	0.264	0.291	—	0.9
k-means	7	0.169	2357	1.508	0.0159	20.63	0.329	0.393	1.758	0.190	0.487	—	3.1
Ward	7	0.086	1716	1.799	0.0150	20.74	0.311	0.304	1.777	0.189	0.408	—	3.2

Gaussian mix.	7	0.050	994	3.052	0.0063	25.66	0.163	0.258	1.868	0.165	0.286	241,245	7.9
Fuzzy c-means	7	0.076	1065	8.066	0.0180	21.59	0.389	0.263	1.521	0.239	0.300	—	0.2
DBSCAN ($\epsilon=0.6$)	5	-0.176	269	1.409	0.1504	3.98	0.598	-0.078	0.725	0.608	0.083	—	1.1
DBSCAN ($\epsilon=0.8$)	4	-0.075	270	2.092	0.1192	7.04	0.840	0.065	0.096	0.969	0.067	—	0.3
RF proximity	2	0.202	911	1.957	0.0167	23.82	0.398	0.299	0.600	0.591	0.154	—	28.7
RF proximity	3	0.161	652	2.263	0.0160	23.82	0.382	0.345	0.942	0.437	0.207	—	13.3
RF proximity	4	-0.043	439	2.649	0.0160	23.82	0.382	0.334	0.990	0.426	0.209	—	0.9

Note. γ is the Pearson gamma between pairwise distances and the co-membership indicator. Entropy and the Herfindahl–Hirschman index measure balance across clusters. BIC is defined only for the Gaussian mixture. Min % is the share of firms in the smallest cluster; for the density-based rows it excludes the firms assigned to noise. Indices requiring pairwise distances are computed on a random subsample of 5,000 firms. The shaded row is the selected configuration.

Figure S5 plots the four indices that disagree most sharply, restricted to the partitional families for legibility. Taken together the four panels make the selection problem visible: silhouette and Calinski–Harabasz decline in k while partition R^2 rises, and the Dunn index is nearly flat for k -means across the whole range.

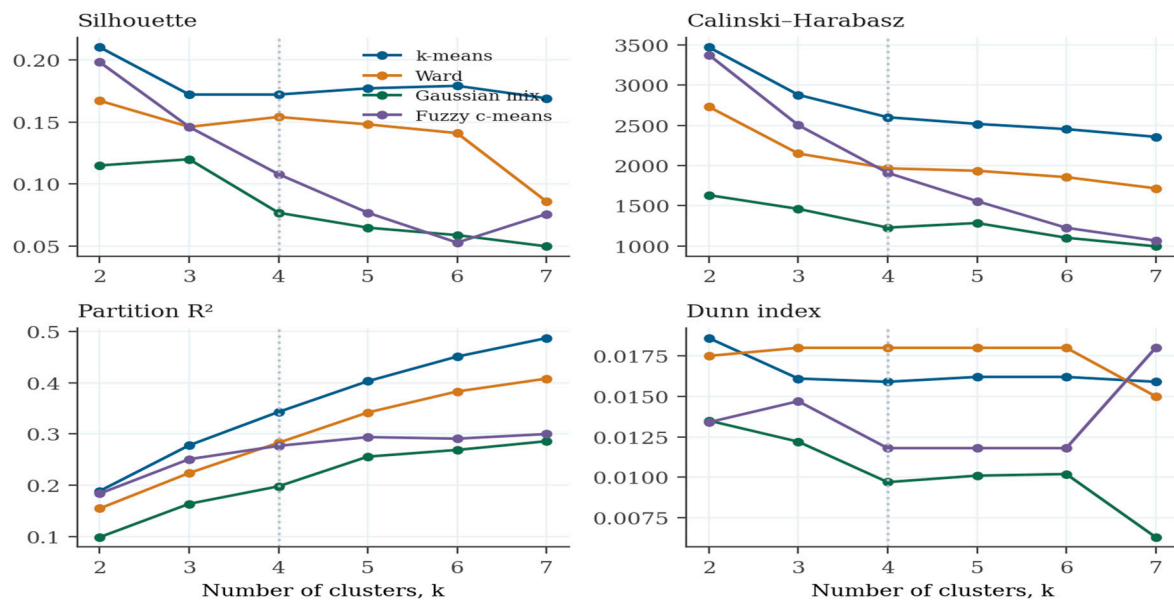


Figure S5. Validity indices across algorithms and number of clusters.

Density-based and random-forest proximity configurations are omitted from the panels and reported in Table S4. The dotted vertical line marks the selected k . Nothing in any single panel identifies it: $k = 4$ is the point at which k -means leads on partition R^2 among the solutions that remain balanced, and the composite ranking below is what makes that consideration operative.

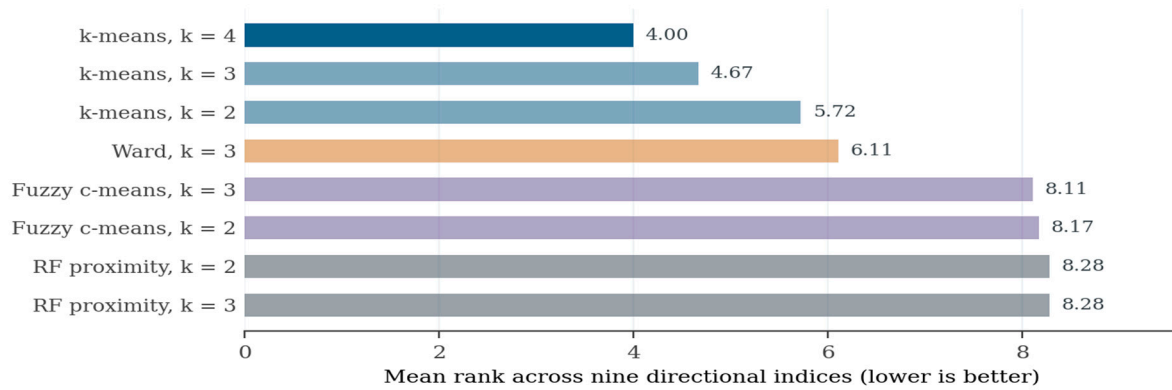


Figure S6. Composite ranking of the retained configurations.

The composite ranking separates $k = 4$ from its nearest rival by a narrow margin, and the two leading configurations were therefore both examined in full. At $k = 3$ the labour-intensive and materials-intensive groups are unchanged and the micro-service and capitalised groups merge; the coefficient pattern of Table S8 survives, with capital turnover still ranging by a factor of three across groups and the labour share still invariant. Fuzzy c-means at $k = 2$ and $k = 3$ rank immediately behind k-means, which is reassuring rather than redundant, since the same structure is reached by a procedure that does not impose hard assignment.

Table S5. Composite ranking (mean rank across nine directional indices; lower is better).

Algorithm	k	Composite rank
k-means	4	4.00
k-means	3	4.67
k-means	2	5.72
Ward	3	6.11
Fuzzy c-means	3	8.11
Fuzzy c-means	2	8.17
RF proximity	2	8.28
RF proximity	3	8.28

Note. Configurations with a smallest cluster below five per cent of firms are excluded, as are density-based solutions classifying more than five per cent of firms as noise. Nine of the eleven criteria are directional and enter the mean rank; maximum diameter and BIC are reported in Table S4 but not ranked, the first because it is bounded by the sample and the second because it is defined for one family only.

Table S6 reports the medians, interpretable in the original units, alongside the standardised means that show each variable's contribution to the separation. The two columns tell different stories about the financial variables. In medians, Cluster 3 is four times as capitalised as Cluster 0, 70.1 against 17.3, and its liquidity ratio is three times as high; in standardised terms both variables are flat across the other three clusters, so their large ranges are generated by a single group. The labour share moves across all four groups, from +0.71 to −0.80, and purchased services from +0.92 to −0.72.

Table S6. Cluster profiles, medians and standardised means.

Variable	C0 med.	C1 med.	C2 med.	C3 med.	C0 std.	C1 std.	C2 std.	C3 std.
ROA (%)	3.03	4.72	7.36	10.23	-0.43	-0.06	0.30	0.60

Equity ratio (%)	17.29	33.22	34.71	70.10	-0.60	0.01	0.01	1.33
Liquidity ratio	1.22	1.08	1.31	3.70	-0.27	-0.39	-0.27	1.76
Capital turnover	1.23	1.03	0.69	0.59	0.38	0.11	-0.31	-0.56
ln(Total assets)	7.58	9.75	4.79	5.60	0.16	0.93	-0.94	-0.53
Labour cost / VA (%)	89.63	67.55	0.63	35.23	0.71	0.06	-0.80	-0.44
Purch. services / Output (%)	29.94	18.88	57.35	34.94	-0.13	-0.72	0.92	0.08
Materials / Output (%)	3.04	47.43	0.81	0.82	-0.42	1.29	-0.54	-0.50
Leased assets / Output (%)	3.69	1.95	0.68	1.15	0.39	-0.20	-0.25	-0.13
Firms	5,067	4,023	3,615	2,224				
Share of firms (%)	33.9	26.9	24.2	14.9				

Note. Medians are of firm-level decade medians in original units. Standardised means are computed on the standardised variables used by the algorithm.

