

Digital Transformation and Firm Performance in Chinese Listed New-Energy Companies: The Potential Mediating Role of Dynamic Capability

Zhe Zhang

zzhe911205@gmail.com

Liaoning Finance and Trade College

Article

Keywords: digital transformation, firm performance, dynamic capability, new energy

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10790068/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

This article examines whether digital-transformation disclosure is associated with return on equity (ROE) among Chinese listed new-energy companies, and whether a statement-based dynamic-capability index mediates that association. An original public-data panel of 74 CSI New Energy Index firms and 495 firm-years (2018–2025) is constructed from East Money statements and CNINFO annual reports. Commercial databases were not used. Digital transformation is measured as $\ln(1 + \text{dictionary frequency})$ in the full annual report. In two-way firm and year fixed-effects models with firm-clustered standard errors, the digital-transformation coefficient on ROE is 0.0069 (SE 0.0104, $P = 0.506$) and is not distinguishable from zero. A firm-cluster bootstrap of the indirect effect yields -0.0046 (95% interval -0.0079 to 0.0085 , $P = 0.792$). Mediation is not supported. Annual-report digital language is treated as a disclosure proxy, not as implemented technology.

1. Introduction

China's new-energy industry occupies a distinctive position in industrial and climate policy. Photovoltaic manufacturing, wind-power equipment, batteries, energy storage and renewable generation have moved from policy-supported niches to globally competitive, capital-intensive sectors. That expansion has been shaped by renewable-energy policy interacting with industrial policy, including local manufacturing targets, feed-in arrangements and subsequent subsidy reform [1]. Listed new-energy companies have therefore had to absorb policy adjustments, module and cell price collapses, upstream material cycles, and more demanding customers in electric-vehicle and grid-storage markets. In that setting, managers and policymakers frequently present digital technologies—industrial internet platforms, intelligent manufacturing, data-driven operations and digital energy services—as instruments of cost control, quality improvement and new-business development.

Digital transformation is not a homogeneous investment with an automatic effect on profitability. The information systems and strategy literatures define it as a process in which digital technologies trigger strategic responses that change organisations, including their value-creation paths [2, 3, 4]. Systematic reviews emphasise organisational redesign and ecosystem embedding rather than isolated information-technology projects [5, 6]. Distinguishing digital transformation from earlier information-technology programmes is important: the former typically redefines the value proposition and organisational identity, whereas the latter more often supports an existing proposition [7, 8, 9]. Strategy research likewise argues that technology alone does not produce performance; complementary managerial choices, culture and capability building are required [10, 11, 12, 13, 14].

Resource-based and dynamic-capability arguments imply that digital tools create value only when firms can sense opportunities, reconfigure processes and absorb the associated costs [15, 16, 17, 18, 19]. Warner and Wäger [20] treat digital transformation as an ongoing process of strategic renewal in which sensing, seizing and transforming microfoundations must be built rather than purchased. Related work shows that first-order dynamic capabilities shape responses to digital disruption [21] and that even

resource-constrained firms can accumulate digital capabilities through platforms and learning [22]. The implication is theoretical rather than tautological: digital disclosure, digital investment and capability reconfiguration are related constructs, not synonyms.

Empirical work on Chinese listed companies has reported associations between annual-report digital language and capital-market or innovation outcomes [23, 24], and some studies report links to accounting performance or productivity in manufacturing samples [25, 26]. Studies that focus on new-energy or energy firms have typically examined total factor productivity, sustainability scores or green outcomes rather than return on equity, and they have usually relied on commercial databases that independent researchers cannot freely reconstruct [27, 28]. Mixed signs are theoretically admissible. Digital programmes may raise long-run productivity while compressing short-run margins; they may also be described more enthusiastically in reports than they are implemented on the shop floor [29, 30].

Three gaps follow. First, there is still limited public data evidence on whether digital-transformation disclosure is associated with ROE in Chinese listed new-energy companies as a distinct industrial population, rather than as a slice of the all-A-share universe. Second, dynamic capability is often invoked as a mechanism but is rarely constructed from fully documented public indicators with an explicit test of the indirect effect. Third, much of the Chinese digital-transformation literature is difficult to reproduce because keyword counts, sample screens and financials are taken from CSMAR or WIND without releasing the underlying annual-report texts.

This article addresses those gaps with a reconstructable design. The research question is associative: among Chinese listed new-energy companies, is digital-transformation disclosure associated with ROE, and does a publicly constructed dynamic-capability index statistically mediate that association? The study does not claim that keyword frequency measures installed technology, and it does not interpret statistical associations as causal effects. The intended contributions are empirical and methodological. Empirically, the paper reports pooled ordinary least squares, firm and year fixed effects, lagged regressors, an ROA robustness outcome, ownership heterogeneity and an exploratory mediation analysis on a manually verified new-energy sample. Methodologically, it shows that a defensible new-energy panel can be assembled from CSI index constituents, East Money statements, CNINFO annual reports and public company profiles, and it documents every exclusion. The policy implication is modest: if disclosure-based digitalisation is only weakly related to contemporaneous ROE, then treating annual-report digital language as a sufficient statistic for industrial upgrading would be a mistake.

The article is written for a readership that includes strategy, information systems and China-economy scholars. It therefore keeps three commitments visible throughout. The first is conceptual: digital transformation, dynamic capability and accounting performance are adjacent but not interchangeable constructs [2, 19, 7]. The second is empirical: every coefficient is accompanied by a standard error, a confidence interval, an exact P value, a sample size and a cluster count [31, 32]. The third is reproducibility: a reader who cannot license CSMAR or WIND can still rebuild the panel from public index lists, public statements and public annual reports. Those commitments constrain what can be claimed.

They also define the contribution. In a literature that often reports large, multi-star digitalisation effects from unreleased commercial extracts, a reconstructable null on contemporaneous ROE is informative rather than embarrassing, provided it is estimated with documented measurement and without silent deletion of observations.

2. Literature review

2.1 Digital transformation and firm performance

Vial [2] synthesises digital transformation as a process in which digital technologies trigger strategic responses that change organisations. Verhoef and colleagues [3] distinguish digitisation, digitalisation and digital transformation, emphasising business-model change rather than isolated information-technology projects. Bharadwaj and colleagues [4] recast information-technology strategy as digital business strategy fused with corporate strategy, with implications for scope, scale, speed and value capture. Institutional and innovation perspectives add that digital change reconfigures organising logics and may require new organisational forms [8, 9, 33]. Hanelt and colleagues [5] conclude from a large review that digital transformation moves firms toward more malleable designs embedded in digital ecosystems. These conceptual contributions caution against treating a count of nouns in an annual report as equivalent to transformation.

In listed-company research, especially in China, the dominant empirical strategy has been to count digital-technology keywords in annual reports and to take the logarithm of one plus the frequency as a proxy [23]. The proxy is attractive because annual reports are public, comparable across firms and rich in strategic language. It is also vulnerable to cheap talk, boilerplate and differences in report length [24, 33]. Evidence on performance is mixed. Some studies find that digital transformation is associated with higher stock liquidity, innovation, productivity or broader performance metrics [23, 24, 27, 25, 26]. Others emphasise short-run profitability pressure from investment, organisational friction and lagged returns. Guo, Chen and Chen [28] examine listed energy companies and relate digital transformation to sustainability performance under a dynamic-capability framing, but they do not isolate new-energy manufacturers and generators as defined here, nor do they take ROE as the primary outcome. Cheng and Cui [29] argue that Chinese manufacturing firms often combine digital and legacy technologies rather than replacing the latter outright, which is especially relevant to photovoltaic and wind-turbine factories. The implication for this article is that a positive, negative or null association with ROE is all theoretically admissible; the task is to estimate the association in a transparent new-energy sample rather than to assume a sign.

A further strand of work asks whether digital transformation in energy firms improves environmental or sustainability outcomes rather than shareholder returns [28]. That literature is complementary rather than substitutable. Sustainability scores, green patents and emissions intensities answer different questions from ROE. Investors, creditors and managers still use accounting ratios when they evaluate listed new-energy issuers, especially during price wars when losses appear quickly in the income

statement. Keeping ROE as the primary outcome is therefore a design choice, not a claim that environmental performance is unimportant. Likewise, total factor productivity studies [27, 25] speak to efficiency in production, which may rise even while ROE falls if capital is expanded faster than profits. The present article does not attempt to adjudicate among those outcome families. It estimates the association that the title names, with the mediator the title names, in the industry.

2.2 New-energy firms as a distinct empirical setting

Chinese new-energy companies differ from a pooled A-share sample in at least three respects. Their capital expenditure and research intensity are high; their product prices are globally determined and cyclical; and their customers and regulators increasingly demand digital monitoring, forecasting and trading capabilities. Zhang and colleagues [1] show that China's renewable-energy approach has long coupled energy policy with industrial policy, producing rapid manufacturing scale-up alongside later problems of overcapacity and subsidy dependence. Those structural features matter for any digitalisation–performance test, because ROE in this industry is dominated by commodity cycles that a slowly moving disclosure score is unlikely to track.

Ren, Zhang and Chen [27] report that digital transformation is associated with higher total factor productivity among A-share new-energy enterprises, with stronger effects among state-owned firms and firms in eastern China. That study, like most in the field, draws financials from commercial databases. Using CSI New Energy Index constituents as a public sampling frame is a complementary approach: the index is designed to cover renewable production, new-energy application, storage and interaction equipment, and its constituent list is observable [34]. Nuclear generation is sometimes labelled “new energy” in Chinese index products, including CSI 399808. The present operational definition follows the pre-specified categories of solar, wind, hydro, geothermal, renewable equipment, photovoltaics, batteries and storage, and hydrogen. Nuclear specialists are therefore excluded, which is a deliberate departure from the index's own universe and is documented in the sample-selection file.

