Computer and Decision Making

An International Journal

www.comdem.org
eISSN: 3008-1416



Financial data analytics without identification: a structured review of a fragmenting field

Dinesh Junghare[1*], Dr. Sachin Hanmant Lad[2], Dr. Deepak Ranjan Sahoo[3], Prof. Rajesh Babarao Sasane[4]

[1*]Ph.D. Scholar. MIT-Art, Design & Technology University, Pune, dineshJunghare@gmail.com, ORCID ID: 0009-0007-9993-7993
[2]Associate Professor, MIT-Art, Design & Technology University, Pune, sachin.lad@mituniversity.edu.in, ORCID ID: 0000-0002-1821-3499
[3]Associate Professor, MIT Art Design and Technology University, Email Id: dcdrsahoo@gmail.com, ORCID ID: 0000-0003-4013-2422
[4]Assistant Professor, MIT Art Design and Technology University, Email Id: mr.rjsh31@gmail.com, ORCID ID: 0000-0002-2745-9562

ARTICLE INFO

Article history:
Received 26 February 2026
Received in revised form 29 July 2026
Accepted 22 September 2026
Available online 29 October 2026

Keywords: financial data analytics; causal identification; systematic review; big data analytics capability; fintech; research methods.

ABSTRACT

We review 224 studies of financial data analytics published between 2020 and 2026 and document a mismatch between the questions the field asks and the designs it uses to answer them. Roughly three in ten titles in the corpus assert a causal relationship — analytics capability improving financial performance, digital inclusion reducing unemployment, ESG orientation raising firm value — while not one names a difference-in-differences, instrumental-variable, regression-discontinuity, or event-study design. The field is also centrifugal: 224 studies are spread across 124 outlets, of which 85 publish a single paper on the subject, and the largest research stream is shrinking rather than consolidating. Its share of annual output falls from 31.0% in 2020–2022 to 15.2% in 2025–2026, displaced by sustainability, fraud detection, and household finance applications. We interpret this pattern as premature thematic diversification: the field moved on from its foundational question about whether analytics investment pays before establishing an evidentiary basis for answering it. We propose an identification agenda built around shocks the sector already supplies — staggered data-protection regimes, open-banking mandates, supervisory technology rollouts, and cloud infrastructure failures — and specify design templates for each. The review is bounded by its reliance on title- and outlet-level coding of an author-compiled corpus, which we discuss at length.

1. Introduction

Between 2020 and 2026, the number of studies published annually on financial data analytics in the corpus assembled for this review grew from 11 to 79. Journals across accounting, information systems, operations research, banking, and computer science now publish work under this heading, and the topic has acquired the conventional markers of a maturing field: dedicated review articles (Triyanto & Ali, 2026; Shahana et al., 2023), bibliometric mappings (Aboelfotoh et al., 2024; Karim et al., 2026), and meta-analyses (Oduro et al., 2023; Santini et al., 2026). Growth of this kind is usually read as consolidation.

The corpus tells a different story. Volume has increased without the accompanying features that make a literature cumulative — a shared dependent variable, a common measurement standard, a set of venues where rival claims meet, or a design convention strong enough that a referee could reject a paper for violating it. What has grown instead is the range of things the phrase “financial data analytics” is asked to cover. In the records retrieved for this study, the term attaches to fraud detection in payment networks, curriculum design for accounting undergraduates, the financial toxicity of stroke among working-age survivors, sentiment extraction from Bloomberg newswires, and the effect of financial strain on bladder health across a twenty-five-year panel. Roughly a quarter of the retrieved records use “financial” or “analytics” in a sense that has no bearing on financial data analytics as a research object.

* *Corresponding author.*
E-mail address: dineshJunghare@gmail.com.

This paper takes that fragmentation seriously as an object of study rather than treating it as noise to be screened away. Reviews of emerging fields typically report what is known and propose what to do next. We report something less comfortable: on the evidence of this corpus, considerably less is known than the volume of publication implies, because the field's dominant claims are causal and its dominant designs are not.

The central exhibit is a simple contrast. Of 224 in-scope studies, 64 (28.6%) make an explicit causal claim in the title itself, using constructions such as "the impact of," "the effect of," "the influence of," "drivers of," "enhances," or "drives." These are not hedged associations. They are assertions about what produces what. Against this, no title in the corpus names a difference-in-differences design, an instrumental variable, a regression discontinuity, a natural experiment, a propensity-score match, a synthetic control, or an event study. Two titles contain the word "causal," and both belong to the same pre-registered study of operational risk, published twice as a report and then as a completed study (Cornwell et al., 2023a, 2023b). Titles understate methods, and we return to this caveat repeatedly. But a field in which the identification strategy is never salient enough to reach the title is a field in which identification is not what distinguishes a good paper from a mediocre one.

The consequence is visible in the field's thematic trajectory. The largest stream in the corpus asks whether investment in analytics capability improves firm financial performance — the question on which the practical case for the entire enterprise rests. That stream accounted for 31.0% of output in 2020–2022 and 15.2% in 2025–2026. It did not shrink because the question was settled. It shrank because attention moved to newer applications: sustainability and green finance rose from 7.1% to 16.0% of output over the same period, credit and risk modelling from 4.8% to 10.4%, and fintech consumer adoption from nothing at all to 7.2%. The field diversified before it converged.

This study makes three contributions. First, it provides a structured descriptive anatomy of the financial data analytics literature over its most active seven-year period, with an explicit coding protocol and reported counts, allowing readers to disagree with our classification and recompute. Second, it diagnoses three structural weaknesses — the identification gap, venue fragmentation, and the unresolved capability–performance chain — and shows that each is visible in the corpus rather than merely asserted. Third, it converts the diagnosis into an agenda, specifying the shocks available in this domain and the design each shock supports, so that the recommendation is operational rather than an exhortation to "employ more rigorous methods."

The paper proceeds as follows. Section 2 defines the boundary of the field, since the boundary problem turns out to be substantive rather than administrative. Section 3 describes the corpus and coding protocol. Section 4 presents the descriptive anatomy. Section 5 synthesises the nine research streams. Section 6 develops the three structural weaknesses. Section 7 sets out the research agenda. Section 8 states the limitations, which are more consequential here than in a conventional review, and Section 9 concludes.

2. What Counts as Financial Data Analytics?

The term resists definition in a way that matters for what follows. Three distinct usages appear in the corpus, and conflating them is one reason the literature accumulates slowly.

In the first usage, financial data is the *input* and analytics is the method. Studies of fraud detection in transaction streams (Luong & Xie, 2026; Almusallam & Qayyum, 2025), volatility forecasting from order-book data (Akyildirim et al., 2021; Ding et al., 2022), and credit scoring from borrower histories (Machado et al., 2025) all belong here. The research question is predictive and the contribution is typically a model that outperforms a benchmark on a held-out sample.

In the second usage, analytics is the *treatment* and a financial outcome is the effect. Studies asking whether big data analytics capability raises firm financial performance (Gupta et al., 2020; Saeed et al., 2023; Itani et al., 2024), whether analytics adoption improves audit quality (Abdelwahed et al., 2024), or whether digital financial inclusion reduces unemployment (Islam et al., 2026) belong here. The research question is causal, and the contribution should be an estimate of an effect.

In the third usage, the loosest of the three, "financial" modifies a variable that happens to be measured in money, and "analytics" means any quantitative method. Studies of the financial toxicity of stroke (Jacobs & Ellis, 2026), the financial burden of Ehlers-Danlos syndromes (Schubart et al., 2025), financial strain and women's bladder health (Brady et al., 2024), and financial stress in neuroendocrine patterning (Hamilton & Steptoe, 2025) fall here. This work is often excellent. It is not financial data analytics in any sense that would let it speak to the first two usages.

The distinction between the first two usages is where the analytical stakes lie. Predictive and causal questions require different designs, different validation standards, and different claims. A model that classifies fraudulent transactions with high recall has demonstrated something real without demonstrating anything causal. A study

reporting that firms with stronger analytics capability report stronger financial performance has demonstrated a correlation that is compatible with capability causing performance, performance funding capability, or a common managerial quality producing both. The literature routinely reports the second kind of result in the vocabulary of the first kind of certainty.

This study restricts attention to the first two usages. We define **financial data analytics** as the application of computational, statistical, or machine-learning methods to data generated by financial activity — transactions, prices, disclosures, credit files, payment flows, audit evidence — or to the organisational adoption of such methods, where the research object is a financial process, agent, or outcome. The definition deliberately admits the accounting and audit literature, which some treatments exclude, because audit evidence is financial data and the audit function is a financial process. It excludes studies where a monetary variable appears as a covariate in an otherwise unrelated design.

3. Corpus and Coding Protocol

3.1 Corpus construction

The corpus for this review is a compiled reference set of 298 records spanning 2020 to 2026, assembled from database searches combining terms for financial activity with terms for analytics, business intelligence, big data, and machine learning. We did not conduct an independent PRISMA-compliant search, and the corpus should be understood as a purposively assembled rather than exhaustively retrieved sample. Section 8 discusses what this costs.

3.2 Screening

Each record was screened against the definition in Section 2 using its title and publication outlet. Records were excluded where a monetary or financial variable appeared incidentally within a design belonging to another field — clinical outcomes research, public health, energy engineering, transport operations, sports medicine, and library and information science accounted for most exclusions. Seventy-four records (24.8%) were excluded on this basis, leaving 224 in-scope studies.

The exclusion rate is itself a finding. A quarter of what a conventional keyword search returns for this topic is not about this topic. Reviews that do not screen aggressively will report a field that appears larger and more interdisciplinary than it is.

3.3 Thematic coding

In-scope studies were assigned to one of nine mutually exclusive themes by applying an ordered set of keyword rules to titles and outlets, with the first matching rule determining assignment. Ordering matters: a study of fraud detection using deep learning matches both the fraud rule and the machine-learning rule, and we resolved such cases in favour of the substantive application rather than the method, since the field is organised by application. The rule set and its ordering are reported in full in the Appendix so that readers can reproduce or contest the classification.

One study resisted assignment and was folded into the closest theme by inspection. Boundary cases are unavoidable in a scheme of this kind; we prefer a transparent rule that misclassifies a handful of studies to a subjective scheme that misclassifies an unknown number.

3.4 Method coding

We attempted to code methodology from titles and abandoned the attempt as unreliable: more than a third of titles gave no usable signal, and the classifications that did emerge were sensitive to arbitrary rule ordering. We therefore report only two method-related counts, both of which depend on explicit and unambiguous title terms: whether a title names a machine-learning or artificial-intelligence technique, and whether a title names a quasi-experimental identification design. Both are conservative — they undercount, because methods are frequently described only in the abstract or body. We use them as lower bounds and interpret them accordingly.