The contingency of the partition with regulatory status is reported below as row percentages, so that each row is the distribution of one register population across the four archetypes. Regulatory status enters no stage of the clustering procedure.

Table S7. Contingency of partition with regulatory status (row percentages).

Regulatory population	C0 labour	C1 materials	C2 micro service	C3 capitalised
Innovative start-ups	16.7	6.1	53.1	24.2
Innovative SMEs	35.7	15.5	27.4	21.3
Conventional SMEs	46.0	46.9	1.6	5.5

Note. $\chi^2(6) = 6,821.9$, $p < 0.0001$, Cramér's $V = 0.4780$. Rows sum to 100.

The full within-cluster estimates follow, plotted first because the coefficients occupy two very different scales: turnover and size are an order of magnitude larger than the cost shares and cannot be read on a common axis. Three features deserve attention beyond the dispersion discussed in Section 5. Liquidity is insignificant in three of the four groups, confirming from an independent direction the inertness already visible in Section 4. Materials are insignificant in the capitalised cluster, where the median materials share is 0.82 per cent of output and the variable has almost no variation to exploit — an absence of data rather than an absence of effect. And the labour share is the only coefficient whose confidence intervals overlap across all four groups.

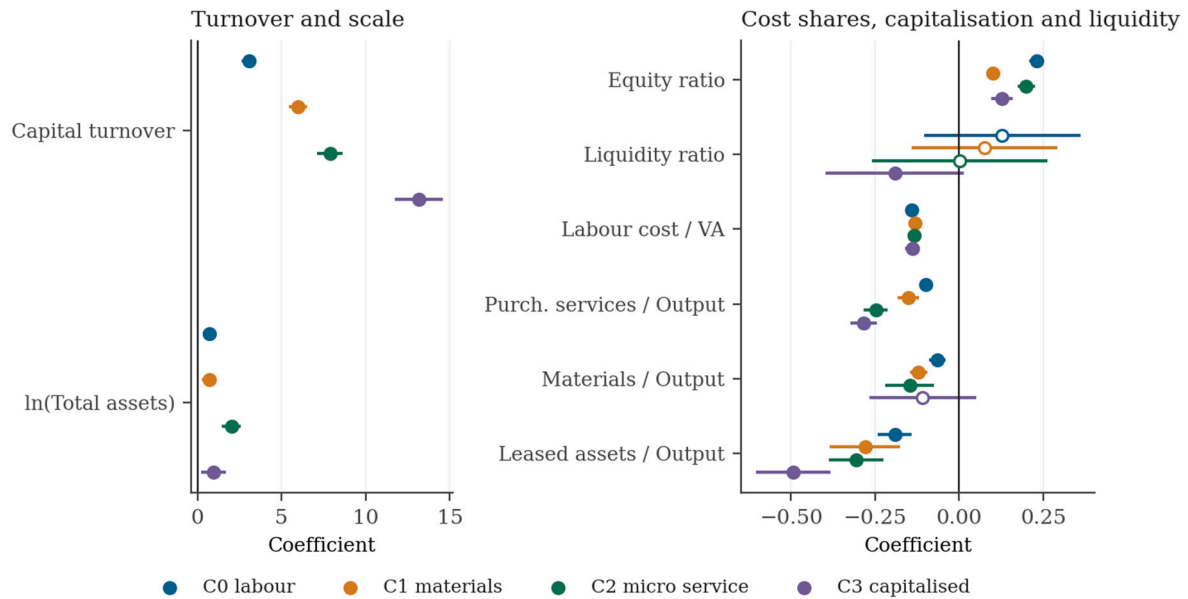


Figure S7. Equation coefficients re-estimated within each cluster.

Hollow markers denote coefficients not significant at the ten per cent level. The intervals are ninety-five per cent and firm-clustered. Table S8 gives the same estimates with standard errors and fit statistics.

Table S8. Equation re-estimated within each cluster. Dependent variable: ROA (%).

Regressor	C0 labour	C1 materials	C2 micro service	C3 capitalised
Equity ratio (%)	0.2314*** (0.0091)	0.1015*** (0.0069)	0.2002*** (0.0104)	0.1280*** (0.0136)
Liquidity ratio	0.1279 (0.1163)	0.0756 (0.1081)	0.0015 (0.1311)	-0.1909* (0.1025)
Capital turnover	3.0860*** (0.1742)	5.9623*** (0.2216)	7.8726*** (0.3498)	13.1552*** (0.6774)
ln(Total assets)	0.7229*** (0.1583)	0.6971*** (0.1736)	2.0210*** (0.2450)	0.9727*** (0.3274)
Labour cost / Value added (%)	-0.1403*** (0.0038)	-0.1317*** (0.0047)	-0.1347*** (0.0061)	-0.1383*** (0.0084)
Purchased services / Output (%)	-0.0980*** (0.0075)	-0.1510*** (0.0139)	-0.2475*** (0.0155)	-0.2852*** (0.0176)
Materials / Output (%)	-0.0646*** (0.0105)	-0.1205*** (0.0104)	-0.1466*** (0.0347)	-0.1081 (0.0789)
Leased assets / Output (%)	-0.1909*** (0.0234)	-0.2800*** (0.0505)	-0.3051*** (0.0390)	-0.4937*** (0.0545)
Observations	34,377	34,632	12,804	10,062
Firms	5,060	4,019	3,612	2,223
Within R ²	0.4293	0.4955	0.4420	0.4759

Note. Two-way fixed effects, firm-clustered standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Firm counts fall marginally below those of Table S6 because a small number of firms contribute no usable within variation once the equation is estimated inside the group.

The Wald statistics test equality of the full eight-element coefficient vector for each pair of clusters, with variance equal to the sum of the two clustered covariance matrices. They are reported for completeness rather than as the decisive evidence: on samples of this size rejection is close to automatic, and the informative quantity is the magnitude of the differences reported in Table 10, not the p-value attached to them.

Table S9. Pairwise Wald tests of coefficient equality across clusters.

Comparison	χ^2	d.f.	p
Cluster 0 vs. Cluster 1	264.42	8	< 0.0001
Cluster 0 vs. Cluster 2	262.68	8	< 0.0001
Cluster 0 vs. Cluster 3	374.34	8	< 0.0001
Cluster 1 vs. Cluster 2	125.42	8	< 0.0001
Cluster 1 vs. Cluster 3	153.27	8	< 0.0001
Cluster 2 vs. Cluster 3	84.45	8	< 0.0001

Note. Eight degrees of freedom. The smallest statistic, 84.45, is between the micro-service and capitalised clusters, which are also the two groups the three-cluster solution merges.

Section S5. Machine-Learning Regression: Figures and Full Results

The first four figures illustrate the machine-learning results of Section 6 of the main text: accuracy under the three transformations, parametric enrichment, permutation importance and calibration by decile. The full results follow.

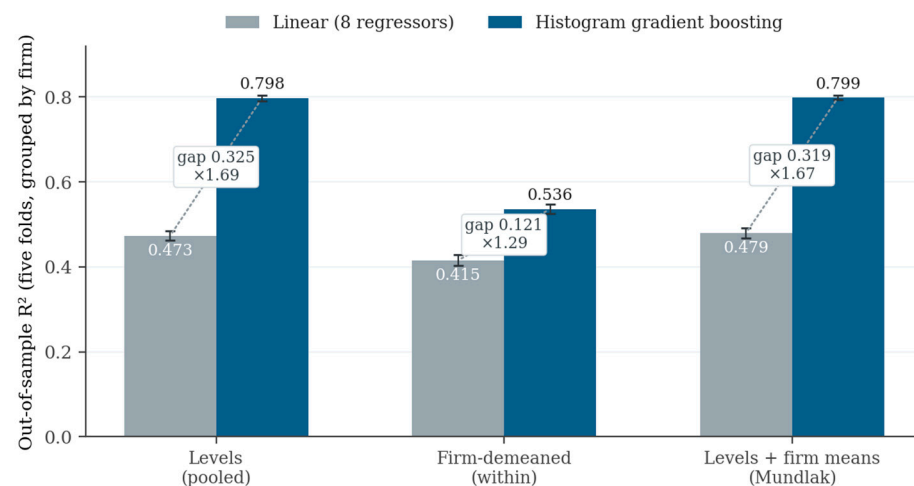


Figure S8. Out-of-sample accuracy under three transformations.

Figure S8 compares the linear and the non-linear fit under the three treatments of between-firm variation reported in Table 11 of the main text. Figure S9 asks how much of the remaining within-firm gap parametric enrichment of the linear model can recover.