CSI's own factsheet describes the New Energy Index as covering renewable-energy production, new-energy application, new-energy storage and new-energy interaction equipment, with a launch date of 10 February 2015 and semi-annual rebalancing [34]. Using that list as a frame is more transparent than using an undocumented commercial concept board, but it is not a census of every A-share firm that ever sold a solar module. Firms that left the index before August 2026 are absent. Firms that entered after listing are present only for post-listing years. Those are ordinary index-frame limitations; they are recorded here so that the sample size of 74 firms is not mistaken for the entire Chinese new-energy listed universe. Combining digital and legacy process technologies is common in this manufacturing setting [29], which again counsels against treating dictionary density as a sufficient description of the factory.

2.3 Measuring digital transformation from annual reports

Wu and colleagues [23] group keywords into artificial intelligence, big data, cloud computing, blockchain and digital-technology applications. Subsequent papers have expanded dictionaries or moved from

counts to machine-learning classifiers. Dictionary methods remain the most reproducible public-data option because the word list can be published in full. Two measurement hazards are especially relevant for new-energy issuers. First, the token “5G” is a substring of capacity units such as “15GW”; naive matching inflates digital scores in photovoltaic and wind reports. This study, therefore, applies a regular expression that rejects watt-unit false positives. Second, nested Chinese phrases are counted independently, following the conventional dictionary procedure, which modestly inflates totals and is treated as a limitation rather than silently corrected. The study searches the full annual report rather than only the management discussion and analysis section. Full-text search increases recall of digital language in risk factors, business descriptions and notes, but it also increases the chance of counting industry-level or policy references that do not describe the firm’s own systems. The resulting index is therefore interpreted as a disclosure-based proxy, consistent with the conceptual warning that digital transformation is not identical to information-technology talk [2, 7, 10].

2.4 Dynamic capability as a potential mediator

Teece, Pisano and Shuen [18] define dynamic capabilities as the firm’s ability to integrate, build and reconfigure competencies to address changing environments. Teece [19, 35] elaborates sensing, seizing and transforming microfoundations and later links them to business-model change. Eisenhardt and Martin [36] recast dynamic capabilities as identifiable processes such as product development and alliances. Wang and Ahmed [37] extract absorptive, adaptive and innovative dimensions that have guided many Chinese empirical indices. Reviews stress both the promise and the ambiguity of the construct [38, 39, 40, 41]. Related foundations include absorptive capacity [42, 43], capability lifecycles [44], managerial cognition [45], deliberate learning [46, 47] and knowledge-based views of the firm [48, 49].

In applied work, absorptive capacity is often proxied by research-and-development intensity; adaptive capacity by growth; and innovative capacity by patents or technical staff. Patents and staff counts are not consistently available from the public sources used here. The present index, therefore, uses three statement-based indicators that can be reconstructed: research-and-development intensity, revenue growth and asset turnover. This is a thin proxy. It is not supply-chain resilience, and it shares variance with contemporaneous performance. Those limitations are central to the mediation interpretation. Digital-capability studies that retain richer qualitative microfoundations [20, 22, 21] cannot be replicated with the public files used here, and the paper does not pretend otherwise.

The theoretical literature is not silent on this measurement problem. Winter [47] distinguishes ad hoc problem solving from genuine dynamic capability, which is a patterned and practised ability. Zollo and Winter [46] emphasise deliberate learning routines. Helfat and Peteraf [44, 45] locate capabilities in lifecycles and in managerial cognition. Peteraf, Di Stefano and Verona [40] warn that two conversations about dynamic capabilities have talked past each other. An equal-weight z-score of three accounting ratios cannot adjudicate those debates. It can, at most, show whether a publicly reconstructable index co-moves with digital disclosure and with ROE. That is a much smaller claim than the claim that sensing, seizing and transforming have been observed. The title of this article, therefore, uses the word potential, and the empirical section keeps it there.

2.5 Ownership, institutions and heterogeneity

State-owned and non-state-owned new-energy firms face different financing conditions, political mandates and digitalisation pressures. Several Chinese studies report stronger digitalisation effects among non-state firms in competitive industries, while others find the reverse in energy samples [27, 24]. Ownership is therefore treated as a heterogeneity dimension rather than as a settled moderator. The public actual-controller field from East Money is a snapshot, not a full ownership panel, so the state-owned indicator is approximately time-invariant and is used in split-sample models rather than as a high-frequency control.

2.6 Econometric practice in the adjacent literature

Panel models with firm and year fixed effects and clustered standard errors are now standard [31, 50, 32]. Short panels with slowly moving regressors have limited within-firm variation, and dynamic specifications can introduce incidental-parameter bias [51]. Mediation in this literature still often follows Baron and Kenny [52], despite later demonstrations that a significant total effect is not required and that bootstrap confidence intervals for the indirect effect are preferable [53, 54, 55, 56]. Sobel standard errors [57] assume normality of the product of coefficients, which is unattractive in small clustered panels. Causal mediation methods [58] further require assumptions that observational digital-disclosure data cannot support. This article reports the three-equation system requested by the research design and a firm-cluster bootstrap for the product of coefficients, while classifying the exercise as exploratory.

3. Theoretical framework and hypotheses

The theoretical starting point is the resource-based view: digital technologies, data assets and complementary organisational routines can be valuable, rare and imperfectly imitable, but only as a bundle [15, 16, 17, 48]. Dynamic-capability theory then explains how firms reconfigure that bundle as markets move [18, 19, 36]. In new-energy manufacturing and generation, digital tools may improve yield, forecast demand, schedule maintenance, integrate storage with the grid, and support new service models. Those uses map onto sensing (data and analytics), seizing (investment in platforms and smart manufacturing) and transforming (reconfiguration of plants, supply chains and service offerings) [19, 20, 35].

Two countervailing mechanisms predict a weak or negative short-run link to ROE. First, digital programmes consume cash, managerial attention and complementary human capital before they raise margins [30, 14]. Second, annual-report digital language may rise with industry fashion even when implementation is shallow [10, 33]. The net association with ROE is therefore an empirical question. Dynamic capability is modelled as a potential mediator rather than as a synonym of digital transformation. Digital disclosure may be associated with higher research intensity, faster revenue adjustment and more intensive asset use; those capability indicators may in turn be associated with ROE. Because the capability index is built from financial ratios that are close to the performance identity,

any indirect effect may partly be mechanical. The paper, therefore, refuses to treat a statistically non-zero product of coefficients as confirmed mediation.

The primary research question is associative. Among Chinese listed new-energy companies, is digital-transformation disclosure associated with ROE after firm and year effects and standard controls? A secondary question asks whether there is an indirect association running from digital-transformation disclosure through a statement-based dynamic-capability index to ROE. Directional hypotheses are stated because the dominant theoretical narrative is positive, but two-tailed tests are used throughout.

H1. Digital-transformation disclosure is positively associated with ROE. H2. Digital-transformation disclosure is positively associated with the dynamic-capability index. H3. The dynamic-capability index is positively associated with ROE after controlling for digital-transformation disclosure. H4. The indirect association ($H2 \times H3$) is positive. H5. The association in H1 differs between state-owned and non-state-owned firms. H1 is also examined with ROA, lagged digital transformation and an alternative dictionary.

These hypotheses are directional because the resource-based and dynamic-capability narratives are directional [15, 18, 20]. They are tested two-tailed because the short-run accounting channel can run the other way, and because disclosure can rise without implementation [10, 33]. Rejecting H1 at $\alpha = 0.05$ would not, on its own, establish that digital systems reduce profits. Failing to reject H1 would not establish that digital systems are irrelevant. The hypotheses organise the tables; they do not license causal slogans. H4 is the mediation claim implied by the title. It requires a non-zero indirect effect whose bootstrap interval excludes zero [53, 55, 56]. A large path b without a non-zero path a does not constitute mediation. That distinction is elementary in the methods literature and is often blurred in applied digitalisation papers that report a mediator coefficient distinguishable from zero and then stop [52, 54].

4. Methods

4.1 Study design

The design is an observational, unbalanced firm-year panel for fiscal years 2018–2025. The unit of analysis is the listed company-year. Estimation is associative. No instrumental variable, regression discontinuity or difference-in-differences design is available from the public sources, so causal language is avoided [50, 58]. Generative artificial intelligence was used to assist coding, literature organisation and drafting. The author verified source access, sample inclusion, statistical output and citations, and accepts full responsibility. The AI system is not an author. No generative-AI imagery is used in figures.

The observational design has consequences that should be stated before the estimates are shown. Managers choose digital disclosure. Firms that are already more profitable, more research-intensive or more exposed to industrial-internet policy may write different annual reports. Firm fixed effects absorb time-invariant confounders, including persistent quality, location and product mix [50]. They do not absorb time-varying omitted variables, such as a sudden capacity expansion that both reduces ROE and prompts a digitalisation paragraph. Lagged digital transformation reduces simultaneous reporting bias

but cannot capture implementation lags of several years [51]. No claim is made that the coefficient on digital transformation is an average treatment effect. The quantity reported is a partial association under an additive linear specification with clustered errors [31, 32]. Readers who require identification strategies based on broadband roll-out, subsidy discontinuities or instrumental variables will not find them here, because those designs cannot be executed honestly with the public files in this package.

4.2 Sample and data sources The submission package includes the company list, sample-selection log, variable/codebook documentation, keyword dictionary, cleaned firm-year files and analysis scripts needed to reconstruct the reported estimates.

No CSMAR, WIND, RESSET or iFinD extract was used. The sampling frame is the CSI New Energy Index (code 399808) constituent list published on Sina Finance, snapshot 28 August 2026 (80 firms) [34]. Company names, industries, listing dates, business scope, actual controllers and websites were taken from East Money F10 company survey and shareholder interfaces and from CNINFO profile functions. Annual financial statements were taken from East Money. Annual-report PDFs were inventoried and downloaded from CNINFO. The operational definition of a new-energy company is a Shanghai or Shenzhen A-share issuer whose principal disclosed business is solar or photovoltaic manufacturing or equipment, wind generation or equipment, hydropower or other renewable generation, geothermal energy, batteries or energy storage, battery materials, hydrogen, or renewable-energy power electronics. New-energy vehicles are not used as a standalone inclusion rule.