4. Descriptive Anatomy of the Field

4.1 Growth

Annual output in the in-scope corpus rose from 11 studies in 2020 to 79 in 2026, with the steepest increase after 2024 (Table 1). Studies published in 2025 and 2026 account for 125 of 224 records, or 55.8% of the corpus. More than half of the literature on financial data analytics, as represented here, was published in its final two years.

Table 1. Annual distribution of in-scope studies

Year	Studies	Share of corpus
2020	11	4.9%
2021	15	6.7%
2022	16	7.1%
2023	30	13.4%
2024	27	12.1%
2025	46	20.5%
2026	79	35.3%
Total	224	100%

Two caveats apply. Compiled corpora over-represent recent work because recent work is more visible in search, and the 2026 figure covers a partial year in some indexing systems while including forward-dated issues in others. The growth pattern is robust to these concerns in direction but not in magnitude, and we do not compute growth rates from it.

Figure 1 plots the same series with cumulative output overlaid. The inflection after 2024 is the visually dominant feature, and the shaded band marks the two years that contain more than half the corpus.

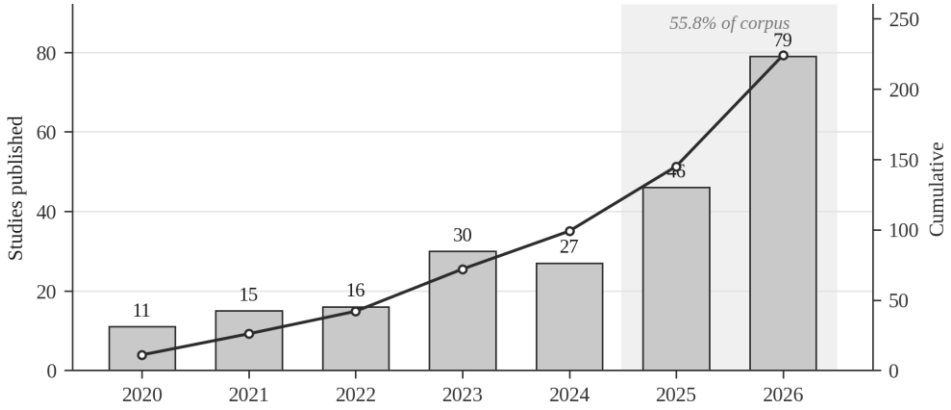


Figure 1. Annual and cumulative publication output, 2020–2026

Note. Bars show in-scope studies published in each year (left axis); the line shows cumulative output (right axis). The shaded band covers 2025–2026, which together account for 125 of 224 studies. Counts are drawn from the compiled corpus described in Section 3 and are subject to the recency bias discussed there and in Section 8.

4.2 Venue dispersion

The 224 studies appear in 124 distinct outlets. Eighty-five of those outlets, 68.5% of the total, publish exactly one study in the corpus. The most frequent venue, *Decision Analytics Journal*, accounts for 13 studies, or 5.8% of the total (Table 2).

Table 2. Most frequent publication outlets

Outlet	Studies
Decision Analytics Journal	13
Procedia Computer Science	10
Technological Forecasting and Social Change	9
Journal of Accounting Education	8
Expert Systems with Applications	7
Finance Research Letters	5
Journal of Financial Reporting and Accounting	4
International Review of Financial Analysis	4
Industrial Marketing Management	4
Borsa Istanbul Review	4
Journal of Open Innovation	4
Journal of Economic Criminology	4

No outlet functions as a field journal. For comparison, a consolidated subfield typically shows a concentrated head — three or four venues carrying a quarter or more of output — followed by a long tail. Here the head is barely distinguishable from the tail, and the leading venue is a general decision-science outlet rather than a finance journal. The four studies in *Finance Research Letters* and four in *International Review of Financial Analysis* represent the field's entire footprint in core finance journals within this corpus.

Figure 2 makes the dispersion visible from two directions. The left panel shows how many outlets publish how many studies, and the distribution is dominated by a single spike at one. The right panel plots cumulative studies against outlets ranked by output. In a field with a recognised core, this curve rises steeply and then flattens, because a few venues carry most of the work. Here it bends only gently away from the diagonal representing perfectly equal dispersion.

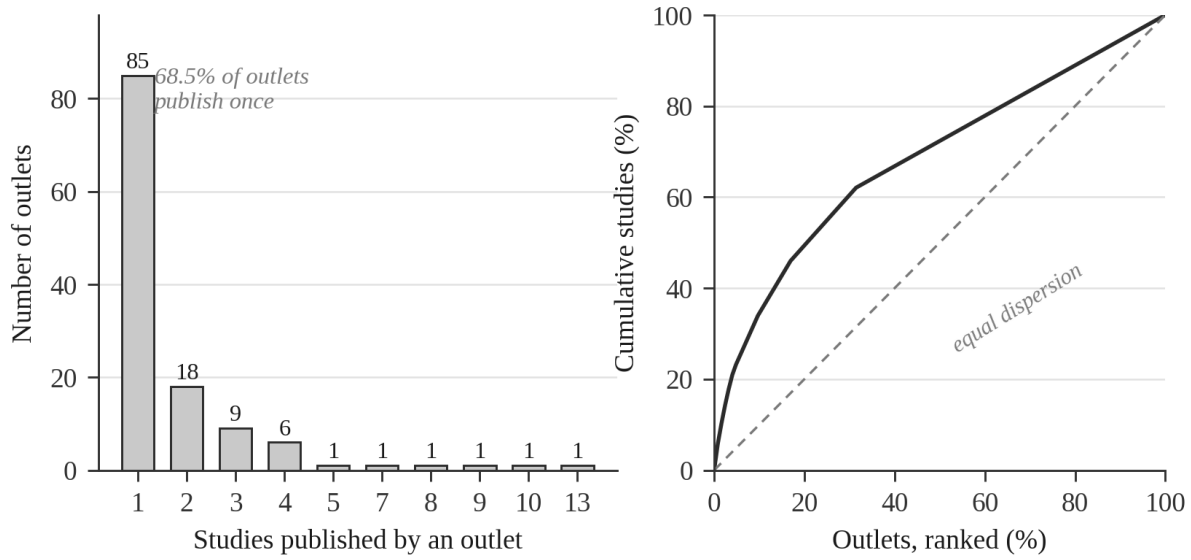


Figure 2. Venue dispersion in the in-scope corpus

Note. Left panel: number of outlets (vertical axis) publishing a given number of studies (horizontal axis); 85 of 124 outlets publish exactly one. Right panel: concentration curve of cumulative studies against outlets ranked from most to least productive; the dashed diagonal represents equal dispersion across outlets.

4.3 Thematic composition

Table 3 reports the thematic distribution. Nine streams are identifiable, and the distribution is flatter than a maturing field would predict: the largest stream holds just over a fifth of output, and six of nine streams fall between 9% and 15%.

Table 3. Thematic composition of the in-scope corpus

Theme	Studies	Share
Analytics capability, firm performance and organisational adoption	46	20.5%
Audit, accounting and assurance analytics	32	14.3%
Market forecasting, pricing and trading	26	11.6%
Sustainability, ESG and green finance	25	11.2%
Household finance, financial literacy and inclusion	23	10.3%
Fraud, financial crime and anti-money laundering	22	9.8%
Credit, risk and financial distress	22	9.8%
Financial data infrastructure and methods	15	6.7%
Fintech, digital finance and consumer adoption	9	4.0%
Data governance, privacy and regulation	4	1.8%
Total	**224**	**100%**

The smallest stream is the most striking. Four studies, 1.8% of the corpus, address the governance, privacy, and

regulatory conditions under which financial data analytics is permitted to operate (Espinosa et al., 2026; Nguyen-Nam et al., 2026; Haq, 2023; Murikah et al., 2026). Everything else in the corpus presumes lawful access to granular financial data about identifiable people. Access is the binding constraint on the entire enterprise, and the field devotes one paper in fifty-six to it.

Figure 3 displays the distribution. The flatness is the point: no stream dominates, and the gap between the largest and the median stream is smaller than one would expect in a literature organised around a shared question.

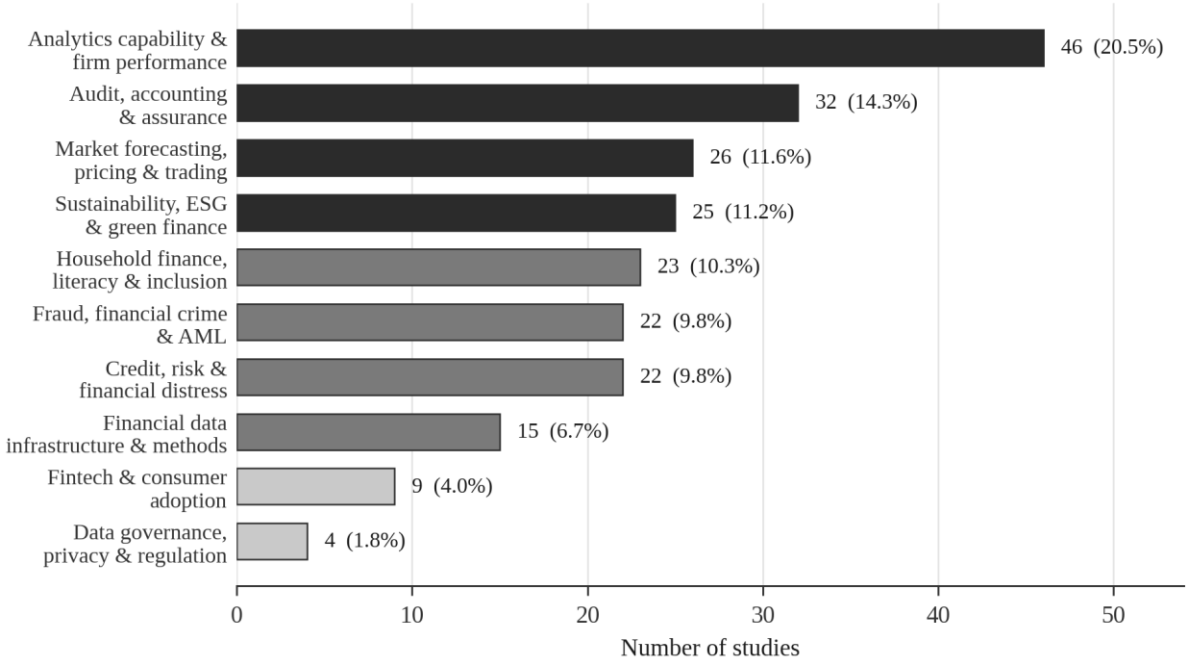


Figure 3. Thematic composition of the in-scope corpus (n = 224)

Note. Bars show the number of studies assigned to each theme under the ordered keyword rules reported in Appendix A; percentages of the corpus appear alongside. Assignment is mutually exclusive, so studies spanning two themes are allocated to the first matching rule, which favours substantive application over method.