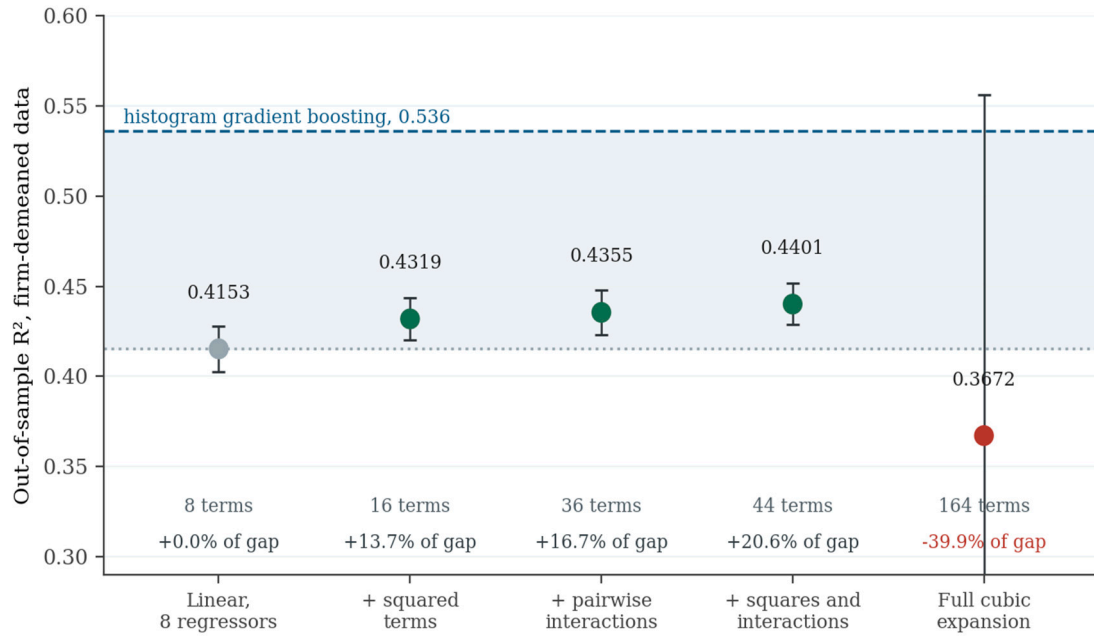


Figure S9. Parametric enrichment against gradient boosting, firm-demeaned data.

Figure S9 shows that squares and interactions recover at most a fifth of the gap and that the full cubic expansion loses ground. Figure S10 turns from fit to the contribution of each predictor.

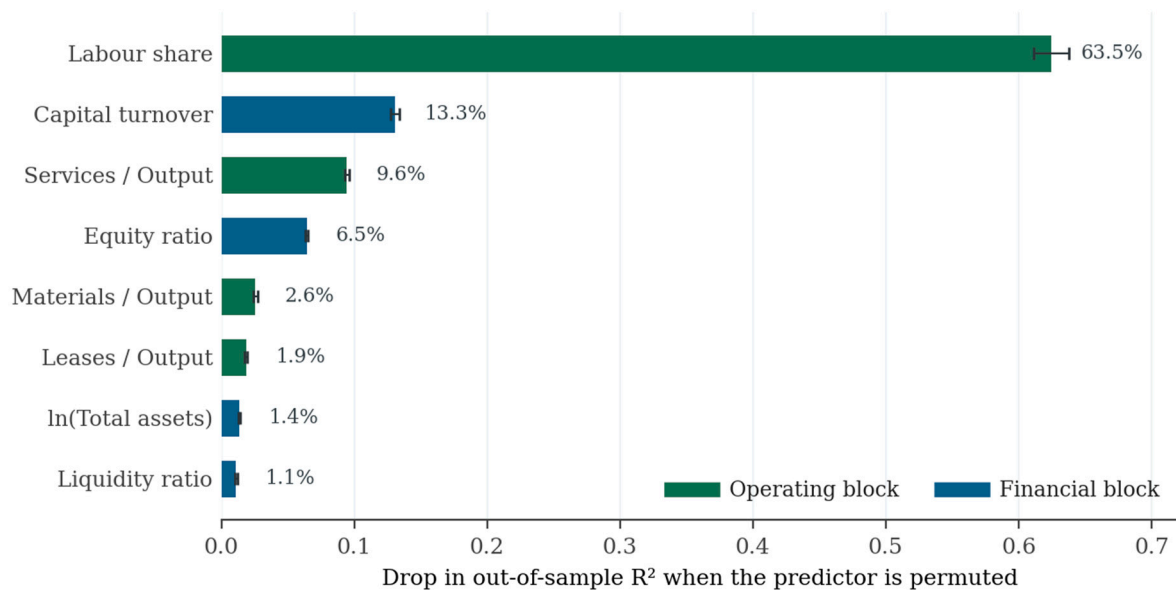


Figure S10. Permutation importance, firm-demeaned data.

Figure S10 ranks the predictors by permutation importance, with the labour share at 63.5 per cent and the equity ratio at 6.5 per cent. Figure S11 locates where in the distribution of profitability the two models differ.

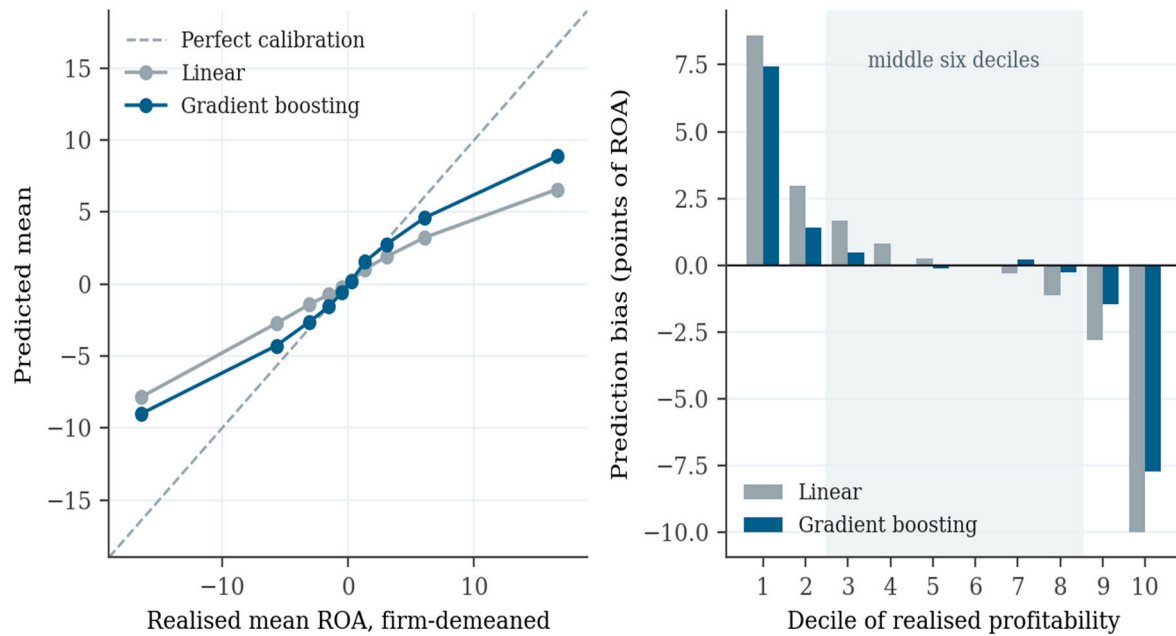


Figure S11. Calibration by decile of realised profitability.

The algorithm comparison is reported on raw levels, as such comparisons are conventionally presented; the like-for-like within-transformed figures of Section 6 are those comparable with the panel estimates. Two columns beyond the headline deserve attention. The cross-fold standard deviation separates stable learners from unstable ones: the leading methods sit between 0.0066 and 0.0082 while AdaBoost reaches 0.1250, meaning its accuracy depends heavily on which firms fall into the test partition. And the timing column matters for replication, since histogram gradient boosting attains the best accuracy in 3.3 seconds against 120.5 for the random forest.

Table S10. Algorithm comparison, raw levels.

Model	R ² (out of sample)	SD across folds	RMSE	MAE	Seconds
Histogram gradient boosting	0.7978	0.0070	5.932	3.476	3.3
Neural network (MLP)	0.7875	0.0082	6.082	3.613	52.5
Random forest	0.7786	0.0078	6.209	3.646	120.5
Extra trees	0.7650	0.0076	6.397	3.803	19.4
Gradient boosting	0.7195	0.0072	6.989	4.299	95.7
k-nearest neighbours	0.6945	0.0069	7.293	4.472	10.7
Decision tree	0.6436	0.0066	7.877	4.893	2.3
Lasso	0.4730	0.0112	9.578	6.210	0.2
Ridge	0.4730	0.0112	9.578	6.212	0.2
Linear regression	0.4730	0.0112	9.578	6.212	0.2
Elastic net	0.4730	0.0111	9.578	6.208	0.2
Linear SVM	0.4551	0.0108	9.740	6.065	1.8
AdaBoost	0.2060	0.1250	11.714	8.510	25.7

Note. Dependent variable ROA (%), same eight regressors and same 91,756 observations as the panel estimates. Five-fold cross-validation grouped by firm. The shaded row is the algorithm carried forward.