From 80 index constituents, six firms were excluded: China National Nuclear Power (601985) and CGN Power (003816) as nuclear specialists; China Baoan (000009) as a conglomerate; the State Power Investment industrial-finance vehicle (000958); Kehua Data (002335), whose principal disclosed businesses are intelligent computing and data-centre power; and the Beijing Stock Exchange issuer 920185, for which CNINFO and East Money annual-report coverage was incomplete. Seventy-four firms were retained. Firm-years before the listing year are omitted from the analysis models. No observation is deleted silently. Survivorship bias is acknowledged: the frame is the current index, not the historical union of all firms that were ever constituents.

4.3 Variable measurement

ROE is net profit attributable to the parent divided by ending equity attributable to the parent. ROA is the same numerator divided by ending total assets. Both use East Money consolidated annual (31 December) statements. Negative equity is flagged; such ROE values are retained in the clean file and winsorised in robustness models.

The primary digital-transformation measure follows the conventional compression of a skewed keyword count [23]:

$$\text{Digital Transformation}_{it} = \ln(1 + \text{KeywordFrequency}_{it}) \quad (1)$$

KeywordFrequency_{it} is the sum of dictionary hits in the full CNINFO annual report for firm *i* in year *t*: a zero frequency, which is observed, maps to a digital-transformation score of zero. Chinese terms are counted as substrings. English abbreviations use regular expressions. The token 5G is matched only when it is not adjacent to digits or watt letters, so that capacity figures such as 15GW are not counted as fifth-generation mobile networks. Reports labelled as summaries, English versions or H-share versions are discarded; where a revision exists, the later filing is kept. Missing reports produce missing digital-transformation scores, not zeros. A robustness measure divides keyword frequency by the count of Chinese characters in the extracted text. A second robustness measure adds an extension dictionary. Keyword frequency is not treated as a measure of implemented technology.

Dynamic capability is the equal-weight mean of the sample z-scores of research-and-development intensity (research expense divided by operating revenue), year-on-year revenue growth, and asset turnover (operating revenue divided by total assets), after 1 per cent and 99 per cent winsorisation of each input. The index is missing if any component is missing. No principal-component rotation is imposed, so weights remain transparent. The index is not a supply-chain-resilience score. Controls are firm size (logarithm of total assets), listing age, leverage (liabilities divided by assets) and current-ratio liquidity. Revenue growth and asset turnover are not used as regression controls because they enter the capability index. State-owned status is used in split samples only.

4.4 Statistical analysis

All tests are two-tailed with $\alpha = 0.05$. Standard errors are clustered by firm [31, 32]. Models include pooled ordinary least squares; year fixed effects; firm fixed effects; two-way firm and year fixed effects; winsorised two-way models; one-year lagged digital transformation; ROA as outcome; state-owned and non-state splits; variance-inflation factors; missingness summaries; and alternative digital-transformation measures. Two-way dummy specifications in a short panel can be rank-deficient; the two-way estimate is therefore reported as an imprecise within association [50, 51].

Mediation estimates the conventional three-equation system [52, 54, 55]. The indirect effect is the product of the digital-transformation coefficient in the capability equation and the capability coefficient in the ROE equation. A firm-cluster bootstrap with 1,000 replications of a year-fixed-effects specification supplies a percentile confidence interval [53, 56], because two-way fixed effects are unstable in small bootstrap draws. Statistical significance is not treated as proof of causality. Multiple comparisons across specifications are discussed as a family of estimates rather than as independent discoveries.

Several distributional and dependence issues were examined rather than assumed away. ROE is left-skewed because of loss-making years in the 2023–2025 photovoltaic downturn; the winsorised digital-transformation robustness model does not change the full-sample inference. Residuals are heteroskedastic across firms of different scales, which is why every reported standard error is clustered by firm rather than computed by ordinary least squares [31, 32]. Within-firm serial correlation is the same rationale: a firm that writes a digitalisation paragraph in 2021 tends to keep similar language in 2022. Year dummies absorb the common 2022–2024 price shock only to the extent that it is parallel across

issuers. It is not fully parallel. Integrated module makers and pure-play lithium chemical firms were hit differently, so remaining heterogeneity is a limitation of two-way additive effects [50]. The rule constrained control variables so that no regressor should mechanically duplicate the dependent variable or the mediator. Size, listing age, leverage and liquidity satisfy that rule. Revenue growth and asset turnover do not, once they enter dynamic capability, so they are omitted from the ROE regressions as controls. Cash holdings, Tobin's Q and board size were considered because they appear in adjacent papers, but a complete public panel of those series was not assembled with the same annual-statement discipline as ROE, and they are therefore not inserted as if they had been. Where a popular control cannot be built from the public files, the correct statement is that it is missing.

Annual-report processing followed a fixed rule set. CNINFO's announcement interface was queried for each stock code for filing dates from 1 January 2018 to 30 August 2026. Each PDF was downloaded, parsed, counted and discarded; the package retains the URL, byte size, page count, character count and keyword vector, not the PDF itself. Two downloads failed on the first pass and were successfully retried; after the retry, 502 of 502 selected reports for the 74 firms were extracted successfully. A worked measurement check on a large photovoltaic issuer's report showed that a naive search for the token 5G produced hits that were, in fact, gigawatt capacity figures. The production regular expression rejects those false positives.

5. Results

5.1 Sample characteristics

The CSI New Energy Index supplied 80 candidate issuers on 28 August 2026 [34]. Six firms were excluded after business-scope reading, as listed in Table 1 and summarised in Fig. 1. The retained sample comprises 74 Shanghai and Shenzhen A-share companies whose principal disclosed activities are photovoltaic manufacturing or equipment (21 firms), batteries and energy storage (20), renewable generation (10), lithium and battery materials (10), wind equipment (5) and related power electronics or mixed equipment (8). The panel is unbalanced because several STAR Market and ChiNext issuers were listed after 2018. After dropping pre-listing years, the analysis file contains 495 firm-years from 2018 to 2025 (45 observations in 2018, rising to 74 in 2025). Digital-transformation scores are non-missing for 494 of those 495 firm-years. Dynamic capability is non-missing for 446 firm-years, the gap arising mainly from first-year revenue growth and occasional missing research-and-development expense. Duplicate firm-years were not present after keeping the last 31 December statement for each issuer-year. Negative equity and extreme ROE values were flagged rather than dropped.

Table 1
Sample selection from the CSI New Energy Index (399808)
snapshot.

Stage	N
CSI 399808 constituents (Sina snapshot, 28 August 2026)	80
Exclude nuclear generators (601985, 003816)	2
Exclude conglomerate (000009)	1
Exclude industrial-finance vehicle (000958)	1
Exclude data-centre power (002335)	1
Exclude BSE listing with incomplete coverage (920185)	1
Retained SSE/SZSE A-share new-energy firms	74
Firm-years 2018–2025 after dropping pre-listing years	495
Firm-years with non-missing digital transformation	494
Firm-years with non-missing dynamic capability	446

5.2 Descriptive statistics and correlations

Table 2 reports descriptive statistics. Mean ROE is 0.091 (SD 0.156; $n = 495$). Mean ROA is 0.045 (SD 0.056). Mean digital transformation, $\ln(1 + \text{dictionary frequency})$, is 1.504 (SD 0.941), corresponding to a typical annual-report frequency of about three to four dictionary hits. The first percentile of digital transformation is 0, and the 99th percentile is 3.64, so a handful of reports are dense in digital language while many remain sparse. Mean dynamic capability is near zero by construction (-0.037 , SD 0.536, $n = 446$). Mean log assets is 24.02, mean listing age is 9.28 years, mean leverage is 0.549 and mean current ratio is 1.63. Mean research-and-development intensity is 0.039. About 22 per cent of firm-years with a classifiable controller are state-owned ($n = 457$). Variance inflation factors for digital transformation, size, listing age, leverage and liquidity are all below 2.2, so collinearity among the regressors used in the baseline models is not a practical concern. Figure 2 displays the distributions of digital transformation and ROE. Figure 3 plots the annual means of both series.

Table 2
Descriptive statistics for the analysis sample (firm-years 2018–2025).

Variable	N	Mean	SD	Min	Median	Max	Missing
roe	495	0.091	0.156	-1.497	0.100	0.562	0
roa	495	0.045	0.056	-0.213	0.038	0.366	0
dt	494	1.504	0.941	0.000	1.609	3.970	1
dt_combined	494	2.309	1.082	0.000	2.398	4.997	1
dc	446	-0.037	0.536	-1.320	0.0000	1.763	49
size	495	24.015	1.179	20.849	23.976	27.606	0
list_age	495	9.285	7.151	0.000	8.000	32.000	0
lev	495	0.549	0.156	0.086	0.579	0.938	0
liquidity	495	1.631	1.164	0.106	1.353	14.401	0
rd_intensity	489	0.039	0.030	0.0000	0.037	0.332	6
rev_growth	450	0.293	0.550	-0.825	0.194	4.278	45
tat	495	0.499	0.257	0.043	0.474	1.869	0
soe	457	0.221	0.415	0.000	0.000	1.000	38

Table 3 reports pairwise correlations. The correlation between digital transformation and ROE is – 0.062. The correlation between digital transformation and ROA is – 0.132. Dynamic capability correlates 0.372 with ROE and 0.469 with ROA, which is expected because the index contains revenue growth and asset turnover. Digital transformation correlates – 0.084 with dynamic capability. Leverage correlates – 0.279 with ROE. These associations are descriptive; they do not hold other variables fixed, and they do not establish mediation.

Table 3
Pairwise correlations. DT is digital transformation; DC is dynamic capability.