4.4 Thematic drift

Comparing the field's first three years with its last two reveals substantial reallocation of attention (Table 4).

Table 4. Thematic shares, 2020–2022 versus 2025–2026

Theme	2020–2022 (n = 42)	2025–2026 (n = 125)	Change
Analytics capability and firm performance	31.0%	15.2%	−15.8 pp
Financial data infrastructure and methods	16.7%	4.8%	−11.9 pp
Audit, accounting and assurance	14.3%	10.4%	−3.9 pp
Sustainability, ESG and green finance	7.1%	16.0%	+8.9 pp
Fintech, digital finance and consumer adoption	0.0%	7.2%	+7.2 pp
Credit, risk and financial distress	4.8%	10.4%	+5.6 pp
Market forecasting, pricing and trading	9.5%	12.8%	+3.3 pp
Household finance, literacy and inclusion	9.5%	11.2%	+1.7 pp
Fraud, financial crime and AML	7.1%	9.6%	+2.5 pp
Data governance, privacy and regulation	0.0%	2.4%	+2.4 pp

Two movements dominate. The capability–performance stream loses half its relative weight, and the infrastructure and methods stream loses more than two-thirds. Sustainability and green finance nearly doubles, and fintech consumer adoption appears from nothing.

Figure 4 presents the reallocation as a shift between two periods, with each theme's 2020–2022 share connected to its 2025–2026 share. Ordering the themes by the size of the shift separates the field's growth areas from the two streams it has moved away from.

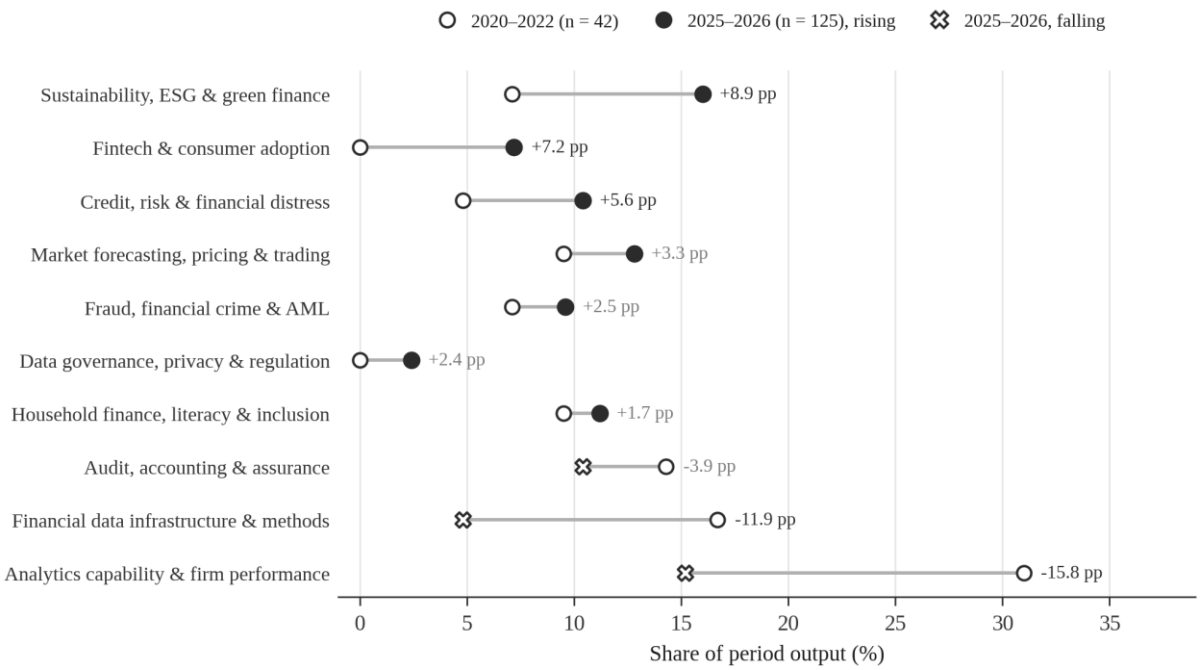


Figure 4. Thematic reallocation between 2020–2022 and 2025–2026

Note. Open circles mark each theme's share of output in 2020–2022 (n = 42); filled circles and crosses mark the 2025–2026 share (n = 125), for rising and falling themes respectively. Labels report the change in percentage points. The early subsample is small, so individual positions are imprecisely estimated; only the two largest movements exceed eleven percentage points.

The decline in infrastructure work has a benign reading. Early studies in the corpus were preoccupied with pipelines, platforms, and processing architecture (Li, J., 2022; Babar, 2025; Gu et al., 2021), and that work may simply have been absorbed into commercial infrastructure, freeing researchers to study applications. The decline in the capability–performance stream admits no such reading, and Section 6.3 takes it up.

4.5 Method salience

Forty-two in-scope titles (18.8%) explicitly name a machine-learning or artificial-intelligence technique. The share rises from 3 of 42 studies in 2020–2022 to 31 of 125 in 2025–2026, from 7.1% to 24.8%. Seventeen studies (7.6%) identify themselves as reviews, bibliometric mappings, or meta-analyses, a high proportion for a field of this age and a further sign that synthesis is outpacing the accumulation of comparable primary results.

No title names a quasi-experimental design. The comparison with the causal-claim count in Section 6.1 is the analytical centre of this paper.

5. Thematic Synthesis

This section characterises each stream: what it has established, where it disagrees with itself, and what it has not attempted. We resist the convention of describing each study in turn, which produces catalogues rather than syntheses.

5.1 Analytics capability, firm performance and organisational adoption

The largest stream asks a single question in many forms: does building analytics capability improve financial outcomes? The theoretical apparatus is largely borrowed, resting on the resource-based view and the dynamic-capabilities framework, with analytics cast as a capability that is valuable, rare, and difficult to imitate (Gupta et al., 2020; Saeed et al., 2023; Arias-Pérez et al., 2021).

The stream is nearly unanimous in reporting positive effects. That unanimity is its central weakness. Studies find analytics capability improving performance directly (Koohang et al., 2023; Haverila et al., 2022), through marketing ambidexterity (Saeed et al., 2023), through supply chain learning and resilience (Iftikhar et al., 2025), through customer-linking and selling capabilities under market turbulence (Itani et al., 2024), through service agility (Edu, 2022), and through intellectual capital in banking (AL-Khatib, 2022). Almost any mediator proposed is found to mediate. A literature in which every hypothesised path is confirmed is a literature with a publication filter, a measurement problem, or both.

The measurement problem is identifiable. Both the independent and dependent constructs are frequently

perceptual and frequently obtained from the same respondent in the same instrument, which makes common-method variance a live alternative explanation for the entire stream. Haverila et al. (2024) partially address this by examining how the respondent's decision-making role conditions the reported relationship — an implicit acknowledgement that who answers the survey changes the answer.

Dissent exists but is quiet. Oduro et al. (2023), meta-analysing the digital technology–performance relationship, find heterogeneity substantial enough to question any single pooled estimate. Zhao, Xie, et al. (2026) offer what they describe as a critical examination of analytics capability in container port performance. Farouk et al. (2025) reframe the question through affordance actualisation, arguing that capability is inert until organisational conditions activate it — which implies that studies regressing performance on capability without those conditions are misspecified. Ain et al. (2025), reviewing measurement of business intelligence and analytics success, find no settled measurement standard. These four papers, taken together, constitute a serious challenge to the stream's central finding. None is positioned as such, and the stream continues largely undisturbed.

5.2 Audit, accounting and assurance analytics

The second-largest stream divides into two subliteratures with little traffic between them.

The first studies professional practice. Its most valuable finding is that adoption has been slower and more contested than technological accounts predict. De Santis and D'Onza (2021) frame audit analytics as a search for legitimacy rather than a search for efficiency, arguing that firms adopt partly to appear modern to clients and regulators. Liew et al. (2021) document the transformation in Big Four practice and find the "how" considerably messier than the "why." Kastrup et al. (2024) show that in financial due diligence, accounting and commercial logics pull analytics use in different directions. Oussii and Zouari (2026) locate both conformity and resistance in response to institutional pressure. This subliterature converges on a conclusion the capability stream would find uncomfortable: adoption is a political and institutional process, and adopting organisations may be responding to legitimacy pressure rather than to expected returns.

The second subliterature is pedagogical, and it is unusually large — eight of the 32 studies in this stream appear in the *Journal of Accounting Education* alone. It covers curriculum integration (Booker et al., 2023; Flint, 2025), teaching cases built around specific tools and standards (Parlier & Lee, 2023; Shan & Richardson, 2025; Altiero et al., 2026), generative AI in managerial accounting instruction (Wiseman et al., 2026), continuous feedback in analytics instruction (Juma'h et al., 2025), and skill formation among early-career accountants (Imjai et al., 2024). The volume signals genuine professional demand. It also means that a substantial fraction of what is published under financial data analytics concerns how to teach it rather than what is true about it.

Evidence on outcomes is thinner than evidence on adoption. Abdelwahed et al. (2024) report that adopting analytics enhances audit quality; Aboelfotoh et al. (2024) map the analytics–reporting-quality relationship bibliometrically rather than testing it. Whether analytics changes what auditors find, as distinct from how auditors work, remains substantially open. Werner et al. (2021), embedding process mining in statement audits, come closest to a mechanism.

5.3 Market forecasting, pricing and trading

This is the stream in which the field's methods are strongest and its economic claims weakest — the mirror image of Section 5.1.

