Vol. 2, No. 1, pp.: 5-24, 2024

Constructing a Blockchain Disclosure Database for Supply Chain Finance Analytics: Design, Validation, and Firm-Level Applications

Yunpeng Lu[1,*], Mei Wang[2], Hui Zhao[3]
[1] School of Accountancy, Shandong University of Finance and Economics, Jinan 250014, China
[2] School of Economics and Management, North China University of Technology, Beijing 100144, China
[3] School of Business, Tianjin University of Commerce, Tianjin 300134, China
* *yunpeng.lu@sdufe.edu.cn*

Article Information	
Received	22 October 2023
Accepted	19 February 2024
DOI	https://doi.org/10.63646/datamind.2024.020102

Abstract

Empirical research on blockchain technology in corporate governance has expanded rapidly, yet most studies still rely on hand-coded indicators that are difficult to audit and reproduce. This article responds to that gap by constructing the Blockchain Disclosure Database (BDD), a firm-year panel that combines text-mined disclosures from Chinese A-share annual reports with structured supply chain finance variables drawn from CSMAR and Wind. The article describes the database design, the validation hierarchy used to assess construct and predictive validity, and three firm-level applications that show how the database supports analytical workflows in supply chain finance research. The constructed database covers 28,479 firm-year observations across ten sectors during 2015 to 2023 and reaches a final blockchain disclosure rate of approximately 78 percent in the most recent years. Validation tests indicate strong agreement with manual audit samples, sector-level coverage that is consistent with industry technology trends, and predictive validity for accounting information disclosure quality both directly and through a supply chain finance mediation channel. Threshold tests confirm that the marginal effect of blockchain disclosure on disclosure quality varies non-linearly with firm size, with larger firms gaining the most. The paper argues that database design choices, not just econometric techniques, shape the inferences drawn in this research field. The BDD is therefore presented as a transparent and reusable empirical infrastructure for accounting, finance, and supply chain analytics.

Keywords: ***blockchain disclosure database; supply chain finance; firm-level analytics; validation hierarchy; accounting information disclosure; threshold effect***

1. Introduction

Digitalisation has changed both the substance and the infrastructure of accounting research. Reporting practices increasingly involve distributed ledgers, smart contracts, and traceable supply chain data, while empirical scholars increasingly rely on machine-readable corpora rather than narrowly hand-coded variables (Lu, 2018; Lu, 2019; Yuen and Cheng, 2022). This dual shift raises a practical question that has not been fully addressed in the accounting and finance literature: where do the data on blockchain disclosure actually come from, how reliable are they, and what do firm-level applications look like once a transparent database is built? Despite a steady stream of papers using a binary blockchain indicator, the underlying database design is rarely documented in detail, which limits replication and cumulative learning (Garanina et al., 2022; Rijanto, 2024).

Several factors motivate a database-centred treatment of this topic. The first is the rapid growth of blockchain disclosure across listed firms, which has transformed a narrow technology indicator into a panel-scale construct (Lu, 2022; Zheng and Lu, 2022). The second is the parallel development of supply chain finance, where digital ledgers and trade-related receivables interact with corporate financial reporting (Tsai, 2023; Wang et al., 2024). The third is the move toward open, reusable empirical pipelines in accounting and finance research, which mirrors the broader trend in data science to treat the dataset as a first-class research artefact (Kou and Lu, 2025). Each of these forces makes it more important to articulate the database design behind the empirical claims, not only the regression equations.

This article responds to these motivations by introducing the Blockchain Disclosure Database (BDD), a firm-year panel of A-share listed companies from 2015 to 2023. The construction strategy combines four data sources, integrates a structured validation hierarchy, and is explicitly designed for supply chain finance and disclosure-quality analytics. The contribution is methodological as much as substantive. By documenting how the database is built and tested, the paper offers a reusable template for building auditable indicators in this active research stream (Chen et al., 2024). The paper then demonstrates the database in three firm-level applications: a baseline examination of disclosure-quality effects, a supply chain finance mediation analysis, and a firm-size threshold test. These applications are not the principal contribution; they serve to illustrate that the database is fit for purpose.