Variable	ROE	ROA	DT	DC	Size	List age	Lev	Liquidity
roe	1.000	0.808	-0.062	0.372	-0.060	-0.130	-0.279	0.147
roa	0.808	1.000	-0.132	0.469	-0.198	-0.142	-0.471	0.317
dt	-0.062	-0.132	1.000	-0.084	0.184	0.133	0.192	-0.141
dt_combined	-0.033	-0.086	0.765	0.123	0.066	0.083	0.105	-0.075
dc	0.372	0.469	-0.084	1.000	-0.358	-0.227	-0.141	0.074
size	-0.060	-0.198	0.184	-0.358	1.000	0.288	0.456	-0.250
list_age	-0.130	-0.142	0.133	-0.227	0.288	1.000	0.097	-0.064
lev	-0.279	-0.471	0.192	-0.141	0.456	0.097	1.000	-0.658
liquidity	0.147	0.317	-0.141	0.074	-0.250	-0.064	-0.658	1.000
rd_intensity	0.029	0.020	0.022	0.515	-0.339	-0.203	-0.106	0.035
rev_growth	0.436	0.557	-0.118	0.638	-0.100	-0.170	-0.056	-0.013
tat	0.229	0.312	-0.048	0.652	-0.172	-0.065	-0.126	0.118
soe	-0.096	-0.220	-0.0086	-0.375	0.219	0.246	0.192	-0.098

5.3 Baseline results

Table 4 reports the digital-transformation coefficient across specifications, and Fig. 4 plots the corresponding confidence intervals. In pooled ordinary least squares with firm-clustered standard errors, the coefficient is -0.0011 (SE 0.0074 , 95% CI -0.0156 to 0.0134 , $t = -0.15$, $P = 0.881$, $n = 494$, 74 firms, $R^2 = 0.104$). The estimate is not statistically distinguishable from zero at $\alpha = 0.05$, two-tailed.

Table 4

Digital-transformation coefficients. The dependent variable is ROE except for M9 (ROA). Standard errors are clustered by firm. Two-tailed tests, alpha = 0.05.

Model	Coef.	SE	P	CI low	CI high	N	Firms	R2
M1 Pooled OLS (ROE)	-0.0011	0.0074	0.8810	-0.016	0.013	494	74	0.104
M2 Year FE (ROE)	0.0048	0.0071	0.5003	-0.0092	0.019	494	74	0.193
M3 Firm FE (ROE)	-0.0009	0.011	0.9357	-0.023	0.021	494	74	0.375
M4 Two-way FE (ROE)	0.0069	0.010	0.5063	-0.013	0.027	494	74	0.418
M5 Winsorised DT, two-way FE	0.0072	0.010	0.4928	-0.013	0.028	494	74	0.418
M6 Lagged DT, year FE	0.0000	0.0075	0.9995	-0.015	0.015	428	72	0.210
M7 DT share, year FE	137.007	91.206	0.1331	-41.753	315.767	494	74	0.195
M8 Combined dictionary, year FE	0.0051	0.0052	0.3296	-0.0051	0.015	494	74	0.194
M9 Two-way FE (ROA)	0.0007	0.0039	0.8585	-0.0069	0.0083	494	74	0.521
M10 SOE, year FE	-0.028	0.011	0.0099	-0.050	-0.0068	101	16	0.419
M10 Non-SOE, year FE	0.0091	0.0064	0.1505	-0.0033	0.022	355	52	0.231

With year fixed effects, the coefficient is 0.0048 (SE 0.0071, 95% CI – 0.0092 to 0.0188, $t = 0.67$, $P = 0.500$, $n = 494$, 74 firms, $R^2 = 0.193$). Size is positively associated with ROE (coefficient 0.0226, SE 0.0113, $P = 0.045$), and leverage is negatively associated with ROE (coefficient – 0.375, SE 0.121, $P = 0.002$). Listing age is negatively associated with ROE (coefficient – 0.0026, SE 0.0013, $P = 0.048$). Liquidity is not distinguishable from zero ($P = 0.164$).

Firm fixed effects without year effects yield a coefficient of – 0.0009 (SE 0.0111, 95% CI – 0.0226 to 0.0208, $t = -0.08$, $P = 0.936$, $n = 494$, 74 firms, $R^2 = 0.375$). Two-way firm and year fixed effects yield 0.0069 (SE 0.0104, 95% CI – 0.0135 to 0.0273, $t = 0.66$, $P = 0.506$, $n = 494$, 74 firms, $R^2 = 0.418$). Rank-deficiency warnings appeared when firm and year dummies were entered jointly in a short panel; the two-way point estimate should therefore be read as a within-firm, within-year association that is imprecisely estimated, not as a fully saturated experimental contrast [50]. Taken together, H1 is not supported in the full sample. Digital-transformation disclosure is not statistically distinguishable from zero in its association with ROE once conventional controls and clustered errors are used.

The jump in R^2 from 0.19 in the year-fixed model to 0.42 in the two-way model shows that firm heterogeneity dominates year shocks. That pattern is consistent with persistent differences in technology, location and product mix among photovoltaic, battery and generation firms. It is also why a short within-firm contrast on digital language has little power: many issuers change their keyword counts slowly, and product prices and capacity cycles dominate ROE in this industry, so that year effects only partly absorb [1].

5.4 Robustness tests and alternative outcome

Table 5 reports robustness checks and the alternative ROA outcome. Winsorising digital transformation at the 1st and 99th percentiles leaves the two-way coefficient at 0.0072 (SE 0.0104, $P = 0.493$, $n = 494$). One-year lagged digital transformation in a year-fixed model is 0.000005 (SE 0.0075, $P = 0.999$, $n = 428$, 72 firms). The length-adjusted measure (keyword frequency divided by Chinese character count) has a large coefficient because the regressor is scaled in thousandths, but it is not distinguishable from zero (coefficient 137.0, SE 91.2, $P = 0.133$, $n = 494$). The combined dictionary produces 0.0051 (SE 0.0052, $P = 0.330$, $n = 494$). None of these robustness models reverses the full-sample conclusion. Replacing ROE with ROA in the two-way specification yields a coefficient of 0.0007 (SE 0.0039, 95% CI – 0.0069 to 0.0083, $t = 0.18$, $P = 0.859$, $n = 494$, 74 firms, $R^2 = 0.521$). H1 is not supported for ROA either.

Table 5
Robustness and alternative outcome. Extract of Table 4, models M5–M9.

Model	Coef.	SE	P	CI low	CI high	N	Firms	R2
M5 Winsorised DT, two-way FE	0.0072	0.010	0.4928	-0.013	0.028	494	74	0.418
M6 Lagged DT, year FE	0.0000	0.0075	0.9995	-0.015	0.015	428	72	0.210
M7 DT share, year FE	137.007	91.206	0.1331	-41.753	315.767	494	74	0.195
M8 Combined dictionary, year FE	0.0051	0.0052	0.3296	-0.0051	0.015	494	74	0.194
M9 Two-way FE (ROA)	0.0007	0.0039	0.8585	-0.0069	0.0083	494	74	0.521

5.5 Mediation analysis

Table 6 and Fig. 5 report the exploratory mediation estimates. On the 445 firm-years with non-missing dynamic capability (73 firms), path a from digital transformation to dynamic capability in the two-way model is – 0.0210 (SE 0.0319, 95% CI – 0.0836 to 0.0416, $t = -0.66$, $P = 0.510$). H2 is not supported. Path b from dynamic capability to ROE, controlling for digital transformation, is 0.2189 (SE 0.0357, 95% CI 0.1489 to 0.2888, $t = 6.14$, $P = 8.47 \times 10^{-10}$). That coefficient is statistically distinguishable from zero, but it is not surprising: the index includes revenue growth and asset turnover, which enter profitability identities. The total effect of digital transformation on ROE in this subsample is 0.0084 (SE 0.0113, $P =$

0.457). The direct effect is 0.0130 (SE 0.0091, $P = 0.154$). The product $a \times b$ is -0.0046 . A firm-cluster bootstrap with 1,000 replications of a year-fixed-effects specification gives a 95% percentile interval of -0.0079 to 0.0085 (two-tailed bootstrap $P = 0.792$) [53, 55]. The interval includes zero. H3 is supported only in the narrow sense that the capability index co-moves with ROE. H4 is not supported. The mediation analysis is exploratory and does not confirm a capability channel from digital disclosure to ROE [54, 58].

Table 6
Exploratory mediation. Indirect effect $a \times b$ with firm-cluster bootstrap (1,000 replications).

Quantity	Estimate
Path a (DT $\rightarrow$ DC)	-0.021
Path b (DC $\rightarrow$ ROE DT)	0.219
Total effect c (DT $\rightarrow$ ROE)	0.0084
Direct effect c'	0.013
Indirect effect $a \times b$	-0.0046
Bootstrap 95% CI low	-0.0079
Bootstrap 95% CI high	0.0085
Bootstrap two-tailed P	0.792
N (firm-years)	445.000
N (firms)	73.000

5.6 Ownership heterogeneity and missing data

Table 7 reports ownership splits. Among 16 state-owned firms (101 firm-years), the year-fixed digital-transformation coefficient is -0.0284 (SE 0.0110, 95% CI -0.0499 to -0.0068 , $t = -2.58$, $P = 0.0099$). Among 52 non-state firms (355 firm-years), the coefficient is 0.0091 (SE 0.0064, 95% CI -0.0033 to 0.0216 , $t = 1.44$, $P = 0.150$). H5 is therefore mixed: the state-owned split is statistically distinguishable from zero and negative, while the non-state split is not distinguishable from zero. With only 16 state-owned clusters, the negative estimate is sensitive to a small number of issuers and to the keyword rule used to classify controllers [32]. It is reported as a heterogeneity check, not as a confirmed difference in treatment effects. Six firms could not be classified and are omitted from the splits.

Table 7
Ownership splits (year fixed effects). The SOE estimate uses 16 firms and is exploratory.