Predictive contributions are substantial and cumulative. Yang et al. (2020) forecast volatility using support vector machines; Akyildirim et al. (2021) and Ding et al. (2022) exploit order imbalance for return and futures-volatility prediction; Ghosh and Jana (2023) build a granular framework for high-frequency variables during an extreme event. Sentiment extraction has become a subfield of its own, running from newswire classification (Vieira et al., 2026) through quantum-enhanced approaches (Bhangu et al., 2026) to sector-specific portfolio construction (Ghiassi et al., 2026). A methodological wing imports physics: physics-informed neural networks for barrier option pricing (Chen, Lü, Tian, & Li, 2025), kernel methods for constant-elasticity-of-variance models under nonlinear volatility (Mishra & Chakraverty, 2025), and singular-value learning for cross-covariance forecasting (Manolakis et al., 2026). Souto and Moradi (2024) contribute a loss function tailored to financial time series, and Naivasha et al. (2026) address the validation of synthetic financial data.

What this work does not do is engage the economic question of whether the predictability it documents survives trading costs, capacity constraints, and competition. Biondo et al. (2024) simulate trading strategies and Anderson and Thoma (2021) interview professional traders about how they actually reason, but these are exceptions. Reviews within the stream (Broby, 2022; Singh et al., 2022; Dhingra et al., 2026; Dote-Pardo & Espinosa-Jaramillo, 2026) catalogue methods more thoroughly than they compare performance on common benchmarks. Without shared evaluation data, incremental accuracy claims cannot be adjudicated, and the stream accumulates models faster than it accumulates knowledge.

5.4 Fraud, financial crime and anti-money laundering

The most technically coherent stream in the corpus, and the one with the clearest problem definition. Contributions cluster around three difficulties: extreme class imbalance, evolving adversarial behaviour, and the tension between detection accuracy and explainability.

Methodologically the stream has moved from graph representation (Henderson, 2020) through generative approaches to imbalance (Aftabi et al., 2023) to explainable ensembles (Luong & Xie, 2026), hybrid feature selection with clustering (Almusallam & Qayyum, 2025), dimensionality reduction (Lokanan, 2026), complex-network anomaly detection for anti-money laundering (Oliveira et al., 2025), and real-time architectures combining sequential and market-behaviour learning (Aktarujjaman et al., 2026). Explainability has moved from a stated desideratum to a design requirement, which reflects the operational reality that a detection flag must be justified to an investigator and eventually to a court.

A separate and less computational subliterature examines the organisational determinants of financial statement fraud — board characteristics and material weaknesses (Manry et al., 2023), CFO gender (Maulidi et al., 2023) — and the institutional geography of financial crime (Sr, 2020; Wu, 2026; Paul, 2026). These two subliteratures rarely cite each other, despite the obvious complementarity: the organisational work identifies where to look, and the computational work supplies the looking.

The stream's binding constraint is data. Detection models are trained on data held by institutions that cannot share it, which makes external replication nearly impossible and leaves reported performance essentially unauditable. Muñoz-Guerrero et al. (2026) release a Latin American fintech behavioural dataset, and the rarity of such releases in this corpus is itself informative. Synthetic data generation (Naivasha et al., 2026) and privacy-preserving computation (Nguyen-Nam et al., 2026) are the plausible routes out, and neither is yet standard practice.

5.5 Credit, risk and financial distress

This stream splits by the level at which risk is measured, and the split is nearly clean.

At the borrower level, the work is applied machine learning: credit risk assessment (Machado et al., 2025), reinforcement learning for credit decisions in manufacturing supply chains (Razaque et al., 2025), distress prediction under missing and imbalanced data (Yu et al., 2024), leverage prediction (Shehadeh et al., 2026), and debt collection prioritisation from contact-centre records (Sánchez et al., 2022). The most interesting contributions widen the data rather than deepen the model — Narayanamurthy et al. (2025) use satellite imagery to reduce uncertainty in institutional crop credit, which addresses the actual constraint in agricultural lending, namely the absence of borrower financial history.

At the system level, the work is closer to macro-financial economics: contagion under interacting channels (Wiersema et al., 2023), online estimation for financial-stability monitoring (Gu et al., 2021), scenario-based flood exposure (Caloia et al., 2026), climate risk in BRICS economies (Javed & Bhutta, 2026), and composite conditions indices (Pestere-Akcay & Durusu-Ciftci, 2026). Xie et al. (2026) provide one of the few studies in the entire corpus with a plausible exogenous source of variation, examining open government data and bank risk-taking.

Cornwell et al. (2023a, 2023b) deserve separate mention. Their pre-registered study of data-driven causal factors in operational risk is, on the evidence of this corpus, the only work that both pre-registers and foregrounds causal inference. That a single study occupies this position across 224 papers is a finding about the field rather than a compliment to the authors, though it is that too.

5.6 Household finance, financial literacy and inclusion

The stream with the strongest theoretical grounding and the weakest connection to analytics. Most of it is survey-based social science on financial capability, decision-making, and well-being: literacy and autonomy in decision-making (Kumar et al., 2023), intra-household allocation of financial power (Bialowolski et al., 2020), digital literacy and retirement planning through saving behaviour (Yadav et al., 2025), literacy and entrepreneurial self-efficacy (Hassan & Mertzanis, 2026), behavioural biases and financial satisfaction (Khalid & Riaz, 2026), and generational patterns in financial cognition (Öztürk et al., 2026). Meta-analytic work covers aging and risk-taking (Ordali & Rapallini, 2026) and trust in financial services (Santini et al., 2023).

Where analytics enters, it is usually as method rather than object: neural networks for describing household financial ratios (Heo et al., 2020), Python-based segmentation of household savings (Cuomo et al., 2023). Du et al. (2026) invert the relationship by reviewing data visualisation as an intervention to improve financial literacy, which is among the more original framings in the corpus and has almost no company.

The stream connects to the rest of the field mainly through digital inclusion, where studies report effects on regional development disparities (Ding & Kang, 2024), unemployment (Islam et al., 2026), and agroecological efficiency (Zhang et al., 2025). These are macro-level claims resting on aggregate panel data, and they are

precisely the claims most in need of identification and least likely to receive it.

5.7 Sustainability, ESG and green finance

The fastest-growing stream, more than doubling its share between the corpus's first and last periods. Three activities recur: constructing ESG measures from unstructured data (Fildisi et al., 2025; Patro et al., 2025), estimating ESG–performance relationships (Zarea et al., 2026; Phalalum & Panmarerng, 2026), and modelling green finance as a driver of environmental or developmental outcomes (Yang, Yu, & Zhang, 2026; Hu et al., 2026; Zhao, R., 2026).

Growth this rapid warrants scepticism about mechanism. Much of the stream studies relationships between two constructed indices — an ESG score and a financial or environmental performance measure — where both are produced by methodological choices the researcher does not control and where rating disagreement across providers is well documented outside this corpus. The AI-driven mapping work (Fildisi et al., 2025) is valuable precisely because it treats measurement construction as the problem rather than assuming it away. Zarea et al. (2026), meta-analysing ESG and banking performance at multiple levels, find the heterogeneity one would expect.

Xiong et al. (2025) offer the stream's most identifiable design, using extreme weather events to examine whether supply chain analytics mitigates negative financial effects. Weather is exogenous to firm analytics investment in a way that ESG orientation is not, and this is the template the rest of the stream needs.

5.8 Financial data infrastructure and methods

Shrinking as a share of output, and internally heterogeneous. It contains streaming architecture (Babar, 2025), visualisation systems (Li, T., 2025), enterprise financial shared services (Chen, Tian, & Li, 2025; Wang, 2026), network representations of financial systems (León & Miguélez, 2021; Ahelegbey et al., 2022; Nascimento et al., 2021), and the emerging large-language-model work (Jindal et al., 2026; Saxena & Rishi, 2025).

The LLM subliterature is very new and, at present, more programmatic than evidential. Jindal et al. (2026) survey applications to financial services and risk intelligence; Saxena and Rishi (2025) design an LLM for financial decision support. Neither establishes that language models outperform incumbent methods on a financial task under controlled comparison. Given how quickly this subliterature is likely to grow, the absence of a shared benchmark at this stage is a problem worth naming early.

5.9 Fintech, digital finance and consumer adoption; and data governance

We treat the two smallest streams together because each is small for an instructive reason.

Fintech adoption research (Yuan et al., 2025; Chou et al., 2026; Osakwe et al., 2025; Santini et al., 2026) applies established technology-acceptance models to new financial products. It is methodologically conventional and, at nine studies, surprisingly thin given the commercial salience of the topic — plausibly because the underlying behavioural questions are also studied under household finance and under capability adoption, splitting the stream across our coding boundaries.

Data governance is small for a worse reason. Four studies address privacy and regulation: differential privacy in banking (Espinosa et al., 2026), privacy-preserving computation (Nguyen-Nam et al., 2026), cross-border financial intelligence exchange under GDPR (Haq, 2023), and inferential privacy under Kenya's data protection regime (Murikah et al., 2026). Akter et al. (2023) address algorithmic bias, though in a marketing rather than financial frame, and Ball et al. (2020) study the organisational tensions created by mandatory public–private data exchange in financial services.

Every predictive model in Sections 5.3 through 5.6 depends on continued lawful access to granular data about identifiable individuals. That access is contested and changing. The field's near-total inattention to the conditions of its own possibility is the most consequential gap the corpus reveals, and — as Section 7 argues — it is also the field's largest untapped source of identifying variation.

6. Three Structural Weaknesses

6.1 The identification gap

Sixty-four in-scope titles, 28.6% of the corpus, make an explicit causal claim. No title names a quasi-experimental design.

The claims are not tentative. They include the impact of digital financial inclusion on regional development quality (Ding & Kang, 2024) and on unemployment (Islam et al., 2026); the effects of big data analytics on decision-making, forecasting, and firm performance (Chatterjee et al., 2023); the influence of digital transformation and budgetary control on SME financial performance (Ebhota et al., 2024); the role of big data analytics in fraud detection (Triyanto & Ali, 2026); and the mechanisms by which green finance enhances urban ecological resilience (Yang, Yu, & Zhang, 2026). Each asserts that one thing produces another.

The designs available to support such claims — difference-in-differences around a policy change, instrumental

variables, regression discontinuity, event studies, synthetic controls — appear in no title in the corpus. Two titles use the word "causal," and both belong to the same pre-registered operational-risk study (Cornwell et al., 2023a, 2023b).

We are careful about what this shows. Titles systematically understate methods, particularly in journals with restrictive title conventions, and several studies in this corpus may well deploy identification strategies that their titles do not advertise. The count is a lower bound and should be read as one. But the inference does not depend on the count being exact. In fields where identification is the currency of credibility, it reaches the title routinely, because it is the thing that makes the paper publishable. That it never does so here indicates that identification is not what this field's referees are selecting on.