Table S10 reports the full comparison of the thirteen algorithms on raw levels. Figure S12 plots the same accuracy measures and sets accuracy against fitting time.

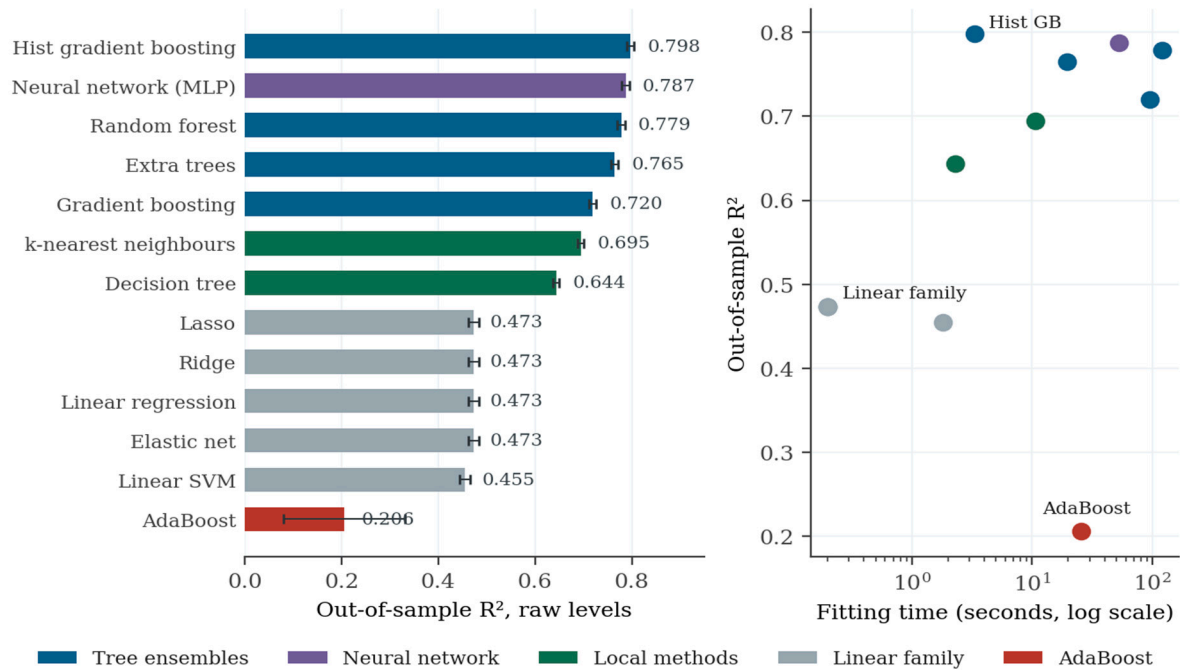


Figure S12. Out-of-sample accuracy by algorithm, and accuracy against fitting time.

Three bands emerge rather than a continuum, and they correspond to what each family can represent: thresholds and interactions, smooth local structure, and additive linearity. The right-hand panel adds the dimension a ranking hides. Accuracy is not bought with computation — the selected algorithm is thirty-six times faster than the random forest it outperforms — and the linear family is separated from the ensembles by a difference in representational capacity rather than in effort. AdaBoost is the one clear failure in the battery, its exponential reweighting chasing the tails of a heavy-tailed target; its failure characterises the distribution rather than the specification, and it is retained in the table for that reason.

The learning curve settles whether the remaining gap is an artefact of sample size. The gradient-boosting test score rises little over the final two thirds of the data while its training score falls towards it, which is the signature of a model that has stopped memorising rather than one starved of observations. The linear model is flat almost from the start.

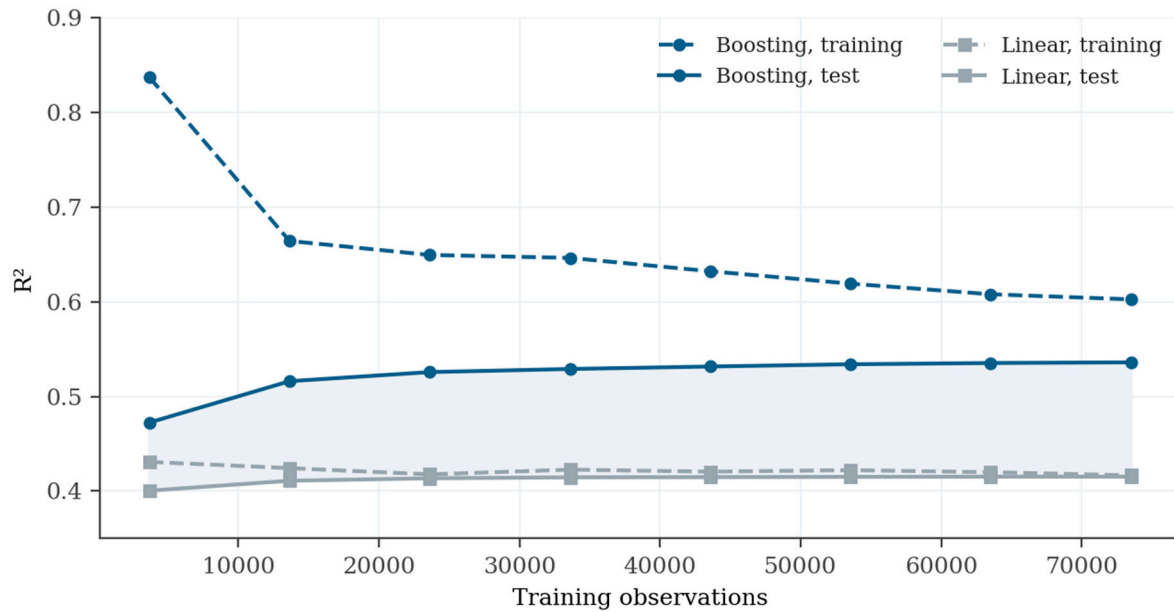


Figure S13. Learning curves, firm-demeaned data.

Figure S13 shows the linear model flat from the first ten thousand observations while the boosting test score converges from above. Table S11 reports the values plotted.

Table S11. Learning curve values.

Training observations	Linear, train	Linear, test	Boosting, train	Boosting, test
3,675	0.4309	0.4005	0.8372	0.4724
13,650	0.4243	0.4110	0.6642	0.5161
23,624	0.4177	0.4136	0.6494	0.5257
33,600	0.4227	0.4146	0.6463	0.5290
43,575	0.4206	0.4148	0.6322	0.5316
53,550	0.4222	0.4151	0.6191	0.5340
63,525	0.4200	0.4153	0.6079	0.5353
73,500	0.4167	0.4153	0.6025	0.5360

Note. The shaded band in Figure S13 is the distance between the two test curves, which is the quantity Section 6 discusses. More data will not close it: the linear test score moves by 0.015 across a twenty-fold increase in training observations.

Interaction strength is measured by the Friedman H-statistic, which reports the share of the joint partial-dependence variation attributable to interaction rather than to the two separate effects. The strongest pair is the equity ratio with capital turnover at 0.042 — real but second-order. A purely additive surface would return zero throughout.

Permutation importance is computed on held-out data, so it measures contribution to genuine prediction rather than in-sample fit. The four lowest predictors together account for 6.9 per cent of predictive content, which means the equation could be reduced to four regressors with modest predictive loss — though not without losing controls the panel specification requires for reasons that have nothing to do with prediction.

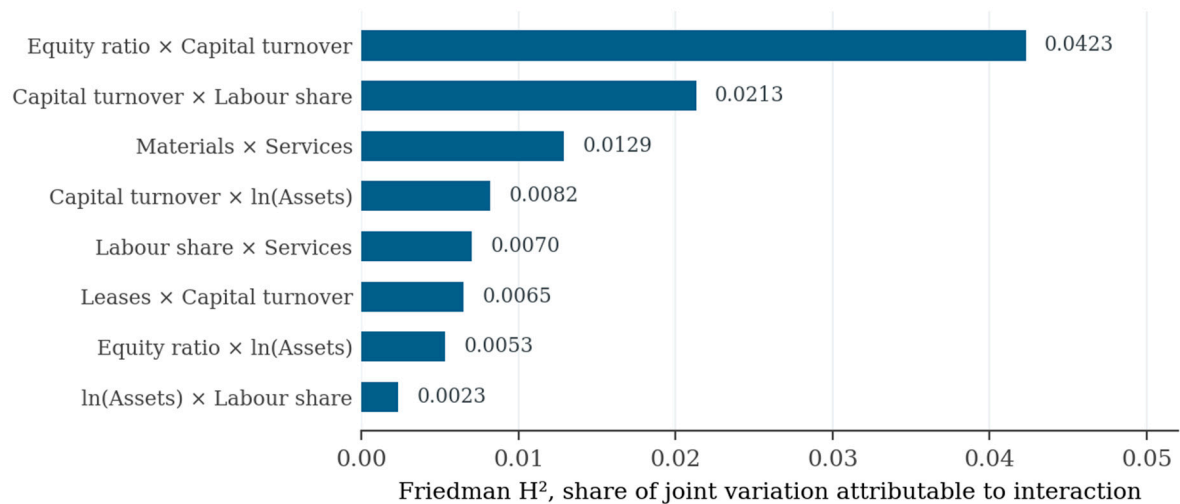


Figure S14. Interaction strength by pair.