Model	Coef.	SE	P	CI low	CI high	N	Firms	R2
M10 SOE, year FE	-0.028	0.011	0.0099	-0.050	-0.0068	101	16	0.419
M10 Non-SOE, year FE	0.0091	0.0064	0.1505	-0.0033	0.022	355	52	0.231

Digital transformation is missing for one analysis firm-year (0.2%). Dynamic capability is missing for 49 firm-years (9.9%), almost entirely because growth is undefined in a firm's first observed year. Research-and-development intensity is missing for six firm-years. State ownership is missing for 38 firm-years. The analysis never recodes those missings to zero. Complete-case mediation, therefore, uses 445 observations, which is disclosed in every mediation statistic.

The empirical pattern is a precisely estimated near-zero full-sample association between annual-report digital language and ROE, a mechanical association between the statement-based capability index and ROE, and no bootstrap evidence of an indirect effect. The only full-sample coefficients that are distinguishable from zero at 0.05 in the year-fixed ROE model are size, listing age and leverage. Those results are consistent with a cyclical, capital-intensive industry in which capital structure and scale matter more for contemporaneous accounting returns than the density of digital nouns in the annual report.

The hypothesis tests can be restated in the language required for this journal. H1 asked whether digital-transformation disclosure is positively associated with ROE. Using two-tailed tests and $\alpha = 0.05$, the full-sample estimates are not statistically distinguishable from zero. The smallest full-sample P value among pooled, year-fixed, firm-fixed and two-way models is 0.500. H2 asked whether disclosure is positively associated with the dynamic-capability index; path a has $P = 0.510$ and is not distinguishable from zero. H3 asked whether the index is positively associated with ROE after controlling for disclosure; path b has $P = 8.47 \times 10^{-10}$ and is distinguishable from zero, subject to the mechanical-construction caveat. H4 asked whether the indirect effect is positive; the bootstrap 95% interval includes zero, and the two-tailed bootstrap P is 0.792, so H4 is not supported [53, 54]. H5 asked whether the H1 association differs by ownership; the state-owned coefficient is distinguishable from zero and negative ($P = 0.0099$, 16 firms), while the non-state coefficient is not ($P = 0.150$, 52 firms). Because the state-owned test uses a small number of clusters and because fourteen digital-transformation coefficients were estimated, H5 is not treated as confirmatory [32]. The overall inferential stance is therefore that the potential mediating role named in the title remains potential: it is specified, estimated and not confirmed in these public data.

6. Discussion

The estimates should be read as associations in a public-data new-energy panel, not as returns to digital investment. Digital-transformation scores are disclosure intensities: firms may increase keyword use because peers, analysts or ministries reward digital language [23, 33]. A non-zero coefficient would still

mix implementation with rhetoric, while a coefficient indistinguishable from zero shows only that annual-report language, as measured here, does not line up with contemporaneous ROE [2, 7].

Two-way fixed effects absorb slow-moving strategy, but a short panel provides limited within-firm variation [51, 50]. Pooled and year-fixed models may therefore reflect unobserved quality. The M4 95% interval of -0.0135 to 0.0273 is the appropriate statement of precision. The paper does not claim a precise zero; it reports an association that is not distinguishable from zero at $\alpha = 0.05$ in these public data.

The dynamic-capability index uses research intensity, growth and turnover, which are close to the ROE identity through margins, asset use and growth. An indirect effect can therefore be partly mechanical. The mediation analysis is consequently labelled exploratory [39, 38, 54]. Stronger tests would require measures not derived from the performance identity, such as plant-level yield, forecast error, or independently audited digital-asset stocks [19, 20].

Industry context also cautions against treating digital disclosure as a sufficient performance statistic. From 2018 to 2021, many photovoltaic and battery issuers expanded capacity under supportive prices; from late 2022, polysilicon, wafer, cell and module prices fell sharply, and several large firms recorded losses in 2024–2025 despite continued discussion of intelligent manufacturing [1]. The coefficient therefore averages across sub-cycles. Subcategory samples are too small for precise three-way interactions, so no underpowered subgroup analysis is reported.

Multiple comparisons are material. Table 4 contains fourteen digital-transformation coefficients, so some 0.05-level findings can occur by chance. The single full-sample rejection is the state-owned split, a pre-specified heterogeneity check based on only 16 clusters [32]. It should therefore not be promoted to a headline causal claim.

6.1 Theoretical implications

For digital-transformation research, the results underline the warning that digital change is a process, not a count of nouns in a report [2, 5, 7]. For dynamic-capability research, they show the cost of operationalising Teece's microfoundations with public financial ratios: the construct becomes estimable, and the mediation claim becomes fragile [19, 40]. For industry studies of Chinese new energy, they treat the sector as a theoretically motivated population rather than as a dummy variable in an all-A-share regression. The resource-based view remains useful: digital keywords without complementary capability are unlikely to be rare or costly to imitate [15, 17]. The empirical contribution is to keep that argument testable with files that a reader can rebuild.

6.2 Practical implications

Managers in photovoltaic, wind, battery and renewable-generation firms should not infer that publishing more digital vocabulary will raise ROE. The measurement design instead suggests separating systems actually deployed, skills and process change, and external reporting [10, 12]. Policymakers using listed-

company text as a digitalisation dashboard should complement it with equipment, patent and energy-system indicators, while investors should treat keyword intensity as a communication choice.

For researchers, the practical implication is reproducibility. Replicators can start from the company list, use the PDF inventory, apply the published dictionary, rebuild ROE from the stored statements, and rerun the analysis. The new-energy setting is substantively important for China's energy transition [1] and exposes the gigawatt-versus-5G measurement trap. Until plant-level digital investment can be linked to audited financials, or implementation is hand-coded rather than inferred from vocabulary, the results do not support treating annual-report keyword density as a performance measure [19, 15].

6.3 Limitations

The sampling frame is a current index snapshot and is therefore subject to survival and index-inclusion rules [34]. Seventy-four firms over eight years provide a small panel for two-way fixed effects. Financials come from East Money's public interface, so restatements may be incomplete. Digital transformation is a dictionary applied to PDF text; image-only pages, unusual fonts and extraction errors can cause missing or under-counted values. Dynamic capability omits patents, technical staff and process measures that commercial databases may contain. State ownership is classified from the latest controller field, no credible instrumental variable is available, and multiple specifications increase the chance of an isolated significant result. English-language and H-share reports are dropped. The period also includes COVID-19, the 2021–2022 energy-price spike and the 2023–2024 photovoltaic price war; year effects do not remove heterogeneous exposure. Hydropower and geothermal firms are thin in the frame, hydrogen is mainly secondary activity, and new-energy vehicle assemblers were not an inclusion rule.

Annual-report measurement error also matters. Text extraction does not perform optical character recognition on image-only pages, so digital content in figures may be missed; repeated slogans in headers may be counted more than once. Such errors can attenuate the digital-transformation coefficient, while the 5G/GW correction prevents capacity figures from being misclassified as fifth-generation mobile activity [50]. These facts reinforce that an indistinguishable coefficient is not evidence that digital systems have no productivity role in new energy [2, 23].

External validity is bounded by listing venue and index membership. Beijing Stock Exchange issuer 920185 is excluded because coverage was incomplete; dual-listed issuers are represented by their A-share Chinese reports; and firms outside CSI 399808 are outside the frame [34]. Statistical power is also limited: with 74 clusters and a slowly moving disclosure score, the study is not well placed to detect small positive ROE effects [32, 51]. Intervals are therefore reported alongside P values, and the paper avoids a rhetoric of precise zeros.

The public-data package is sufficient to estimate this short new-energy panel but cannot reproduce every commercial-data study. Patent stocks, research-staff ratios, digital intangible-asset breakdowns and historical controller panels may remain behind paywalls. The study leaves such variables missing rather

than imputing them and narrows its claims accordingly. The bootstrap interval for the indirect effect includes zero; it is reported transparently rather than altered to produce a confirmed mediation result [53, 55, 54].

7. Conclusion

This article assembled a public-data panel of Chinese listed new-energy companies and asked whether digital-transformation disclosure is associated with ROE and whether a statement-based dynamic-capability index mediates that association. Using 74 firms and 495 firm-years from 2018 to 2025, the full-sample association is not statistically distinguishable from zero (two-way fixed-effects coefficient 0.0069, SE 0.0104, $P = 0.506$). The bootstrap interval for the indirect effect includes zero ($P = 0.792$). H1, H2 and H4 are not supported; H3 reflects mechanical co-movement between the capability index and ROE, and H5 is mixed and exploratory. The contribution is the reconstructable sample, documented dictionary, gigawatt false-positive correction, and conservative mediation stance.

The public-data design answers the research question without over-claiming. Digital-transformation disclosure constructed from CNINFO annual reports is not statistically distinguishable from zero in its full-sample association with ROE, and the statement-based capability index has no bootstrap-supported indirect effect. The numerical results are accompanied by documentation of how the panel and measures were constructed. Future work should link plant-level digital investment to audited financials or hand-code implementation rather than infer it from vocabulary [20, 7].

The study period ends in 2025, so the latest annual reports are included, but long-run returns cannot be assessed fully. Digital programmes announced in 2023 may not affect 2025 ROE even if they later raise yields. The one-year lagged specification is also indistinguishable from zero ($P = 0.999$), but longer lags remain an open question [50]. Plant-level commissioning dates, MES logs or independently audited digital-asset stocks would permit stronger tests.

Nothing in the estimates licenses a policy that subsidises digital-keyword density or dismisses shop-floor digitalisation as irrelevant. Text-based dashboards should not be treated as performance scorecards [10, 12, 2]. Performance in this industry remains linked to prices, utilisation, leverage and scale, while the public ratios used here cannot isolate the underlying capability mechanisms [19, 1].

This interpretation is consistent with the resource-based and dynamic-capability views, which locate advantage in bundles and processes rather than vocabulary [15, 18, 36], and with the information-systems distinction between discussing digital technology and redefining a value proposition [7, 4].