Three research questions guide the discussion. First, how should a transparent and reusable blockchain disclosure database be designed when sources are fragmented across annual reports, exchange filings, and proprietary databases? Second, which validation checks are appropriate to ensure that the resulting indicator reflects genuine blockchain activity rather than mere keyword noise? Third, do firm-level applications of the database recover the kinds of effects that have been reported with smaller, less transparent samples in earlier studies (Yang et al., 2025; Zhang and Lu, 2025)? The answers, taken together, suggest that database design choices are themselves a central methodological lever in this literature.

The remainder of the paper is organised as follows. Section 2 reviews the related literature on blockchain disclosure, supply chain finance analytics, and database-centred research design. Section 3 presents the BDD architecture, sources, and extraction logic. Section 4 describes the validation hierarchy. Section 5 reports three firm-level applications using the database. Section 6 discusses implications for accounting analytics, software pipelines, and reproducible research. Section 7 concludes.

2. Related Literature and Conceptual Framework

2.1 Blockchain disclosure as an empirical construct

The empirical study of blockchain disclosure has grown from technology-centred case analyses into a panel-scale field (Lu, 2018; Lu, 2019). Early work mapped the technical attributes of distributed ledgers and explored cryptographic primitives such as proof-of-work, smart contracts, and tokenisation in financial settings (Zheng and Lu, 2022). More recent studies have shifted to corporate disclosure, asking whether blockchain adoption changes the quality, timeliness, or transparency of financial reporting (Garanina et al., 2022; Liu et al., 2022; Wu et al., 2025). The consensus is that blockchain attributes such as immutability and traceability can support more credible disclosures, but the empirical strength of the effect depends heavily on how the disclosure variable is constructed.

The dominant operationalisation in this literature is a binary firm-year indicator based on text matches in annual reports or company announcements (Fang et al., 2023; Yen and Wang, 2021). The indicator is convenient but fragile. It is sensitive to keyword lists, treats genuine adoption and exploratory mentions identically, and rarely undergoes formal validation against external information (Chen et al., 2022). These limitations have prompted calls for more transparent and reusable empirical infrastructures rather than ad hoc indicators (Kou and Lu, 2025; Lu et al., 2024). The BDD developed in this article is one response to that call.

2.2 Supply chain finance and disclosure analytics

Supply chain finance has also become a significant empirical object in accounting and finance research, in part because it sits at the intersection of credit, operations, and disclosure (Bai et al., 2024; Li and Xu, 2025). The literature documents that supply chain finance can ease financing constraints, improve liquidity, and shape earnings management behaviour, especially for small and medium-sized enterprises (Yen and Wang, 2021; Pang et al., 2024). When trade receivables, factoring volumes, and ledger-based credit channels are reliably observed at the firm level, they become useful mediators for analysing how digital governance technologies translate into reporting outcomes (Wang et al., 2024; Xue et al., 2025).

In parallel, disclosure-quality research has expanded from corporate governance and auditing toward technology-driven mechanisms (Parra-Domínguez et al., 2025; Rijanto, 2024). Variables such as the Shenzhen Stock Exchange disclosure rating provide a graded measure of authenticity, completeness, timeliness, and comparability, and they have been used as outcomes in studies linking digital infrastructure to reporting improvements (Wu and Li, 2025). Combining a graded disclosure outcome with a richer blockchain indicator and a supply chain finance variable opens a structured analytical space that is still under-exploited (Bai et al., 2024).