Figure 5 sets the two trends against each other. The left panel tracks the share of titles each year that name a machine-learning or artificial-intelligence method against the share naming an identification design; the first rises and the second is flat at zero throughout. The right panel states the contrast in levels for the corpus as a whole.

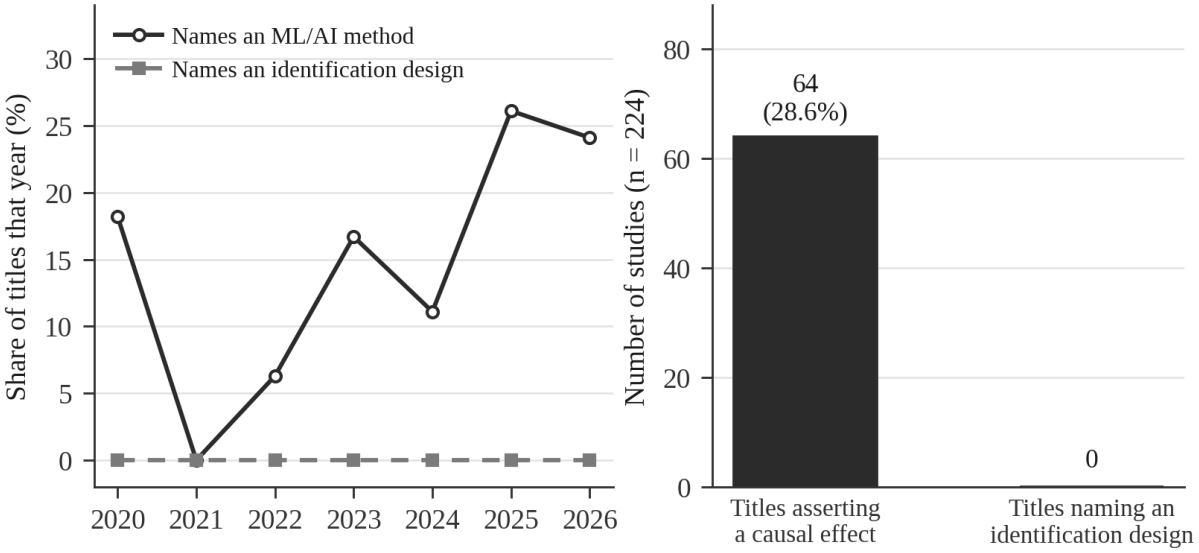


Figure 5. Method salience versus causal ambition

Note. Left panel: share of that year's in-scope titles naming a machine-learning or AI method (solid line) and naming a quasi-experimental identification design (dashed line). Annual shares before 2023 rest on fewer than twenty studies and are correspondingly noisy. Right panel: counts across the full in-scope corpus. Keyword lists for both classifications are reported in Appendix A. Both counts are lower bounds, since titles understate methods.

The consequence is concrete rather than abstract. Consider the most policy-relevant claim in the corpus: that digital financial inclusion reduces unemployment. Regions adopting digital financial infrastructure differ systematically from regions that do not — in urbanisation, mobile penetration, banking density, state capacity, and pre-existing growth trajectory. Every one of these plausibly affects unemployment directly. A panel regression with region and year fixed effects addresses time-invariant differences and common shocks, and leaves untouched the possibility that inclusion arrives first in regions already on an improving trajectory. Without variation that is plausibly independent of that trajectory, the estimate is a description of where inclusion happened, and policy conclusions drawn from it are unwarranted.

6.2 Centrifugal fragmentation

Fragmentation in this corpus is structural rather than merely descriptive.

The venue evidence is in Section 4.2: 124 outlets for 224 studies, 68.5% of outlets appearing once, and no journal exceeding 6% of output. Three mechanisms follow from this. First, referee pools do not overlap, so a study weak by finance standards may be evaluated by information-systems referees, and vice versa. Second, methodological norms cannot propagate, because there is no venue whose standards other venues track. Third, rival findings do not meet: the audit-adoption literature's finding that adoption responds to legitimacy pressure (De Santis & D'Onza, 2021) directly undercuts the capability stream's assumption that adoption responds to expected returns, and the two are published in journals that do not read each other.

The construct evidence is in Section 2. When "financial data analytics" can denote fraud detection, curriculum design, and the household financial consequences of chronic illness, the term coordinates nothing. A quarter of

retrieved records had to be screened out on these grounds.

The synthesis evidence is in Section 4.5. Seventeen studies (7.6%) identify as reviews, bibliometric mappings, or meta-analyses. In a field of this age, that is a high ratio, and it is what one observes when researchers sense incoherence and respond by mapping rather than by testing. This paper is, we acknowledge, the eighteenth such study.

6.3 The abandoned foundational question

The capability–performance stream fell from 31.0% of output in 2020–2022 to 15.2% in 2025–2026. The decline would be unremarkable if the question had been answered.

It has not been. The stream reports overwhelmingly positive effects using designs that cannot distinguish those effects from reverse causality — profitable firms fund analytics — or from common-method variance, since capability and performance are frequently measured perceptually in the same instrument from the same respondent. Its own dissenters make this case. Ain et al. (2025) find no settled way to measure analytics success. Oduro et al. (2023) find heterogeneity large enough to undermine any pooled estimate. Farouk et al. (2025) argue that capability without activating conditions produces nothing, which implies specification error throughout. Zhao, Xie, et al. (2026) query the relationship directly in a port setting.

What should have followed is a consolidation phase: agreement on measures, archival rather than perceptual outcomes, and designs exploiting variation in analytics adoption that is not chosen by the firm. What followed instead was reallocation to ESG, fraud, and household applications — newer questions, with newer data, and the same designs.

This is the pattern Section 6.2 predicts. In a fragmented field, no venue has the standing to declare a question unresolved and demand better evidence before publishing more of the same. Attention moves to whatever is novel, and the foundational claim on which the field's practical relevance rests remains, on the evidence assembled here, unestablished.

7. A Research Agenda

Recommending better identification is easy and useless without saying where the variation would come from. This section names shocks available in financial data analytics research and the design each supports. All are drawn from institutional features of the sector, and each is observable in the settings the corpus already studies.

7.1 Staggered data-protection and disclosure regimes

Data protection law has arrived in different jurisdictions at different times: GDPR across the EU, and subsequently comparable regimes in Brazil, India, Kenya, South Africa, China, and a growing set of US states. These regimes directly change what financial firms may do with customer data.

This is the classic staggered-adoption setting. The treatment is legislated at a level above the firm, adoption timing is driven by legislative calendars rather than by the analytics maturity of any firm, and both treated and not-yet-treated units are observable. Outcomes are available in archival form: credit approval rates, model performance where disclosed, fraud loss rates, marketing efficiency, and the financial performance measures the capability stream currently obtains from surveys. Estimation should use an estimator robust to heterogeneous treatment effects across cohorts rather than ordinary two-way fixed effects, and pre-trend evidence must be presented with attention to the power of the pre-test, not merely its outcome.

Haq (2023) and Murikah et al. (2026) study these regimes qualitatively. Neither exploits their timing. The variation is sitting unused.

7.2 Open banking and payment infrastructure mandates

Open banking mandates, real-time payment rails, and interoperability requirements have been introduced on staggered schedules across the UK, EU, India, Brazil, Australia, and elsewhere. Each represents a discrete expansion in the data available to competing institutions.

Because these mandates typically apply above a size threshold — institutions above a certain asset level must comply, those below need not — they support a regression-discontinuity design in the size dimension as well as a difference-in-differences design in time. Firms just above and just below the threshold are similar in everything except the obligation. Comparing them identifies the effect of mandated data access on lending volumes, pricing dispersion, entry, and incumbent profitability, which speaks directly to the capability–performance question with none of the perceptual measurement the stream currently relies on.

7.3 Supervisory technology deployment

Supervisory authorities adopt analytics at different times and with different intensity across jurisdictions and, within federal systems, across regulators. Wu (2026) discusses supervisory technology in cross-strait money laundering investigation without exploiting deployment timing.

Staggered SupTech rollout supports difference-in-differences estimation of the effect of regulatory analytics on

detected misconduct, reporting behaviour, and compliance cost. The design requires care with the outcome, because detected fraud can rise when detection improves even as underlying fraud falls — the same interpretive problem that arises when enforcement intensity changes. Distinguishing detection from incidence requires either an outcome unaffected by detection effort or an explicit model of the detection process, and the fraud stream (Section 5.4) has the technical capacity to build one.

7.4 Infrastructure failures as negative shocks

Cloud provider outages, ransomware incidents, core banking migration failures, and data-centre disruptions temporarily remove analytics capacity from specific institutions at specific times. Achuthan and Khobragade (2026) study market reaction to breach disclosure, which is adjacent but distinct.

These events support event-study designs around unanticipated capability loss. The identifying logic is attractive because the timing is genuinely unanticipated by the affected firm, and the shock is to capability specifically rather than to the firm's general condition. Two threats require attention: outages are not random across firms, since poorly managed institutions suffer more of them, which argues for restricting attention to shared-infrastructure failures where the cause is upstream of the firm; and disclosure of an outage is itself a choice, which introduces selection into the observed sample.

7.5 Measurement infrastructure

Three commitments would do more for cumulative knowledge than any single study.

The first is shared benchmarks. The forecasting and fraud streams cannot adjudicate accuracy claims without common evaluation data. Where real data cannot be shared, validated synthetic benchmarks are the alternative, and the corpus already contains the components: synthetic financial time series validation (Naivasha et al., 2026), privacy-preserving computation (Nguyen-Nam et al., 2026), differential privacy for banking (Espinosa et al., 2026), and at least one open behavioural dataset (Muñoz-Guerrero et al., 2026). Assembling these into a benchmark is a coordination problem, not a research problem.

The second is archival replacement of perceptual measures. The capability stream's dependent variable should come from financial statements, not from the respondent who also rated the independent variable. Where capability itself must be measured, observable proxies — data-related job postings, technology spend disclosures, patent and cloud contracting activity — are preferable to self-assessment.

The third is pre-registration for confirmatory work. Cornwell et al. (2023a, 2023b) demonstrate that it is feasible in this domain. In a stream where nearly every hypothesised mediator is confirmed, pre-registration is the cheapest available correction.

7.6 Propositions

We state four propositions that follow from the diagnosis and are testable with the designs above.

P1. Estimated effects of analytics capability on financial performance will be substantially smaller when identified from exogenous variation in analytics adoption than when estimated from cross-sectional variation in self-reported capability.

P2. The effect of analytics capability on performance is conditional on complementary organisational conditions, such that the unconditional effect is close to zero in firms lacking them — the implication of Farouk et al. (2025), stated as a testable restriction.