Figure S14 shows that interaction strength is small for every pair of regressors. Table S12 reports the underlying H-statistics.

Table S12. Friedman H-statistic by pair.

Pair	H ²
Equity ratio × Capital turnover	0.0423
Capital turnover × Labour share	0.0213
Materials / Output × Services / Output	0.0129
Capital turnover × ln(Total assets)	0.0082
Labour share × Services / Output	0.0070
Leases / Output × Capital turnover	0.0065
Equity ratio × ln(Total assets)	0.0053
ln(Total assets) × Labour share	0.0023

Note. Computed on a 3,000-observation subsample of the held-out fold. The result corroborates the parametric exercise from the other direction: interactions recovered 16.7 per cent of the gap against 13.7 for squared terms alone, so neither dominates and both leave four fifths unexplained.

No pair reaches an H-statistic of 0.05, so the non-linearity the learner captures is univariate curvature rather than joint dependence. Table S13 reports the permutation importance underlying Figure S10.

Table S13. Permutation importance, firm-demeaned data.

Predictor	Drop in R ²	SD	Share of total (%)
Labour share	0.6247	0.0132	63.5
Capital turnover	0.1309	0.0034	13.3
Services / Output	0.0946	0.0017	9.6
Equity ratio	0.0642	0.0010	6.5
Materials / Output	0.0257	0.0018	2.6
Leases / Output	0.0187	0.0010	1.9
ln(Total assets)	0.0137	0.0007	1.4

Liquidity ratio	0.0110	0.0011	1.1
-----------------	--------	--------	-----

Note. Twenty permutations per predictor on the held-out fold. Shares are of the summed drop across the eight predictors.

Accuracy by subgroup is reported against both partitions used in the paper, the regulatory one of Section 4 and the estimated one of Section 5. The gap between linear and non-linear fits is widest in the materials-intensive cluster and narrowest in the micro-service one, where the linear form is nearly adequate. The absolute levels matter as much as the gaps: between 38 and 48 per cent of within-firm variation is linearly predictable in every subgroup, a more even picture than the pooled figures suggest.

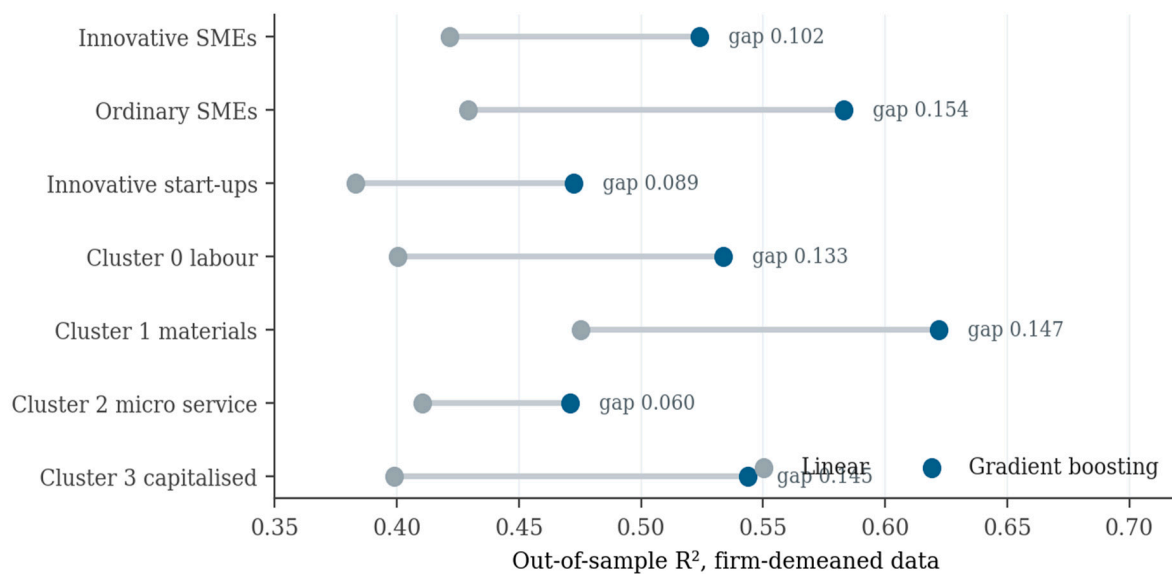


Figure S15. Out-of-sample accuracy by subgroup, firm-demeaned data.

Figure S15 shows the gap between the two models subgroup by subgroup. Table S14 reports the values plotted.

Table S14. Out-of-sample accuracy by subgroup, firm-demeaned data.

Subgroup	Observations	Linear R ²	Boosting R ²	Gap
Innovative SMEs	19,576	0.4215	0.5238	0.1023
Conventional SMEs	59,435	0.4291	0.5829	0.1537
Innovative start-ups	12,864	0.3830	0.4723	0.0893
Cluster 0 labour	34,377	0.4004	0.5336	0.1331
Cluster 1 materials	34,632	0.4752	0.6218	0.1466
Cluster 2 micro service	12,804	0.4106	0.4710	0.0604
Cluster 3 capitalised	10,062	0.3988	0.5436	0.1447

Note. Each subgroup is fitted and evaluated separately with the same five-fold grouped cross-validation. Observation counts for the regulatory populations differ marginally from Table 2 because the subgroup fits require a complete lag-free record.

Finally the decile calibration underlying Figure S11. Reading down the two bias columns shows that both models are well calibrated in the middle six deciles and that essentially all of the difference between them arises in the two deciles at each end.

Table S15. Prediction bias by decile of realised profitability.

Decile	Actual mean	Linear mean	Boosting mean	Linear bias	Boosting bias
1	-16.46	-7.86	-9.02	+8.60	+7.43
2	-5.70	-2.73	-4.28	+2.97	+1.42
3	-3.09	-1.42	-2.63	+1.66	+0.45
4	-1.55	-0.74	-1.57	+0.82	-0.02
5	-0.48	-0.22	-0.59	+0.26	-0.11
6	0.26	0.23	0.22	-0.04	-0.04
7	1.34	1.04	1.57	-0.30	+0.22
8	3.04	1.90	2.76	-1.14	-0.28
9	6.04	3.22	4.59	-2.82	-1.45
10	16.59	6.58	8.88	-10.01	-7.71

Note. Firm-demeaned ROA on the held-out fold. Bias is predicted minus actual, so positive values indicate over-prediction. Deciles are of realised profitability, so the first and last rows are the observations on which any specification of this equation performs worst.

Section S6. The Accounting Identity: Figures and Robustness of the Diagnostic

The three figures illustrate the tests of Section 7 of the main text. Table S16 then repeats the central comparison of the identity diagnostic across partitions and specifications.

The value-added ratio used in these tests is not among the variables of the estimating equation and was computed separately, by reapplying the sample construction of Section 3 of the main text to the source extractions: positive value of production, positive value added, and value of production of at least one per cent of total assets. The resulting 96,052 firm-year observations on 15,501 firms exceed the estimating sample of Table 2 by 4.7 per cent, because the ratio requires only two of the nine variables and is defined for firm-years that the full specification drops for missingness elsewhere; the median of the ratio changes by less than 0.01 when the sample is restricted to firm-years with all nine variables present. For Table 14, cluster membership is reconstructed by reapplying the procedure of Section 5 — decade medians of the nine standardised variables, k-means at $k = 4$ — to the 94,258 firm-year observations on 15,220 firms for which all nine variables are present, a sample 2.7 per cent larger than that of Table 2. The reconstruction reproduces the published partition closely: cluster shares are 33.2, 27.0, 24.0 and 15.8 per cent against 33.9, 26.9, 24.2 and 14.9 in Table 8, and the profile medians agree to within a few tenths on every variable. The comparison in Table 14 is therefore between a value-added ratio computed on the reconstructed partition and coefficients estimated on the original one; the conclusion turns on a ratio of 2.8 against 1.07 and is not sensitive to a discrepancy of this size.