Declarations

Competing interests

The author declares no competing interests.

Ethical Approval

This article does not contain any studies with human participants performed by any of the authors. Ethical approval was not required because the study analysed publicly available secondary firm-level data and did not involve human participants or identifiable personal data.

Informed Consent

This article does not contain any studies with human participants. Informed consent was therefore not applicable.

Author Contribution

Z.Z. conceived and designed the study, collected and analyzed the data, interpreted the results, and wrote the main manuscript text. Z.Z. reviewed and approved the final manuscript.

Data Availability

The analysis uses publicly accessible secondary data from the CSI New Energy Index (399808), East Money company and financial-statement information, and CNINFO annual-report and company-disclosure records. The cleaned firm-year dataset, variable/codebook documentation, digital-transformation keyword dictionary, company list, and sample-selection log are provided with the submission for peer review. The upstream annual reports and company disclosures remain publicly accessible through CNINFO (<https://www.cninfo.com.cn/>), and financial/company information is publicly accessible through East Money (<https://www.eastmoney.com/>). CSI New Energy Index (399808) constituent information is publicly accessible through Sina Finance (https://vip.stock.finance.sina.com.cn/corp/go.php/vII_NewestComponent/indexid/399808.phtml). CSMAR and WIND were not used.

Code Availability

Cleaning and analysis were written in Python 3 with pandas, statsmodels, pypdf, requests and matplotlib. Derived CSV files allow the statistical results to be rerun without repeating PDF downloads. The supporting code/scripts and documentation are provided with the submission for peer review.

References

1. Zhang S, Andrews-Speed P, Zhao X, He Y (2013) Interactions between renewable energy policy and renewable energy industrial policy: a critical analysis of China's policy approach to renewable

- energies. *Energy Policy* 62:342–353. <https://doi.org/10.1016/j.enpol.2013.07.063>
2. Vial G (2019) Understanding digital transformation: a review and a research agenda. *J Strateg Inf Syst* 28:118–144. <https://doi.org/10.1016/j.jsis.2019.01.003>
 3. Verhoef PC, Broekhuizen T, Bart Y, Bhattacharya A, Dong JQ, Fabian N, Haenlein M (2021) Digital transformation: a multidisciplinary reflection and a research agenda. *J Bus Res* 122:889–901. <https://doi.org/10.1016/j.jbusres.2019.09.022>
 4. Bharadwaj A, El Sawy OA, Pavlou PA, Venkatraman N (2013) Digital business strategy: toward a next generation of insights. *MIS Q* 37:471–482. <https://doi.org/10.25300/MISQ/2013/37:2.3>
 5. Hanelt A, Bohnsack R, Marz D, Antunes Marante C (2021) A systematic review of the literature on digital transformation: insights and implications for strategy and organisational change. *J Manag Stud* 58:1159–1197. <https://doi.org/10.1111/joms.12639>
 6. Lanzolla G, Lorenz A, Miron-Spektor E, Schilling M, Solinas G, Tucci CL (2020) Digital transformation: what is new if anything? Emerging patterns and management research. *Acad Manag Discov* 6:341–350. <https://doi.org/10.5465/amd.2020.0149>
 7. Wessel L, Baiyere A, Ologeanu-Taddei R, Cha J, Blegind Jensen T (2021) Unpacking the difference between digital transformation and IT-enabled organisational transformation. *J Assoc Inf Syst* 22:102–129. <https://doi.org/10.17705/1jais.00655>
 8. Yoo Y, Henfridsson O, Lyytinen K (2010) The new organising logic of digital innovation: an agenda for information systems research. *Inf Syst Res* 21:724–735. <https://doi.org/10.1287/isre.1100.0322>
 9. Nambisan S, Lyytinen K, Majchrzak A, Song M (2017) Digital innovation management: reinventing innovation management research in a digital world. *MIS Q* 41:223–238. <https://doi.org/10.25300/MISQ/2017/41:1.03>
 10. Kane GC, Palmer D, Phillips AN, Kiron D, Buckley N (2015) Strategy, not technology, drives digital transformation. *MIT Sloan Manag Rev* 14:1–25
 11. Fitzgerald M, Kruschwitz N, Bonnet D, Welch M (2014) Embracing digital technology: a new strategic imperative. *MIT Sloan Manag Rev* 55:1–12
 12. Matt C, Hess T, Benlian A (2015) Digital transformation strategies. *Bus Inf Syst Eng* 57:339–343. <https://doi.org/10.1007/s12599-015-0401-5>
 13. Hess T, Matt C, Benlian A, Wiesböck F (2016) Options for formulating a digital transformation strategy. *MIS Q Exec* 15:123–139
 14. Sebastian IM, Ross JW, Beath C, Mocker M, Moloney KG, Fonstad NO (2017) How big old companies navigate digital transformation. *MIS Q Exec* 16:197–213
 15. Barney J (1991) Firm resources and sustained competitive advantage. *J Manag* 17:99–120. <https://doi.org/10.1177/014920639101700108>
 16. Wernerfelt B (1984) A resource-based view of the firm. *Strateg Manag J* 5:171–180. <https://doi.org/10.1002/smj.4250050207>

17. Peteraf MA (1993) The cornerstones of competitive advantage: a resource-based view. *Strateg Manag J* 14:179–191. <https://doi.org/10.1002/smj.4250140303>
18. Teece DJ, Pisano G, Shuen A (1997) Dynamic capabilities and strategic management. *Strateg Manag J* 18:509–533. [https://doi.org/10.1002/\(SICI\)1097-0266\(199708\)18:7%3C509::AID-SMJ882%3E3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7%3C509::AID-SMJ882%3E3.0.CO;2-Z)
19. Teece DJ (2007) Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strateg Manag J* 28:1319–1350. <https://doi.org/10.1002/smj.640>
20. Warner KSR, Wäger M (2019) Building dynamic capabilities for digital transformation: an ongoing process of strategic renewal. *Long Range Plann* 52:326–349. <https://doi.org/10.1016/j.lrp.2018.12.001>
21. Karimi J, Walter Z (2015) The role of dynamic capabilities in responding to digital disruption: a factor-based study of the newspaper industry. *J Manag Inf Syst* 32:39–81. <https://doi.org/10.1080/07421222.2015.1029380>
22. Li L, Su F, Zhang W, Mao JY (2018) Digital transformation by SME entrepreneurs: a capability perspective. *Inf Syst J* 28:1129–1157. <https://doi.org/10.1111/isj.12153>
23. Wu F, Hu HZ, Lin HY, Ren XY (2021) Enterprise digital transformation and capital market performance: empirical evidence from stock liquidity. *Manag World* 37(7):130–144 (in Chinese)
24. Wu Y, Li Z (2024) Digital transformation, entrepreneurship, and disruptive innovation: evidence of corporate digitalisation in China from 2010 to 2021. *Humanit Soc Sci Commun* 11:163. <https://doi.org/10.1057/s41599-023-02378-3>
25. Peng Y, Tao C (2022) Can digital transformation promote enterprise performance?—From the perspective of public policy and innovation. *J Innov Knowl* 7:100198. <https://doi.org/10.1016/j.jik.2022.100198>
26. Chen C, Zhang Y, Wang S (2023) Digital transformation and firm performance: a case study on China's listed companies in 2009–2020. *Digit Econ Sustain Dev* 1:18. <https://doi.org/10.1007/s44265-023-00018-x>
27. Ren Y, Zhang X, Chen H (2022) The impact of new energy enterprises' digital transformation on their total factor productivity: empirical evidence from China. *Sustainability* 14:13928. <https://doi.org/10.3390/su142113928>
28. Guo L, Chen F, Chen L (2025) How does digital transformation affect corporate sustainability performance? Evidence from listed energy companies in China. *PLoS ONE* 20:e0325898. <https://doi.org/10.1371/journal.pone.0325898>
29. Cheng C, Cui H (2024) Combining digital and legacy technologies: firm digital transformation strategies—evidence from Chinese manufacturing companies. *Humanit Soc Sci Commun* 11:1021. <https://doi.org/10.1057/s41599-024-03498-0>
30. Brynjolfsson E, McAfee A (2014) The second machine age: work, progress, and prosperity in a time of brilliant technologies. W. W. Norton, New York