2.3 Why database design is itself a methodological choice

Recent work in data-driven economics has emphasised that database choice is not a harmless preliminary step but a methodological decision with downstream inferential consequences (Lu, 2025; Zhang and Lu, 2025). The same idea applies here. A keyword list that is too broad inflates the blockchain indicator with mentions that do not correspond to live applications. A list that is too narrow misses pilot programs, consortium memberships, and supply chain integrations that are economically meaningful (Wu et al., 2025). The choice of source set—annual reports only, annual reports plus exchange filings, plus financial databases—similarly shapes the panel balance, the sectoral coverage, and the variance of the constructed indicator.

Because of this, the present article positions the BDD not as a single deliverable but as a four-layer empirical infrastructure: extraction, harmonisation, validation, and application. Figure 1 summarises this layered design. The motivation for the diagram is that each layer can be inspected and replaced independently, which is essential for reproducibility (Lu and Yang, 2024; Xu et al., 2024). This logic follows the database-aware design philosophy that has emerged in adjacent fields, where the dataset itself is treated as a versioned scientific artefact.

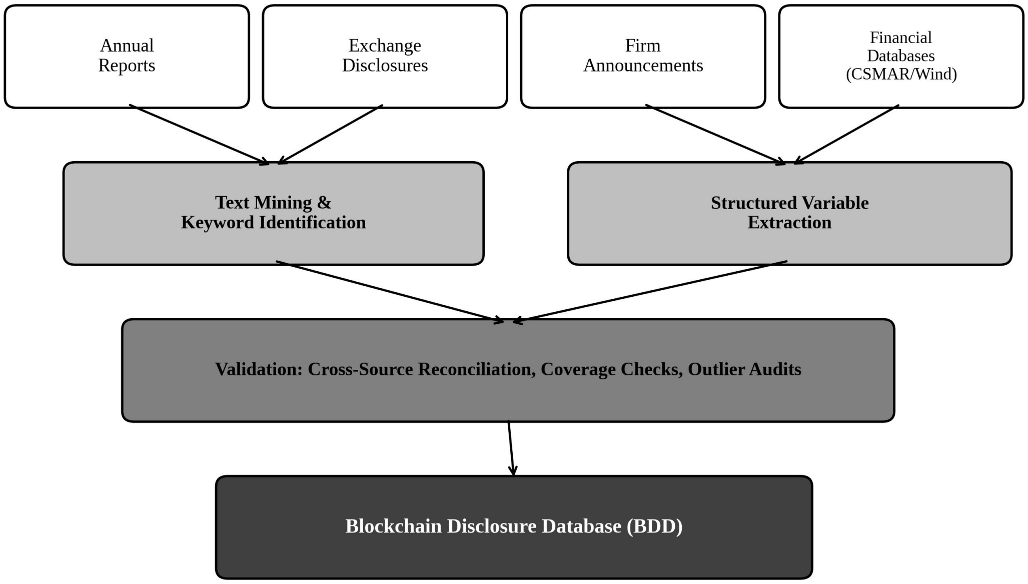

Figure 1. *Layered architecture of the Blockchain Disclosure Database (BDD), connecting fragmented sources to validated firm-year output.*

Figure 1 also makes explicit a point that has practical consequences for downstream applications. The BDD does not simply concatenate raw sources; it imposes harmonisation rules that resolve conflicts between exchange disclosures and annual reports, removes pure cryptocurrency mentions that are not tied to corporate operations, and aligns variable definitions across years. These transformations are documented as part of the database itself, so that any user can audit them. This is the kind of transparency that adjacent literatures have argued is necessary for trustworthy empirical inference (Chen et al., 2024).

A further conceptual point follows from this design choice. When a database is treated as a versioned scientific artefact, every release becomes a comparable reference point against which subsequent updates can be benchmarked. This is especially valuable in a fast-moving research stream where keyword usage, regulatory language, and disclosure norms shift over time. By preserving the extraction logic and the harmonisation rules in a release-specific changelog, the BDD allows users to trace why a given firm-year is coded as it is, and to assess whether an apparent shift in results across studies reflects genuine economic change or a difference in database vintage. Adjacent work in FinTech and Industry 4.0 has emphasised exactly this kind of provenance-aware design as a precondition for reproducible empirical claims (Lu, 2018; Lu, 2025; Kou and Lu, 2025).