P3. Data-protection regimes reduce measured predictive model performance in credit and fraud applications while leaving realised loss rates approximately unchanged, because incumbent models are overfitted to features that the regime removes.

P4. Mandated data-sharing regimes reduce the performance advantage of large incumbents relative to entrants, because the advantage derives from proprietary data access rather than from superior analytical capability.

P3 and P4 are the propositions most likely to be wrong, which is the reason to state them.

8. Limitations

The limitations of this review are consequential enough to condition every claim above, and we state them plainly rather than deferring them to a closing paragraph.

The corpus is compiled, not systematically retrieved. It is a purposively assembled reference set rather than the output of a documented PRISMA protocol with reported search strings, databases, and inter-rater screening. Coverage is therefore unknown, and the composition may reflect the assembly process as much as the field. Two specific biases are likely: over-representation of recent work, which search engines surface preferentially, and over-representation of Elsevier and Emerald titles relative to society journals in finance and accounting. The sparse representation of core finance journals reported in Section 4.2 should be read with this in mind — it may partly reflect the corpus rather than the field, and readers should treat that particular finding as the weakest in the paper.

Coding is title- and outlet-level. We did not read full texts. Thematic assignment therefore rests on what titles reveal, and the method counts in Section 4.5 are lower bounds by construction. The identification finding in Section 6.1 is the claim most exposed to this limitation, since a study can implement a difference-in-differences design without naming it in the title. We have argued that the salience of identification in titles is itself informative, but this is an inference about publication norms, not a direct measurement of methods, and a full-text audit of a random subsample would be the natural next step. We expect such an audit would find some quasi-experimental work; we would be surprised if it found much.

The thematic scheme forces single assignment. Many studies belong to two streams — a machine-learning fraud detection paper is both a fraud study and a methods study — and our ordered rules resolve these by fiat. Different orderings would produce different distributions. The direction of the thematic drift reported in Table 4 is robust to plausible reorderings; the exact percentages are not.

Period comparison rests on small early samples. The 2020–2022 subsample contains 42 studies, so shares in Table 4 are estimated imprecisely and single studies move them by more than two percentage points. We have drawn conclusions only from the two largest movements, both exceeding eleven percentage points.

The screening boundary is contestable. Excluding studies where financial variables appear incidentally is a judgement, and a researcher interested in the financial consequences of illness would reasonably draw the boundary elsewhere. Our exclusions and their rationale are stated in Section 3.2 so that the boundary can be redrawn.

None of these limitations is fatal to the central argument, because that argument turns on a pattern too large to be an artefact of coding: a field making causal claims in nearly three of ten titles while never advertising a causal design, dispersed across 124 outlets, abandoning its foundational question mid-stream. But each limitation bounds how precisely the pattern can be described, and we would not defend any single percentage in this paper to the decimal place.

9. Conclusion

This review characterises the financial data analytics literature between 2020 and 2026 as expanding rapidly and consolidating not at all. Output grew sevenfold over the period, and more than half of the corpus was published in its final two years. That growth is distributed across 124 outlets, more than two-thirds of which publish a single study on the subject, and across nine research streams of nearly equal size.

The field's central methodological problem is a mismatch between ambition and design. Nearly three in ten studies claim causal effects in their titles; none advertises a design capable of supporting such a claim. The most consequential expression of this problem is the fate of the field's foundational question. Whether analytics capability improves financial performance was the largest research stream in 2020–2022 and had lost half its relative weight by 2025–2026, not because the answer was established but because attention moved elsewhere. The stream's own internal critics — on measurement, heterogeneity, and the conditions under which capability becomes productive — have not been answered so much as outrun.

The field's most consequential substantive gap is data governance, which accounts for 1.8% of the corpus. Every predictive application reviewed here depends on lawful access to granular data about identifiable people. The conditions of that access are changing across jurisdictions on staggered schedules. This is simultaneously the field's largest blind spot and its largest untapped source of identifying variation, which is a coincidence worth acting on.

Two implications follow. For researchers, the binding constraint is no longer method sophistication but research design; the field has adopted machine learning far faster than it has adopted causal inference, and the returns to a well-identified estimate now substantially exceed the returns to another model architecture. For editors and referees, who collectively determine what gets published in a field with no central venue, the practical lever is to require that causal language in a title be matched by a design in the paper, and to treat perceptual measurement of both sides of a hypothesised relationship as a threat to validity rather than a methodological convention.

We began by noting that publication growth is usually read as consolidation. In this literature it is not. The volume is real, the technical competence is real, and the accumulated stock of credible causal knowledge is considerably smaller than either would suggest.

References

1. Abdelwahed, A. S., Abu-Musa, A. A. E. S., Badawy, H. A. E. S., & Moubarak, H. (2024). Investigating the impact of adopting big data and data analytics on enhancing audit quality. Journal of Financial Reporting and Accounting, 23(2), 472–495.

2. Aboelfotoh, A., Zamel, A. M., Abu-Musa, A. A., Frendy, Sabry, S. H., & Moubarak, H. (2024). Examining the ability of big data analytics to investigate financial reporting quality: A comprehensive bibliometric analysis. Journal of Financial Reporting and Accounting, 23(2), 444–471.
3. Achuthan, K., & Khobragade, S. (2026). Cyberattacks, perception, and market impact: Financial and social media dynamics of breach disclosure. Computers & Security, 165, 104862.
4. Aftabi, S. Z., Ahmadi, A., & Farzi, S. (2023). Fraud detection in financial statements using data mining and GAN models. Expert Systems with Applications, 227, 120144.
5. Ahelegbey, D. F., Cerchiello, P., & Scaramozzino, R. (2022). Network based evidence of the financial impact of Covid-19 pandemic. International Review of Financial Analysis, 81, 102101.
6. Ain, N. U., DeLone, W. H., & Vaia, G. (2025). Measuring the success of business intelligence and analytics systems: A literature review. Technovation, 146, 103277.
7. Aktarujjaman, M., Moniruzzaman, M., Uddin, M. S., Ahmed, A., Ahmed, M., Mridha, M. F., & Aung, Z. (2026). A real-time analytics framework for financial crime detection using sequential and market behavior learning. Decision Analytics Journal, 19, 100716.
8. Akter, S., Sultana, S., Mariani, M., Wamba, S. F., Spanaki, K., & Dwivedi, Y. K. (2023). Advancing algorithmic bias management capabilities in AI-driven marketing analytics research. Industrial Marketing Management, 114, 243–261.
9. Akyildirim, E., Sensoy, A., Gulay, G., Corbet, S., & Salari, H. N. (2021). Big data analytics, order imbalance and the predictability of stock returns. Journal of Multinational Financial Management, 62, 100717.
10. AL-Khatib, A. W. (2022). Intellectual capital and innovation performance: The moderating role of big data analytics: Evidence from the banking sector in Jordan. EuroMed Journal of Business, 17(3), 391–423.
11. Almusallam, N., & Qayyum, J. (2025). A hybrid feature selection and clustering-based ensemble learning approach for real-time fraud detection in financial transactions. Computers, Materials and Continua, 85(2), 3653–3687.
12. Altiero, E. C., Nickell, E. B., Spoolstra, A., Wallace, D. M., & Hornik, S. (2026). Riding the revenue recognition wave: An ASC 606 case using GenAI and data analytics. Journal of Accounting Education, 74, 101040.
13. Anderson, I., & Thoma, V. (2021). The edge of reason: A thematic analysis of how professional financial traders understand analytical decision making. European Management Journal, 39(2), 304–314.
14. Arias-Pérez, J., Coronado-Medina, A., & Perdomo-Charry, G. (2021). Big data analytics capability as a mediator in the impact of open innovation on firm performance. Journal of Strategy and Management, 15(1), 1–15.
15. Babar, M. (2025). A hybrid approach to financial big data analysis using extended ensemble learning and optimized spark streaming. Journal of Open Innovation: Technology, Market, and Complexity, 11(3), 100602.
16. Ball, K., Canhoto, A., Daniel, E., Dibb, S., Meadows, M., & Spiller, K. (2020). Organizational tensions arising from mandatory data exchange between the private and public sector: The case of financial services. Technological Forecasting and Social Change, 155, 119996.
17. Bhangu, K. S., Sehra, S. S., & Singh, J. (2026). A novel quantum-enhanced sentiment analysis for financial news classification. Expert Systems with Applications, 328, 132841.
18. Bialowolski, P., Cwynar, A., & Weziak-Bialowolska, D. (2020). Financial management, division of financial management power and financial literacy in the family context – evidence from relationship partner dyads. International Journal of Bank Marketing, 38(6), 1373–1398.
19. Biondo, A. E., Mazzarino, L., & Pluchino, A. (2024). Trading strategies and financial performances: A simulation approach. International Review of Financial Analysis, 95, 103426.
20. Booker, D. D., Pelzer, J. R. E., & Richardson, J. R. (2023). Integrating data analytics into the auditing curriculum: Insights and perceptions from early-career auditors. Journal of Accounting Education, 64, 100856.
21. Brady, S. S., Arguedas, A., Huling, J. D., Hellemann, G., Lewis, C. E., Fok, C. S., Eeden, S. K. V. D., & Markland, A. D. (2024). Financial strain across 25 years and women's bladder health: A life course perspective. American Journal of Obstetrics and Gynecology, 230(1), 77.e1–77.e12.
22. Broby, D. (2022). The use of predictive analytics in finance. The Journal of Finance and Data Science, 8, 145–161.
23. Caloia, F. G., van Ginkel, K. C. H., & Jansen, D.-J. (2026). Floods and financial stability: Scenario-based evidence from below sea level. Journal of Financial Stability, 84, 101545.
24. Chatterjee, S., Chaudhuri, R., Gupta, S., Sivarajah, U., & Bag, S. (2023). Assessing the impact of big data analytics on decision-making processes, forecasting, and performance of a firm. Technological Forecasting and Social Change, 196, 122824.
25. Chen, Y., Lü, X., Tian, H., & Li, R.-H. (2025). Physics-informed neural network for barrier option pricing in coupled financial quantitative system with varying interest rate and volatility. Engineering Analysis with Boundary Elements, 180, 106457.
26. Chen, Y., Tian, S., & Li, H. (2025). Integrating AI and process reengineering in financial shared services. International Journal of Information Systems in the Service Sector, 16(1).
27. Chou, T.-H., Chang, F.-K., Tran, T. H., & Liang, K. (2026). Unveiling consumer intentions in internet-only banking: A machine learning–assisted analytical approach. Chinese Management Studies, 20(5), 1302–1326.
28. Cornwell, N., Bilson, C., Gepp, A., Stern, S., & Vanstone, B. J. (2023a). Modernising operational risk management