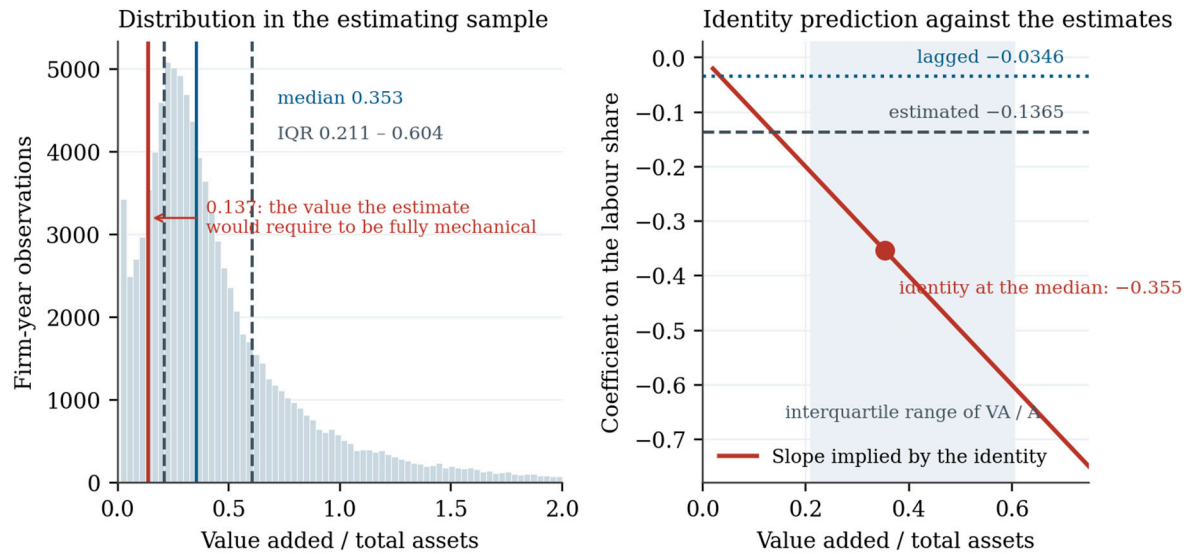


Figure S16. The distribution of value added over total assets, and the slope the identity implies.

Figure S16 shows the distribution of value added over total assets and the slope the identity implies at each point of it. Figure S17 turns to the lagged specification, in which the identity is broken.

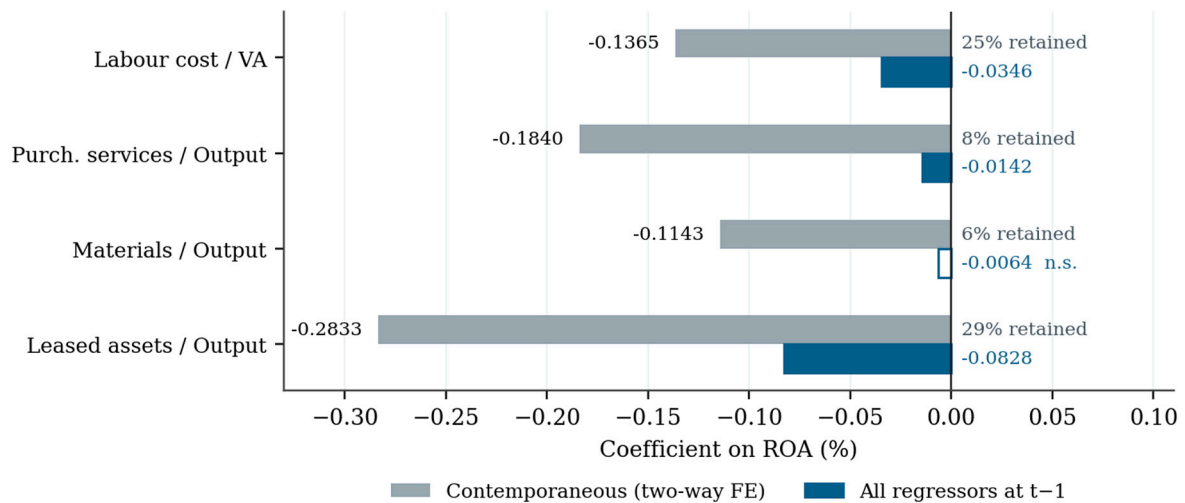


Figure S17. What survives when the identity is broken by lagging.

Figure S17 shows that a quarter of the labour-share coefficient survives when the identity is broken by lagging. Figure S18 compares the identity prediction with the estimate across the four operating archetypes.

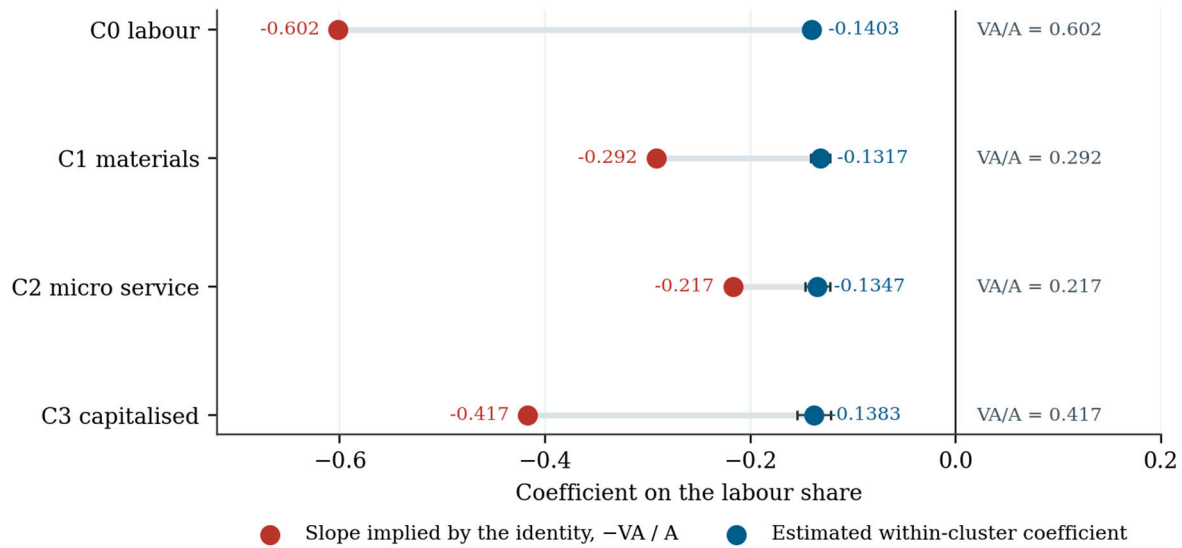


Figure S18. The identity prediction against the estimate, by archetype.

The worked example uses the pooled sample and the archetypes. The robustness table repeats the comparison on the pooled sample under three definitions of the implied slope, on the three regulatory populations, on the four archetypes, on four sub-periods and on the lagged specification.

Table S16. Robustness of the identity diagnostic across partitions and specifications.

Partition or specification	Median VA/A	Implied slope	Estimated coefficient	Ratio, estimate / implied	Range of VA/A across groups	Range of coefficient across groups
Pooled, unconditional identity	0.353	-0.353	-0.1365***	0.39	—	—
Pooled, total derivative	—	-0.375	-0.1365***	0.36	—	—
Pooled, weighted by within-firm variance of s	—	-0.283	-0.1365***	0.48	—	—
Pooled, first and third quartile of VA/A	0.211 / 0.604	-0.211 / -0.604	-0.1365***	0.65 / 0.23	—	—
Regime: innovative start-ups	0.213	-0.213	-0.1311***	0.62		
Regime: innovative SMEs	0.304	-0.304	-0.1170***	0.38		
Regime: conventional SMEs	0.394	-0.394	-0.1632***	0.41	1.8×	1.4×
Archetype: C0 labour	0.602	-0.602	-0.1403***	0.23		
Archetype: C1 materials	0.292	-0.292	-0.1317***	0.45		
Archetype: C2 micro service	0.217	-0.217	-0.1347***	0.62		
Archetype: C3 capitalised	0.417	-0.417	-0.1383***	0.33	2.8×	1.07×
Sub-period 2014–2016	0.392	-0.392	-0.1823***	0.47		
Sub-period 2017–2019	0.388	-0.388	-0.1579***	0.41		
Sub-period 2020–2022	0.326	-0.326	-0.0841***	0.26		
Sub-period 2023–2025	0.336	-0.336	-0.0737***	0.22	1.2×	2.5×

Lagged regressors (identity broken)	0.035 required	—	-0.0346***	—	—	—
-------------------------------------	----------------	---	------------	---	---	---

Note. The implied slope is $-VA/A$ evaluated at the group median. Pooled and archetype rows use Tables 4, 13 and 14; regime rows use the medians of Table 13 and the coefficients of Table A3; sub-period rows use the rebuilt estimating sample of Appendix A.1 and the coefficients of Table A7; the lagged row uses Table A2. *** $p < 0.01$. The ranges are the ratio of the largest to the smallest group value. Across regimes the coefficient varies less than VA/A and does not follow its ordering; across archetypes it is nearly invariant while VA/A varies almost threefold; across sub-periods it varies more than VA/A , so its decline over time is not produced by the identity either.