31. Petersen MA (2009) Estimating standard errors in finance panel data sets: comparing approaches. *Rev Financ Stud* 22:435–480. <https://doi.org/10.1093/rfs/hhn053>
32. Cameron AC, Miller DL (2015) A practitioner's guide to cluster-robust inference. *J Hum Resour* 50:317–372. <https://doi.org/10.3368/jhr.50.2.317>
33. Hinings B, Gegenhuber T, Greenwood R (2018) Digital innovation and transformation: an institutional perspective. *Inf Organ* 28:52–61. <https://doi.org/10.1016/j.infoandorg.2018.02.004>
34. China Securities Index Co., Ltd (2026) CSI New Energy Index (399808) factsheet, 31 July 2026. <https://www.csindex.com.cn>. Accessed 30 August 2026
35. Teece DJ (2018) Business models and dynamic capabilities. *Long Range Plann* 51:40–49. <https://doi.org/10.1016/j.lrp.2017.06.007>
36. Eisenhardt KM, Martin JA (2000) Dynamic capabilities: what are they? *Strateg Manag J* 21:1105–1121. [https://doi.org/10.1002/1097-0266\(200010/11\)21:10/11%3C1105::AID-SMJ133%3E3.0.CO;2-E](https://doi.org/10.1002/1097-0266(200010/11)21:10/11%3C1105::AID-SMJ133%3E3.0.CO;2-E)
37. Wang CL, Ahmed PK (2007) Dynamic capabilities: a review and research agenda. *Int J Manag Rev* 9:31–51. <https://doi.org/10.1111/j.1468-2370.2007.00201.x>
38. Ambrosini V, Bowman C (2009) What are dynamic capabilities, and are they a useful construct in strategic management? *Int J Manag Rev* 11:29–49. <https://doi.org/10.1111/j.1468-2370.2008.00251.x>
39. Barreto I (2010) Dynamic capabilities: a review of past research and an agenda for the future. *J Manag* 36:256–280. <https://doi.org/10.1177/0149206309350776>
40. Peteraf M, Di Stefano G, Verona G (2013) The elephant in the room of dynamic capabilities: bringing two diverging conversations together. *Strateg Manag J* 34:1389–1410. <https://doi.org/10.1002/smj.2078>
41. Helfat CE, Finkelstein S, Mitchell W, Peteraf M, Singh H, Teece D, Winter SG (2007) *Dynamic capabilities: understanding strategic change in organisations*. Blackwell, Malden, MA
42. Cohen WM, Levinthal DA (1990) Absorptive capacity: a new perspective on learning and innovation. *Adm Sci Q* 35:128–152. <https://doi.org/10.2307/2393553>
43. Zahra SA, George G (2002) Absorptive capacity: a review, reconceptualisation, and extension. *Acad Manag Rev* 27:185–203. <https://doi.org/10.5465/amr.2002.6587995>
44. Helfat CE, Peteraf MA (2003) The dynamic resource-based view: capability lifecycles. *Strateg Manag J* 24:997–1010. <https://doi.org/10.1002/smj.332>
45. Helfat CE, Peteraf MA (2015) Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strateg Manag J* 36:831–850. <https://doi.org/10.1002/smj.2247>
46. Zollo M, Winter SG (2002) Deliberate learning and the evolution of dynamic capabilities. *Organ Sci* 13:339–351. <https://doi.org/10.1287/orsc.13.3.339.2780>
47. Winter SG (2003) Understanding dynamic capabilities. *Strateg Manag J* 24:991–995. <https://doi.org/10.1002/smj.318>

48. Grant RM (1991) The resource-based theory of competitive advantage: implications for strategy formulation. *Calif Manag Rev* 33:114–135. <https://doi.org/10.2307/41166664>
49. Grant RM (1996) Toward a knowledge-based theory of the firm. *Strateg Manag J* 17:109–122. <https://doi.org/10.1002/smj.4250171110>
50. Wooldridge JM (2010) *Econometric analysis of cross-section and panel data*, 2nd edn. MIT Press, Cambridge, MA
51. Nickell S (1981) Biases in dynamic models with fixed effects. *Econometrica* 49:1417–1426. <https://doi.org/10.2307/1911408>
52. Baron RM, Kenny DA (1986) The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 51:1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
53. Preacher KJ, Hayes AF (2008) Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav Res Methods* 40:879–891. <https://doi.org/10.3758/BRM.40.3.879>
54. Zhao X, Lynch JG Jr, Chen Q (2010) Reconsidering Baron and Kenny: myths and truths about mediation analysis. *J Consum Res* 37:197–206. <https://doi.org/10.1086/651257>
55. Hayes AF (2018) *Introduction to mediation, moderation, and conditional process analysis: a regression-based approach*, 2nd edn. Guilford Press, New York
56. MacKinnon DP, Lockwood CM, Williams J (2004) Confidence limits for the indirect effect: distribution of the product and resampling methods. *Multivar Behav Res* 39:99–128. https://doi.org/10.1207/s15327906mbr3901_4
57. Sobel ME (1982) Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol Methodol* 13:290–312. <https://doi.org/10.2307/270723>
58. Imai K, Keele L, Tingley D (2010) A general approach to causal mediation analysis. *Psychol Methods* 15:309–334. <https://doi.org/10.1037/a0020761>

Figures

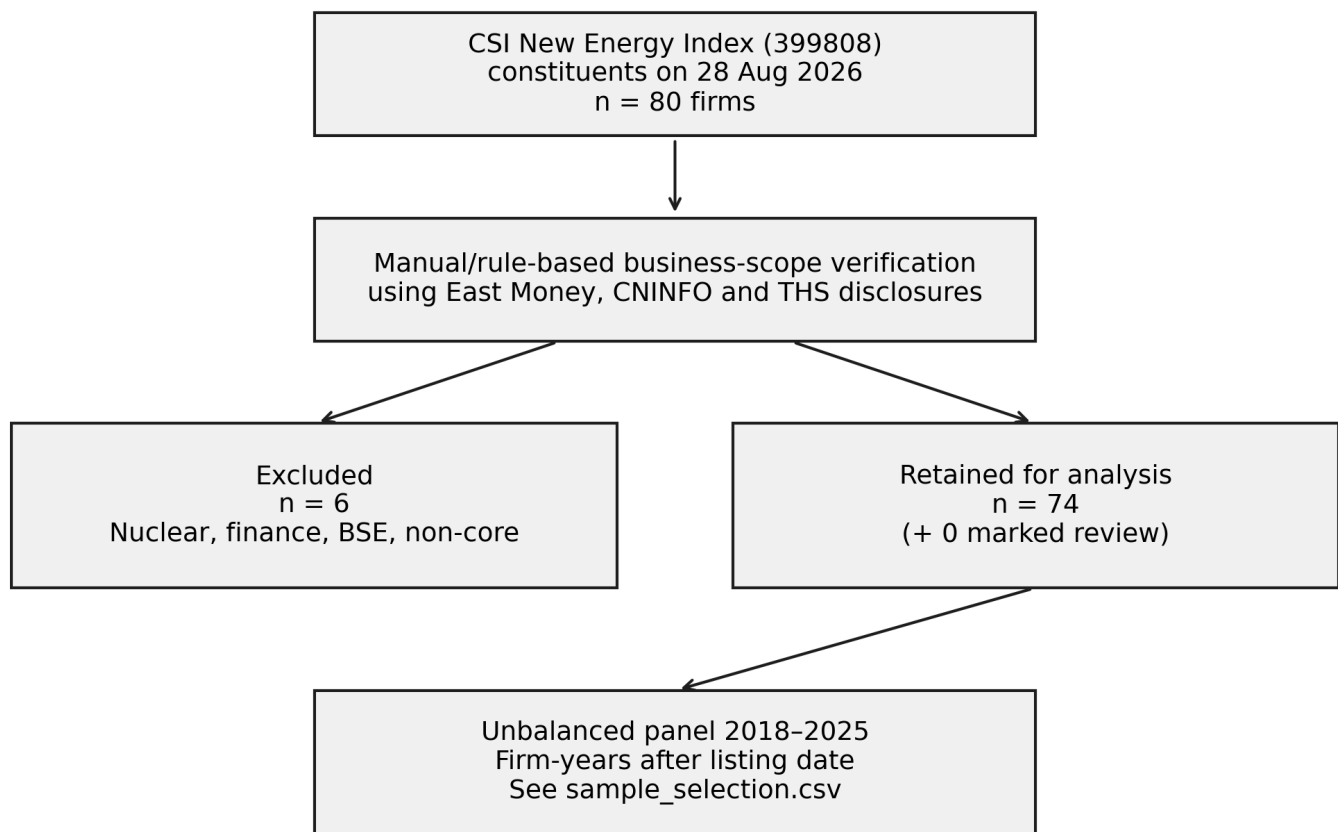


Figure 1

Sample-selection flowchart. The candidate universe is the CSI New Energy Index (399808) constituent list on 28 August 2026 (n = 80). Six firms were excluded; 74 firms and 495 firm-years remain.

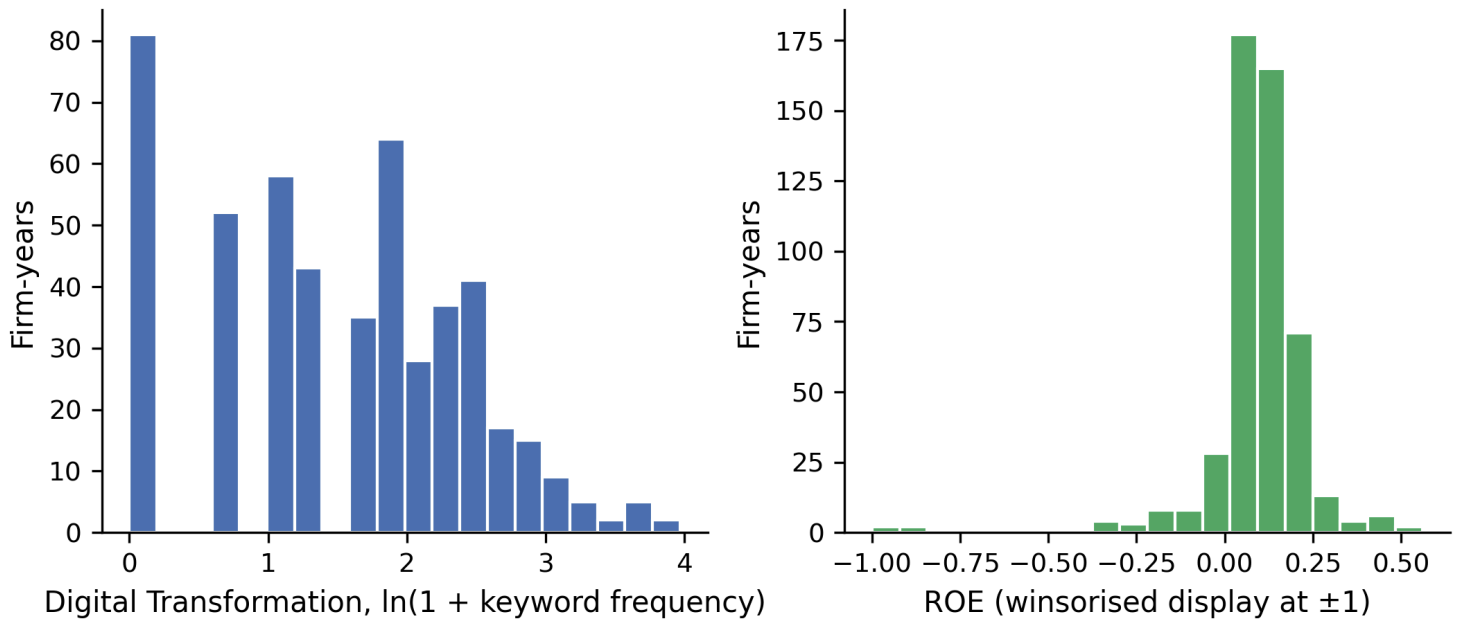


Figure 2

Distributions of digital transformation, $\ln(1 + \text{dictionary frequency})$, and ROE in the analysis sample. ROE is clipped at ± 1 for display only; models use the unclipped series.