in financial institutions via data-driven causal factors analysis: A pre-registered report. Pacific-Basin Finance Journal, 77, 101906.
29. Cornwell, N., Bilson, C., Gepp, A., Stern, S., & Vanstone, B. J. (2023b). Modernising operational risk management in financial institutions via data-driven causal factors analysis: A pre-registered study. Pacific-Basin Finance Journal, 79, 102011.
30. Cuomo, M. T., Tortora, D., Colosimo, I., Celsi, L. R., Genovino, C., Festa, G., & Rocca, M. L. (2023). Segmenting with big data analytics and Python: A quantitative exploratory analysis of household savings. Technological Forecasting and Social Change, 191, 122431.
31. De Santis, F., & D'Onza, G. (2021). Big data and data analytics in auditing: In search of legitimacy. Meditari Accountancy Research, 29(5), 1088–1112.
32. Dhingra, V., Sharma, A., & Goel, A. (2026). Index tracking in financial markets: A comprehensive review of optimization, statistical, and data-driven modeling approaches. Knowledge-Based Systems, 343, 116011.
33. Ding, G., & Kang, N. (2024). The impact of digital financial inclusion on China's regional disparities in the quality of economic development: Based on the relational data paradigm. Economic Analysis and Policy, 81, 629–651.
34. Ding, S., Cui, T., & Zhang, Y. (2022). Futures volatility forecasting based on big data analytics with incorporating an order imbalance effect. International Review of Financial Analysis, 83, 102255.
35. Dote-Pardo, J., & Espinosa-Jaramillo, M. T. (2026). A review of mathematical models for pricing, risk, and optimization in cryptocurrency analytics. Decision Analytics Journal, 18, 100670.
36. Du, M., Amor, R., Ma, K.-L., & Wünsche, B. C. (2026). Data visualization for improving financial literacy: A systematic review. Visual Informatics, 10(1), 100272.
37. Ebhota, O. S., Hongxing, Y., & Sampene, A. K. (2024). Investigating the influence of digital transformation, budgeting and budgetary control on the financial performance of SMEs. Scientific African, 26, e02429.
38. Edu, A. S. (2022). Positioning big data analytics capabilities towards financial service agility. Aslib Journal of Information Management, 74(4), 569–588.
39. Espinosa, D., Pérez, J. F., & Gauthier-Umaña, V. (2026). Differential privacy strategies for data analytics in the banking sector. Information Sciences, 754, 123606.
40. Farouk, F. M., Schinckus, C., & Smith, S. (2025). Realising value from big data analytics: The process of affordance actualisation. Digital Business, 5(1), 100108.
41. Fildisi, B., Vakaj, E., Dridi, A., Imran, A. S., & Azad, R. M. A. (2025). Integrating AI-driven analytics for enhanced ESG mapping: Aligning local and global perspectives. Sustainable Futures, 10, 101231.
42. Flint, M. S. (2025). Expansion of analytical methods in auditing education. Journal of Accounting Education, 70, 100948.
43. Ghiassi, M., Shah, K., Agrawal, M., & Schmitzer, R. (2026). An intelligent text analytics approach to sentiment-based portfolio decision making in the electric vehicle market. Decision Analytics Journal, 19, 100719.
44. Ghosh, I., & Jana, R. K. (2023). A granular machine learning framework for forecasting high-frequency financial market variables during the recent black swan event. Technological Forecasting and Social Change, 194, 122719.
45. Gu, X., Mamon, R., Duprey, T., & Xiong, H. (2021). Online estimation for a predictive analytics platform with a financial-stability-analysis application. European Journal of Control, 57, 205–221.
46. Gupta, S., Drave, V. A., Dwivedi, Y. K., Baabdullah, A. M., & Ismagilova, E. (2020). Achieving superior organizational performance via big data predictive analytics: A dynamic capability view. Industrial Marketing Management, 90, 581–592.
47. Hamilton, O. S., & Steptoe, A. (2025). Financial stress and sleep duration in immune and neuroendocrine patterning: An analytical triangulation in ELSA. Brain, Behavior, and Immunity, 127, 396–408.
48. Haq, M. Z. (2023). How does the General Data Protection Regulation (GDPR) affect financial intelligence exchange with third countries? Journal of Money Laundering Control, 27(1), 158–170.
49. Hassan, M., & Mertzanis, C. (2026). Digital financial literacy and entrepreneurial self-efficacy: A balanced scorecard analysis of new venture performance. Technology in Society, 88, 103438.
50. Haverila, M., Gani, M. O., Dina, F. A., & Mohiuddin, M. (2024). The impact of user-centric measures on the financial performance in the context of big data marketing analytics (BDMA) applications: The influence of a decision-making role. Journal of Systems and Information Technology, 27(2), 237–264.
51. Haverila, M., Haverila, K. C., Mohiuddin, M., & Su, Z. (2022). The impact of quality of big data marketing analytics (BDMA) on the market and financial performance. Journal of Global Information Management, 30(1).
52. Henderson, R. (2020). Using graph databases to detect financial fraud. Computer Fraud & Security, 2020(7), 6–10.
53. Heo, W., Lee, J. M., Park, N., & Grable, J. E. (2020). Using artificial neural network techniques to improve the description and prediction of household financial ratios. Journal of Behavioral and Experimental Finance, 25, 100273.
54. Hu, Y., Wang, X., Li, L., & Gao, Y. (2026). The impact of green finance on high-quality agricultural development: Evidence from China. China Finance Review International.
55. Iftikhar, A., Do, Q., Stevenson, M., & Aslam, H. (2025). Big data analytics and supply chain learning: A serial mediation model for enhancing resilience and financial performance. European Management Journal.

56. Imjai, N., Swatdikun, T., Rungruang, P., Basiruddin, R., & Aujirapongpan, S. (2024). Empowering generation Z accountants in the era of data complexity and open innovation: Nurturing big data analytics, diagnostic, and forensic accounting skills. Journal of Open Innovation: Technology, Market, and Complexity, 10(2), 100308.
57. Islam, F., Chowdhury, M. A. F., & Abdullah, M. (2026). Impact of digital financial inclusion and technological innovation on unemployment. Journal of Financial Economic Policy, 18(4), 654–680.
58. Itani, O. S., Kalra, A., & Rostami, A. (2024). How does big data affect organizational financial performance in turbulent markets? The role of customer-linking and selling capabilities. Technological Forecasting and Social Change, 201, 123221.
59. Jacobs, M. M., & Ellis, C. (2026). Financial toxicity of stroke: Racial and ethnic disparities among working-age and elderly stroke survivors. Journal of the National Medical Association, 118(4), 716–725.
60. Javed, M. H., & Bhutta, D. N. T. (2026). Climate change risk and financial stability in BRICS countries: The moderating role of climate finance. The North American Journal of Economics and Finance, 84, 102636.
61. Jindal, S., Upadhyay, A., Srivastava, A., Arora, S., & Sharma, V. (2026). Financial services and risk intelligence powered by large language models. In K. J. Singh, S. Mahajan, D. S. Kapoor, & K. Thakur (Eds.), LLMs in practice (pp. 345–364). Elsevier.
62. Juma'h, A. H., Boumediene, S., & Blankenberger, B. (2025). Progressing student learning in auditing analytics: The impact of continuous feedback on learning outcomes. Journal of International Education in Business, 18(4), 465–483.
63. Karim, R., Roshid, M. M., Dhar, B. K., & Waaje, A. (2026). Mapping green fintech and sustainability transitions: A bibliometric analysis of digital finance research. Sustainable Futures, 12, 102074.
64. Kastrup, T., Grant, M., & Nilsson, F. (2024). Data analytics use in financial due diligence: The influence of accounting and commercial logic. Qualitative Research in Accounting & Management, 22(2), 158–185.
65. Khalid, I., & Riaz, A. (2026). Impact of behavioral biases and social factors on financial satisfaction: A moderated-mediated model. Journal of Economic and Administrative Sciences.
66. Koohang, A., Sargent, C. S., Zhang, J. Z., & Marotta, A. (2023). Big data analytics: From leadership to firm performance. Industrial Management & Data Systems, 123(12), 2976–2996.
67. Kumar, P., Pillai, R., Kumar, N., & Tabash, M. I. (2023). The interplay of skills, digital financial literacy, capability, and autonomy in financial decision making and well-being. Borsa Istanbul Review, 23(1), 169–183.
68. León, C., & Miguélez, J. (2021). Securities cross-holding in the Colombian financial system: A topological approach. Studies in Economics and Finance, 38(4), 786–806.
69. Li, J. (2022). Preliminarily explore the steps of financial big data analytics. Procedia Computer Science, 208, 471–477.
70. Li, T. (2025). Design of financial data analysis and visualization system combining fuzzy c-means and convolutional neural network. Systems and Soft Computing, 7, 200309.
71. Liew, A., Boxall, P., & Setiawan, D. (2021). The transformation to data analytics in Big-Four financial audit: What, why and how? Pacific Accounting Review, 34(4), 569–584.
72. Lokanan, M. E. (2026). Using dimensionality reduction to detect financial crime. Journal of Financial Crime, 33(4), 270–286.
73. Luong, N. D. A., & Xie, S. (2026). Explainable ensemble machine learning for financial transaction fraud detection: Insights from XGBoost and deep neural networks. The Journal of Finance and Data Science, 12, 100195.
74. Machado, M. R., Chen, D. T., & Osterrieder, J. R. (2025). An analytical approach to credit risk assessment using machine learning models. Decision Analytics Journal, 16, 100605.
75. Manolakis, E., Bongiorno, C., & Mantegna, R. N. (2026). Physics-informed singular-value learning for cross-covariances forecasting in financial markets. Finance Research Letters, 110, 110602.
76. Manry, D., Huang, H.-W., & Yan, Y.-C. (2023). Financial statement fraud litigation, material weaknesses, and board characteristics. Accounting Research Journal, 36(4–5), 349–368.
77. Maulidi, A., Shonhadji, N., Fachruzzaman, Sari, R. P., Nuswantara, D. A., & Widuri, R. (2023). Are female CFOs more ethical to the occurrences of financial reporting fraud? Theoretical and empirical evidence from cross-listed firms in the US. Journal of Financial Crime, 30(5), 1342–1366.
78. Mishra, B., & Chakraverty, S. (2025). Kernel based physics-informed machine learning for approximating CEV model under nonlinear volatility regimes in real-world financial environments. Chaos, Solitons & Fractals, 200, 117154.
79. Muñoz-Guerrero, L. E., Ceballos, Y. F., & Trejos-Rojas, L. D. (2026). A comprehensive dataset of customer behavior in Latin American fintech: 12-month transactional and demographic data for churn analysis. Data in Brief, 65, 112484.
80. Murikah, W., Nthenge, J. K., & Musyoka, F. M. (2026). Reconciling cultural values with algorithmic profiling through Kenya's data protection lens: A conceptual framework for inferential privacy. Telematics and Informatics Reports, 22, 100318.
81. Naivasha, P., Musumba, G., Gikunda, P., & Wandeto, J. (2026). Model-based evaluation of synthetic financial time series data: A comparative study with multi-metric validation. Array, 29, 100684.
82. Narayanamurthy, G., Jayanth, R. S. S., Moser, R., Schaefers, T., & Nagendra, N. P. (2025). Data-driven digital

transformation for uncertainty reduction – Application of satellite imagery analytics in institutional crop credit management. International Journal of Production Economics, 280, 109498.