Two results of Table S16 qualify the main text. First, the ratio of estimate to identity lies between 0.22 and 0.65 in every row, so the conclusion that the coefficient is an attenuated identity does not depend on the partition. Second, the cross-group test is decisive only across archetypes. Across regimes the ranges are closer, 1.8 against 1.4, and the ordering is only partly preserved; across sub-periods the coefficient moves more than the identity predicts. In both cases the variation of the coefficient is not the variation the identity would produce, which is the claim the test is designed to check.

Section S7. Results by Method and Their Relation to the Literature

Table S17 states, for each method, what was found and whether it aligns with, qualifies or contradicts the existing evidence. Section 8 of the main text discusses the comparison, and each empirical section closes with a comparison with related studies.

Table S17. Results by method and their relation to the existing literature.

Method	Result obtained	Relation to the literature	Reading	Studies concerned
Panel econometrics	Capitalisation is positively and significantly associated with ROA under all five static estimators (+0.121 to +0.177)	In line	Confirms the sign that is the settled result of the small-firm literature, on a population of unlisted firms it has rarely covered	Matias & Serrasqueiro (2017); Dalci (2018); Quoc Trung (2021); Kalash (2023); Youssef et al. (2023); Gonçalves et al. (2024); Bhawna & Sahay (2025)
	The same coefficient turns to -0.068 when regressors are lagged and to between -0.12 and -0.28 under five instrument sets; the sign is not identified	In opposition	Contradicts the practice of reporting one instrumented column as confirmation; the disagreement between treatments is the result, not a footnote	Canarella & Miller (2018); Vijayakumaran & Vijayakumaran (2019); Samal & Yadav (2025); Dsouza et al. (2025); Saiz-Sepulveda et al. (2026)
	Profitability persists at 0.271 per year, which is why lagged capitalisation cannot instrument	Contrasting	Accepts the persistence finding and turns it against the identification strategy; the same literatures use, a	Hirsch & Hartmann (2014); Yang et al. (2015); Hirsch et al. (2021); Sinha &

	current capitalisation (Hansen J = 77.18)		connection neither strand draws	Vodwal (2022); Choi et al. (2024)
	Coefficient vectors differ across regulatory regimes (Wald 87.18, 181.94, 138.08 on 8 d.f.)	Contrasting	Certification tracks a difference in the structure of profit determination, not only in its average level, which treatment-effect designs cannot detect	Vannoni (2019); Manaresi et al. (2021); Migliaccio & Pavone (2021); Angilella et al. (2023); Aiello et al. (2024); Anderloni & Harasheh (2025)
Accounting identity	$ROA = (VA/A)(1 - s) - D/A$, so $\partial ROA / \partial s = -VA/A$ exactly; four of the eight regressors are subtracted items divided by the aggregates they are subtracted from	In opposition	Cost shares cannot be read as behavioural coefficients in an equation whose dependent variable is built from the same statement, which the operating-side literature does without comment	Campisi et al. (2019); Grau & Reig (2021); Bhattu-Babajee & Seetanah (2022); Chadha et al. (2023); Tripathi et al. (2024)
	Median VA/A is 0.353, so the identity implies -0.355 against an estimate of -0.1365; the lagged coefficient of -0.0346*** would require 0.035	Contrasting	The identity over-predicts the estimate by a factor of two to three, so the coefficient is the identity attenuated rather than the identity, and a quarter of it survives the identity being broken	Grau & Reig (2021); Prakash & Nauriyal (2020); Yousaf (2025)
	Across the four archetypes VA/A ranges by a factor of 2.8 while the coefficient ranges by 1.07, from -0.132 to -0.140	In opposition	Neither the convention that coefficient stability evidences structure nor the objection that it is arithmetic can be settled by assertion; here the identity fixes a magnitude, the magnitude is tested, and the stability is not mechanical	Canarella & Miller (2018); Das et al. (2022); Kalash (2023); Sultana et al. (2026)
Unsupervised clustering	Four operating archetypes emerge from accounting data alone; k-means at k = 4 selected on eleven criteria, ARI 0.9375 over 50 resamples	In line	Reproduces the finding that firms fall into stable financial-operational types, with a selection protocol stricter than the single-index norm	Linares-Mustarós et al. (2018); Bayaraa et al. (2019); Anton et al. (2021); Ljungkvist & Andersén (2021); Juntunen et al. (2022); Kristóf & Virág (2022)

	Silhouette falls monotonically in k for every partitioning method, so silhouette-based selection returns two clusters by construction	In opposition	Challenges the dominant selection practice: the number of groups reported in much of this literature is an artefact of the index chosen	Bayarara et al. (2019); Salina et al. (2020); Bai & Harith (2023); Williams et al. (2025); Vařaničová et al. (2025)
	The turnover coefficient ranges $\times 4.3$ across clusters and the labour share $\times 1.1$; the partition does not coincide with the register (Cramér's $V = 0.478$)	Contrasting	The turnover dispersion is a genuine failure of the pooled vector; the labour-share invariance survives the cross-cluster test and is a property of the relationship, so the partition serves as a diagnostic and not only as a description	Davcik & Grigoriou (2020); Godley et al. (2021); Salles-Filho et al. (2023); Mladenova et al. (2025)
Machine-learning regression	Tree ensembles beat the linear family on the same data (R^2 0.798 against 0.473 on levels); histogram gradient boosting selected on a five-indicator battery	In line	Confirms the accuracy ordering repeatedly reported for accounting data, including the weakness of linear SVM and the instability of AdaBoost	Shetty et al. (2022); Pap et al. (2022); Gavurova et al. (2022); Giudici et al. (2023); Papíková & Papík (2024); Cheraghali & Molnár (2026)
	On firm-demeaned data the gap is 0.121, not 0.325; the ratio falls from 1.69 to 1.29	In opposition	The headline advantage reported in levels comparisons is mostly firmal heterogeneity that fixed effects already absorb, not functional form	Vajjhala & Strang (2024); Mahmood et al. (2025); Sultana et al. (2026); Sengar et al. (2026)
	Permutation importance assigns 63.5% of predictive content to the labour share and 6.5% to capitalisation; H^2 peaks at 0.042	Contrasting	Part of the 63.5% is the accounting relation a learner reproduces almost exactly; the 6.5% is not, since capitalisation is mechanically unrelated to the outcome and is therefore informative by its smallness	Antar & Tayachi (2025); Ariza-Garzón et al. (2024); Balzano & Magrini (2026)

Note. "In line" denotes a result that reproduces the established finding; "contrasting" a result that qualifies or reframes it; "in opposition" a result that contradicts the established finding or the practice that produces it. The second block reports the tests of Section 7, one of which reverses a reading advanced earlier in this paper. Studies listed are those from the reviewed corpus most directly concerned.

Section S8. Policy-Based Identification: Investor Tax Deductions and Registration

Section S3 concludes that identification of the capitalisation coefficient requires exogenous variation in capital structure. The Italian scheme offers one candidate. Investors in certified firms may deduct from personal income tax a statutory share of the equity they contribute, and that share has risen in steps: for innovative start-ups from 19 per cent to 30 per cent in 2017 (Law 232/2016), to 50 per cent in 2020 (Decree-Law 34/2020) and to 65 per cent in 2025 (Law 193/2024); for innovative SMEs, whose incentive required separate authorisation, to 30 per cent in 2017 and to 50 per cent in 2021. Because the rate changes by statute rather than by the firm's choice, it is a natural candidate instrument for the equity ratio, and registration in the special section provides a dated event around which capitalisation and profitability can be traced.

Two designs exploit this variation: an instrumental-variable design in which the deduction rate in force, interacted with registration, instruments the equity ratio, and an event study around registration as an innovative SME. Both require registration dates, which were taken from the business register extractions of 27 July 2026 and matched by tax code to 4,977 of the 5,357 start-ups and 2,752 of the 2,819 innovative SMEs of the estimating sample rebuilt from the source extractions (Appendix A.1). Certified firms without a registration date are excluded, so the estimates rest on 90,597 observations from 14,660 firms and the fixed-effects benchmarks differ slightly from those of Table 2.

Section S8.1. Instrumental Variables with the Deduction Rate

The instrument varies within firms, at registration and at each change of rate, and across regulatory populations within years, so it survives firm and year fixed effects. Registration itself enters as an exogenous control, so that the instrument is identified from the rate alone and not from the fact of certification. Table S18 reports the first stage, the reduced form and the second stage, with the fixed-effects estimate on the same observations as benchmark; the second row sets the innovative-SME rate to zero before 2017, since the dates of that incentive are the less certain, and the last two rows compare each certified population with the conventional SMEs separately.