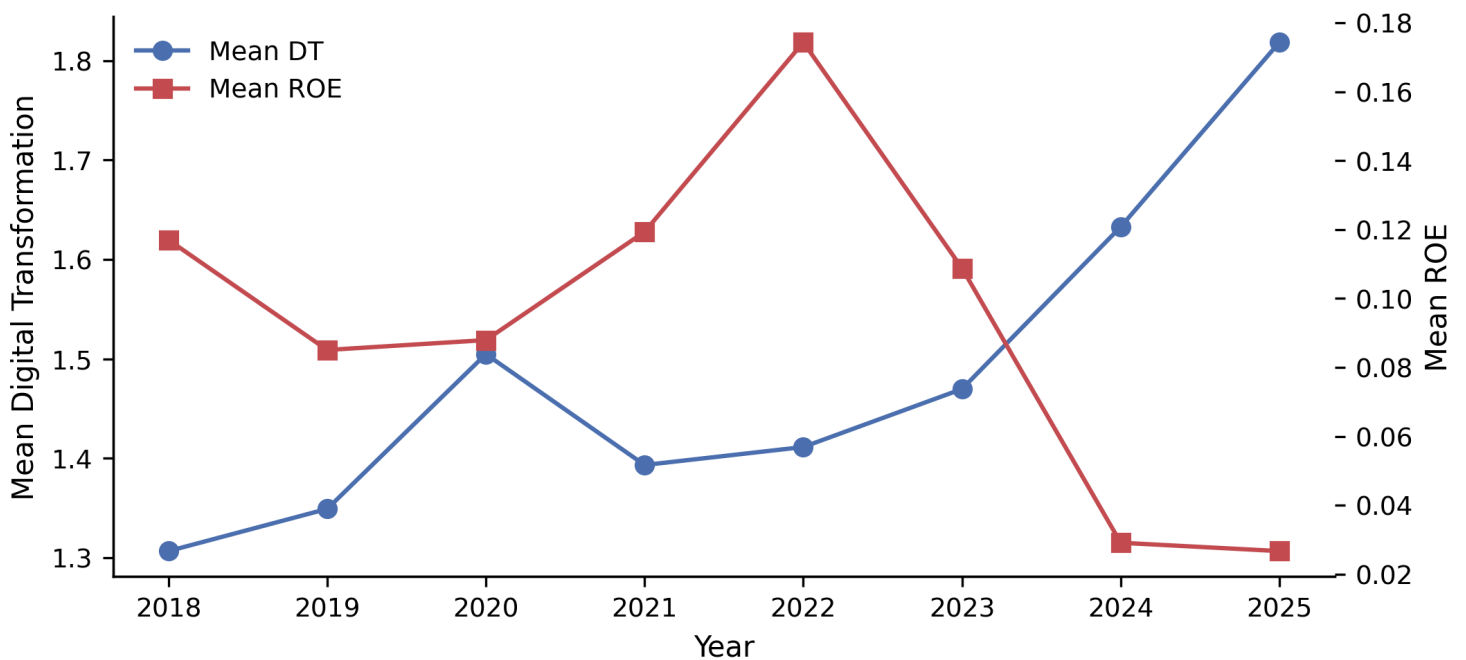


Figure 3

Annual means of digital transformation (left axis) and ROE (right axis), 2018–2025. Means are unweighted firm-year averages in the analysis sample.

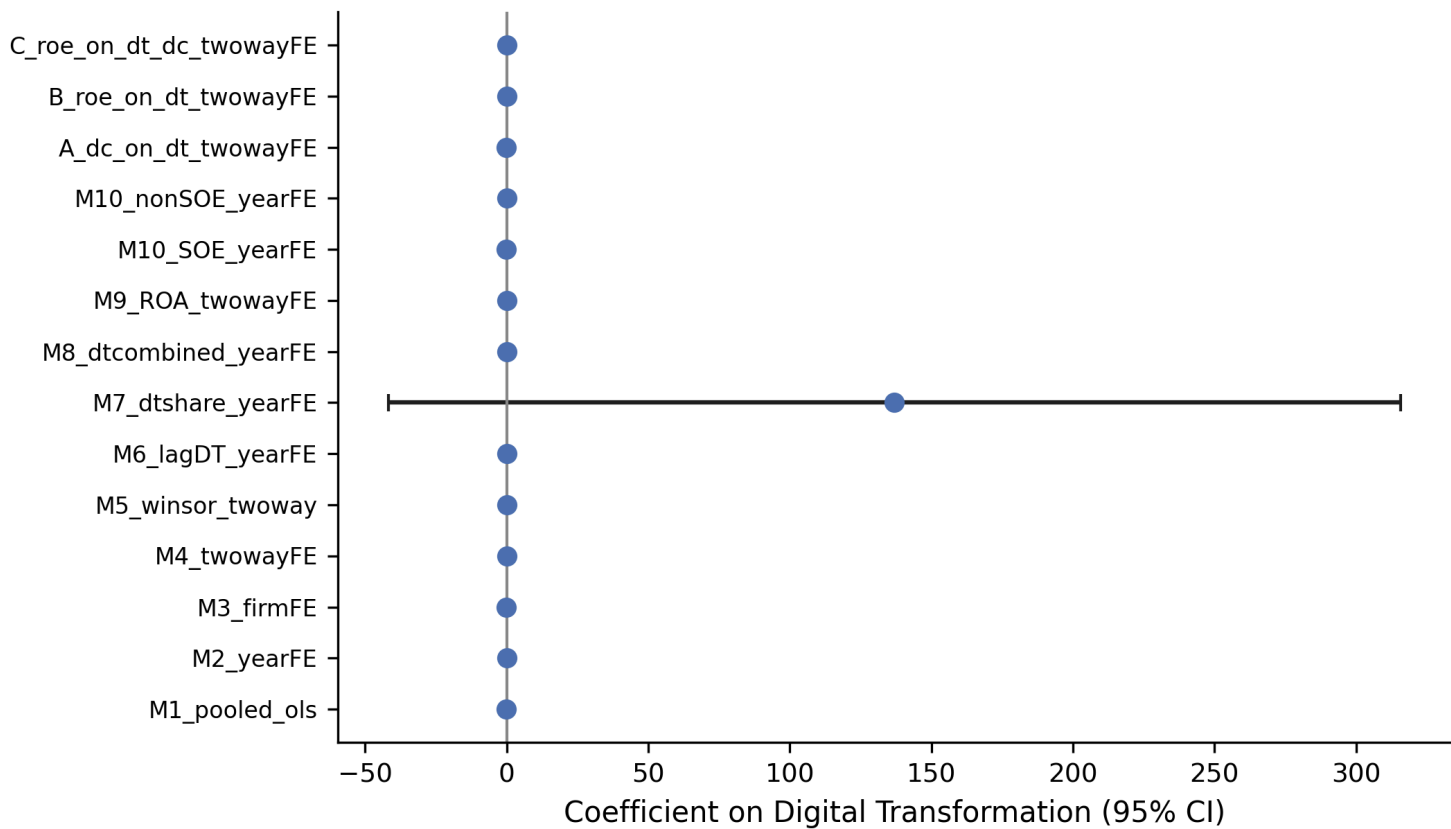


Figure 4

Digital-transformation coefficients and 95% firm-clustered confidence intervals across specifications are reported in Table 4. The vertical line marks zero.

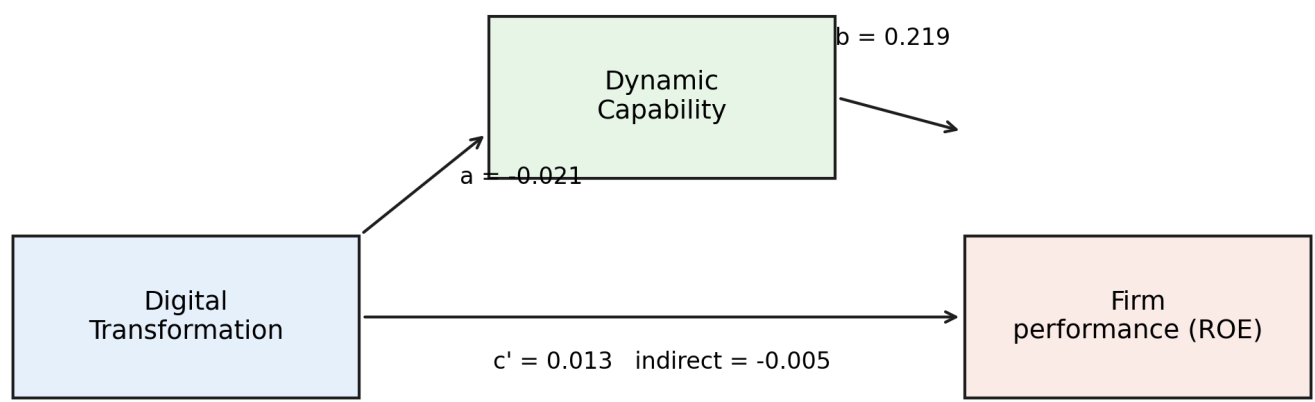


Figure 5

Exploratory mediation paths. The bootstrap interval for the indirect effect $a \times b$ includes zero.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.docx](#)