83. Nascimento, D. C., Pimentel, B. A., Souza, R. M. C. R., Costa, L., Gonçalves, S., & Louzada, F. (2021). Dynamic graph in a symbolic data framework: An account of the causal relation using COVID-19 reports and some reflections on the financial world. Chaos, Solitons & Fractals, 153, 111440.
84. Nguyen-Nam, T., Tran-Truong, P. T., Nguyen-Ngoc, P., Son, H. X., Phan, T. H. T., Nguyen, T. T., Le, K. H., & Nguyen, C. G. (2026). Privacy-preserving computation in financial systems: A systematic survey and quantitative analysis. Information Systems, 142, 102777.
85. Oduro, S., Nisco, A. D., & Mainolfi, G. (2023). Do digital technologies pay off? A meta-analytic review of the digital technologies/firm performance nexus. Technovation, 128, 102836.
86. Oliveira, R. M. A., Sant'Anna, A. M. O., & Ferreira, P. H. (2025). Complex networks-based anomaly detection for financial transactions in anti-money laundering. Forensic Science International: Digital Investigation, 55, 302005.
87. Ordali, E., & Rapallini, C. (2026). Aging and financial risk-taking: A meta-analysis. Journal of Economic Psychology, 113, 102888.
88. Osakwe, C. N., Ogunmokun, O. A., Elgammal, I., & Kwarteng, M. A. (2025). Individuals' attitudes and their adoption intentions of central bank digital currency: Combining theories and analytics for deeper insights. Technological Forecasting and Social Change, 210, 123857.
89. Oussii, A. A., & Zouari, N. F. (2026). Data analytics adoption in financial auditing: Institutional pressures, conformity and resistance. Journal of Accounting & Organizational Change.
90. Öztürk, M., Yıldırım, D., & Eren, M. (2026). Financial cognition, behavior, and well-being in a generational context: A Bayesian-SEM perspective from Türkiye. Borsa Istanbul Review, 26, 100841.
91. Parlier, J., & Lee, L. (2023). Inventory analytics: A teaching case using Excel and Alteryx. Journal of Accounting Education, 63, 100848.
92. Patro, P. K., Quaye, E., Acquaye, A., Jayaraman, R., & Salah, K. (2025). Supply chain carbon finance indexing with generative AI and advanced data analytics techniques. Journal of Cleaner Production, 502, 145387.
93. Paul, S. (2026). Macro–meso–micro analytics for trade-based money laundering risk screening in Bangladesh: A conceptual entity-graph extension. Journal of Economic Criminology, 13, 100244.
94. Pestere-Akcay, M., & Durusu-Ciftci, D. (2026). A composite financial conditions index for Türkiye: Construction, dynamics, and policy implications. Borsa Istanbul Review, 26(4), 100822.
95. Phalalum, V., & Panmarerng, K. (2026). AI-driven analysis of ESG-linked financial determinants for sustainable growth in Thailand: A hybrid SEM framework. Journal of Open Innovation: Technology, Market, and Complexity, 12(1), 100729.
96. Razaque, A., Beishenaly, A., Kalpeyeva, Z., Uskenbayeva, R., & Nikolaevna, M. A. (2025). A reinforcement learning and predictive analytics approach for enhancing credit assessment in manufacturing. Decision Analytics Journal, 15, 100560.
97. 97. Saeed, M., Adiguzel, Z., Shafique, I., Kalyar, M. N., & Abrudan, D. B. (2023). Big data analytics-enabled dynamic capabilities and firm performance: Examining the roles of marketing ambidexterity and environmental dynamism. Business Process Management Journal, 29(4), 1204–1226.
98. Santini, F. de O., Eberle, L., Ladeira, W. J., Milan, G. S., Graciola, A. P., & Sampaio, C. H. (2023). Trust in the financial services context: A meta-analysis. International Journal of Bank Marketing, 41(7), 1693–1714.
99. Santini, F. de O., Jafar, S. H., Ladeira, W. J., & Rasul, T. (2026). Drivers of robo-advisors adoption in financial services: A meta-analytical study. International Journal of Bank Marketing, 44(3), 439–453.
100. Saxena, A., & Rishi, B. (2025). Designing an artificial intelligence-enabled large language model for financial decisions. Management Decision, 63(12), 4135–4153.
101. Schubart, J. R., Schaefer, E. W., Mills, S. E., Knight, D. R. T., Shen, C., & Francomano, C. A. (2025). The financial impact of Ehlers-Danlos syndromes on patients in the United States in 2022. Mayo Clinic Proceedings: Innovations, Quality & Outcomes, 9(1), 100586.
102. Shahana, T., Lavanya, V., & Bhat, A. R. (2023). State of the art in financial statement fraud detection: A systematic review. Technological Forecasting and Social Change, 192, 122527.
103. Shan, Y., & Richardson, V. J. (2025). Querying and visualizing financial data: A teaching case applying STEM skills. Journal of Accounting Education, 72, 100978.
104. Shehadeh, M., Al-Nadabi, A. K., Yaseen, H., & Migdady, H. (2026). AI-based ensemble learning model for financial leverage prediction and risk management in the banking sector. Journal of Financial Reporting and Accounting.
105. Singh, V., Chen, S.-S., Singhania, M., Nanavati, B., Kar, A. K., & Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries – A review and research agenda. International Journal of Information Management Data Insights, 2(2), 100094.
106. Souto, H. G., & Moradi, A. (2024). Wasserstein distance loss function for financial time series deep learning. Software Impacts, 20, 100639.
107. Sr, K. R. H. (2020). Channels of corruption in Africa: Analytical review of trends in financial crimes. Journal of Financial Crime, 27(1), 294–306.

108. Tiwari, M., Gepp, A., & Kumar, K. (2025). Using data analytics to distinguish legitimate and illegitimate shell companies. Journal of Economic Criminology, 7, 100123.
109. Triyanto, D. N., & Ali, S. (2026). The role of big data analytics in financial fraud detection: A systematic literature review. International Review of Economics & Finance, 109, 105453.
110. Vieira, P. R., Matos, P. V., Shingirayi, B. S., & Korompli, S. (2026). Sentiment analysis of financial news on Bloomberg and its impact on the Euro Stoxx 50. Review of Accounting and Finance, 25(3), 434–455.
111. Wang, W. (2026). Big data-driven efficiency optimization in enterprise financial shared services. Information Resources Management Journal, 39(1).
112. Werner, M., Wiese, M., & Maas, A. (2021). Embedding process mining into financial statement audits. International Journal of Accounting Information Systems, 41, 100514.
113. Wiersema, G., Kleinnijenhuis, A. M., Wetzer, T., & Farmer, J. D. (2023). Scenario-free analysis of financial stability with interacting contagion channels. Journal of Banking & Finance, 146, 106684.
114. Wiseman, D., Reinking, J., Tian, Y., & Hornik, S. (2026). Integrating generative AI, data analytics, and critical thinking into the managerial accounting curriculum: A case-based approach. Journal of Accounting Education, 74, 101041.
115. Wu, Y.-X. (2026). Addressing challenges in investigating money laundering cases facilitated by underground banking across the Taiwan strait: The role of supervisory technology (SupTech). International Journal of Law, Crime and Justice, 85, 100830.
116. Xie, L., Li, Y., & Cao, L. (2026). Open government data and bank risk-taking. Economic Modelling, 164, 107786.
117. Xiong, Y., Han, R., Ma, X., Lam, H. K. S., & Lyons, A. (2025). Mitigating the negative financial effects of extreme weather events through supply chain analytics. International Journal of Production Economics, 279, 109441.
118. Yadav, M., Banerji, P., & Garg, A. (2025). Digital financial literacy and retirement planning with mediating effect of saving behaviour. American Journal of Business, 40(3), 127–143.
119. Yang, R., Yu, L., Zhao, Y., Yu, H., Xu, G., Wu, Y., & Liu, Z. (2020). Big data analytics for financial market volatility forecast based on support vector machine. International Journal of Information Management, 50, 452–462.
120. Yang, Z., Yu, H., & Zhang, J. (2026). Mechanisms of green finance enhancing urban ecological resilience: Leveraging green innovation diffusion and digital financial network. Journal of Innovation & Knowledge, 16, 101017.
121. Yu, L., Li, M., & Liu, X. (2024). A two-stage case-based reasoning driven classification paradigm for financial distress prediction with missing and imbalanced data. Expert Systems with Applications, 249, 123745.
122. Yuan, Y.-P., Tan, G. W.-H., & Ooi, K.-B. (2025). What shapes mobile fintech consumers' post-adoption experience? A multi-analytical PLS-ANN-fsQCA perspective. Technological Forecasting and Social Change, 217, 124162.
123. Zarea, M., Prior, D., & Hegazy, M. (2026). ESG and banking financial performance: A multi-level meta-analysis review. The Journal of Economic Asymmetries, 33, e00467.
124. Zhang, Y., Cui, R., Wei, X., & Chen, T. (2025). How digital financial inclusion enhances agroecological efficiency: Impact effects and mechanism tests. International Review of Economics & Finance, 97, 103755.
125. Zhao, G., Xie, X., Jones, P., & Wu, X. (2026). A critical examination of the impact of big data analytics capability on container port performance. Technology in Society, 88, 103469.
126. Zhao, R. (2026). Green innovation strategy, sustainable financial inclusion, and the firms' green transition. Finance Research Letters, 91, 108827.