Table S18. Policy-intensity instrument: statutory investor deduction rate interacted with registration.

Sample	Obs. (firms)	First stage	F	Reduced form, ROA	2SLS, equity ratio	FE, equity ratio
All three populations	90,597 (14,660)	−3.396* (1.933)	3.09	−5.556*** (1.150)	1.636* (0.904)	0.167*** (0.004)
Innovative-SME rate set to zero before 2017	90,597 (14,660)	−3.047* (1.757)	3.01	−4.853*** (1.050)	1.593* (0.882)	0.167*** (0.004)
Innovative and conventional SMEs only	78,441 (9,683)	0.903 (2.274)	0.16	−3.021** (1.261)	−3.345 (8.925)	0.159*** (0.005)
Start-ups and conventional SMEs only	70,925 (11,908)	−9.709*** (3.638)	7.12	−11.112*** (2.372)	1.145*** (0.443)	0.172*** (0.005)

Note. Firm and year fixed effects removed by alternating projections; the seven remaining regressors of Table 2 and a registration indicator enter as exogenous controls. The instrument is the deduction rate in force, expressed as a fraction, multiplied by an indicator equal to one from the year of registration. First stage: coefficient of the instrument in the equation for the equity ratio; reduced form: coefficient of the instrument in the equation for ROA; F: squared cluster-robust *t* statistic of the excluded instrument, which with a single

instrument coincides with the Kleibergen–Paap statistic. Standard errors clustered by firm in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The instrument fails on each count that matters. It is weak: the first-stage F is 3.1 in the pooled sample, 7.1 in the start-up comparison and 0.16 in the innovative-SME comparison, below any conventional threshold. It is wrongly signed: a higher deduction coincides with a lower, not a higher, equity ratio, which indicates that the rate steps capture something other than investor response, since the cohorts registered after 2017 and after 2020 differ from earlier ones and the 2020 step coincides with the pandemic year. And its exclusion restriction cannot be defended: registration bundles the deduction with other benefits, such as simplified access to the public guarantee fund and, for start-ups, exemptions from registration fees and flexible labour rules, all of which can affect profitability directly, while an equity injection raises total assets and so lowers ROA mechanically, through the denominator that the identity of Appendix B makes explicit. The second-stage estimate of 1.64, ten times the fixed-effects benchmark and imprecise, is what a weak and invalid instrument produces, not an estimate of the effect of capitalisation. The event study that follows asks a less ambitious question, what happens to the two variables around registration, and answers it more credibly.

Section S8.2. Event Study around Registration as an Innovative SME

Registration as an innovative SME lends itself to an event study better than registration as a start-up, because start-ups register at or near incorporation and have few accounts before it, whereas innovative SMEs register at a median age of six years and are observed on both sides of the event. The specification regresses the equity ratio and ROA on indicators for the years relative to registration, binned at four years on either side, with the year before registration as reference, firm and year fixed effects, and the conventional SMEs as never-registered controls; start-ups are excluded. Because many innovative SMEs register at about five years of age, a large share are likely to be former start-ups whose pre-registration years are the early years of their life. The specification is therefore repeated on the 697 firms registered after the age of ten, whose pre-registration period is a mature one. Figure S19 plots the coefficients for both samples.

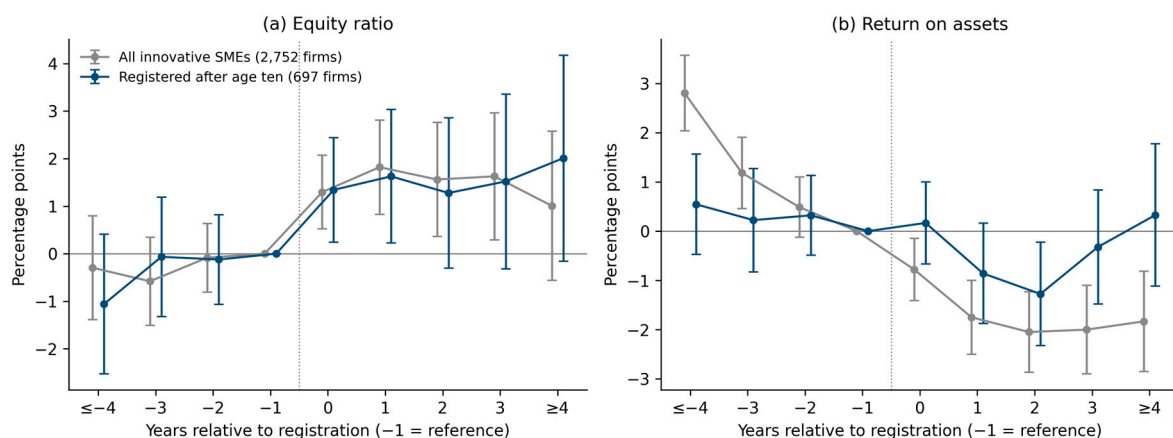


Figure S19. Equity ratio and ROA around registration as an innovative SME. Coefficients with 95 per cent confidence intervals, firm and year fixed effects, conventional SMEs as controls, standard errors clustered by firm.

Panel (a) contains the one clean result of the section. The equity ratio is flat before registration and rises by about 1.3 percentage points in the year of registration and by about 1.5 points on average over the following years, in both samples, with pre-

registration coefficients that are jointly insignificant. Panel (b) shows why the result cannot be carried over to profitability. In the full sample ROA falls steadily for four years before registration, and the pre-registration coefficients are jointly significant, so the parallel-trends condition fails for the outcome of interest. Among firms registered after the age of ten the pre-trend disappears, and ROA falls by about one percentage point in the two years after registration and then recovers. Table S19 reports the coefficients and the joint tests behind both panels.

Table S19. Event-study coefficients around registration as an innovative SME, in percentage points.

Years from registration	Equity ratio, all	ROA, all	Equity ratio, registered after age 10	ROA, registered after age 10
≤ -4	-0.29 (0.56)	2.80*** (0.39)	-1.06 (0.75)	0.55 (0.52)
-3	-0.58 (0.47)	1.18*** (0.37)	-0.07 (0.64)	0.22 (0.54)
-2	-0.09 (0.37)	0.49 (0.31)	-0.12 (0.48)	0.32 (0.41)
-1 (reference)	0	0	0	0
0	1.29*** (0.40)	-0.78** (0.32)	1.34** (0.56)	0.17 (0.42)
1	1.82*** (0.51)	-1.75*** (0.38)	1.63** (0.72)	-0.86* (0.52)
2	1.56** (0.61)	-2.05*** (0.42)	1.28 (0.81)	-1.28** (0.54)
3	1.63** (0.68)	-2.00*** (0.46)	1.52 (0.94)	-0.32 (0.59)
≥ 4	1.01 (0.80)	-1.83*** (0.52)	2.01* (1.11)	0.33 (0.74)
Pre-registration, joint $\chi^2(3)$	1.93 [p = 0.587]	55.95 [p < 0.001]	3.18 [p = 0.365]	1.41 [p = 0.704]
Mean, years 0 to 3	1.58*** (0.47)	-1.65*** (0.32)	1.44** (0.67)	-0.57 (0.43)
Observations (firms)	78,441 (9,683)		65,328 (7,628)	
Registered firms	2,752		697	

Note. Firm and year fixed effects; start-ups excluded; conventional SMEs are never-registered controls. The joint test is the Wald test that the three pre-registration coefficients are zero; the mean is the average of the coefficients for years 0 to 3, with its standard error. Standard errors clustered by firm in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. With never-registered controls and binned endpoints, the heterogeneity bias of two-way fixed-effects event studies under staggered timing is limited, but the estimates are not robust to it by construction.

Section S8.3. What the Policy Variation Establishes

Taken together, the two designs yield one robust finding and one bounded one, and neither identifies the coefficient of the equity ratio in the profitability equation. The robust finding is that registration as an innovative SME is followed by a rise in capitalisation of about one and a half percentage points, without anticipation. The bounded finding concerns profitability among mature registrants: the average change in ROA over the four years from registration is -0.57 percentage points, with a standard error of 0.43. If the fixed-effects coefficient of about 0.17 were causal, a rise of 1.44 points in the equity ratio would raise ROA by about 0.24 points. The estimate points the other way, but its 95 per cent interval, from -1.41 to 0.26, does not exclude that value, and the change cannot in any case be attributed to capitalisation, because registration changes more than the balance sheet. What the event study shows therefore leans against a positive causal effect of capitalisation rather than for it, which is the reading the instrumental-variable estimates of Section S3 also support, but it does not establish it.

The conclusion of Section S3 is extended rather than revised: the one source of policy variation in the Italian scheme does not provide a valid instrument for capital structure in these data. Credible identification would require firm-level records of equity rounds, such as equity-crowdfunding campaigns or venture-capital transactions with their dates and amounts, or a design exploiting the sixty-month limit on start-up status, which ends the status on a date fixed by law rather than chosen by the firm.