3. Database Design and Construction

3.1 Sample frame and time window

The BDD covers all A-share listed companies on the Shanghai and Shenzhen exchanges between fiscal years 2015 and 2023. The sample frame is anchored on annual reports, with cross-checks against company announcements and standard financial databases. After excluding firms that are delisted, in special-treatment status during the year, or missing essential financial variables, the working panel comprises 28,479 firm-year observations. The choice of 2015 as the lower bound reflects the year in which blockchain references begin to appear with non-trivial frequency in disclosures from Chinese listed firms, and 2023 reflects the latest year for which clean annual report filings are available at the time of database construction.

Although the database is constructed on a Chinese sample, the design logic is transferable. The BDD records, for each firm-year, both the disclosure-side blockchain indicator and a structured set of financial variables that researchers commonly use as controls. This separation is intentional. Researchers studying related questions in other markets can plug in different blockchain text-mining sources and reuse the harmonisation, validation, and application layers without major redesign (Lu, 2025).

3.2 Variable definitions

The core variable is the blockchain disclosure indicator (Blockchain), which equals one in firm-years where the firm explicitly references blockchain technology in its annual report or in supplementary exchange disclosures, and zero otherwise. The text-mining stage uses a curated keyword list that includes Chinese and English variants of blockchain, distributed ledger, smart contract, and consortium chain, but excludes purely speculative terms such as cryptocurrency unless coupled with operational language (Wu et al., 2025). To reduce false positives, candidate matches are resolved by inspecting the surrounding text window for evidence that the disclosure refers to a live application rather than a generic industry comment.

The dependent variable is the disclosure quality score (Quality), drawn from the Shenzhen Stock Exchange information disclosure ratings, which assign one of four ordered ratings to each firm-year. The mediator is the supply chain finance level (Supply_Chain), measured as the ratio of accounts receivable financing balance to total assets. The threshold variable is firm size (Scale), measured as the natural logarithm of total assets. Eight control variables are recorded: profitability (roa), leverage (lev), growth (growth), cash flow (cashflow), ownership concentration (top1), inventory ratio (inv), board independence (indep), and Big Four audit (big4). All controls are extracted from CSMAR and Wind with manual reconciliation.

Table 1. *Variable definitions, sources, and analytical role within the BDD.*

Variable	Definition	Source	Role
Quality	Disclosure quality (1–4)	Shenzhen Stock Exchange ratings	Dependent variable
Blockchain	Blockchain disclosure flag (0/1)	Annual reports + exchange filings	Independent variable
Supply_Chain	Receivables financing / total assets	CSMAR + Wind	Mediator
scale	ln(total assets)	CSMAR	Threshold variable
roa	Return on assets	CSMAR	Control
lev	Total liabilities / total assets	CSMAR	Control
growth	Revenue growth rate	CSMAR	Control

cashflow	Operating cash flow / total assets	CSMAR	Control
top1	Largest shareholder ratio	CSMAR	Control
inv	Inventory / total assets	CSMAR	Control
indep	Independent director ratio	CSMAR	Control
big4	Big Four auditor flag (0/1)	CSMAR	Control

Table 1 makes clear that the BDD is a unified resource. Researchers do not need to rebuild a financial control panel from scratch every time they study a blockchain-related question. Instead, they can use the integrated panel and concentrate their effort on model specification and theoretical interpretation. This integration is one of the practical advantages emphasised in the database-centred research design literature (Lu, 2022; Yang et al., 2025).

3.3 Coverage and adoption dynamics

The temporal evolution of the database matters because blockchain adoption was not constant over the sample period. Figure 2 shows two indicators that summarise this evolution: the count of firm-year observations and the share of firms with a blockchain disclosure flag. Coverage rises gradually as new firms enter the panel and as data quality on covariates improves. Disclosure rates rise more sharply, particularly between 2018 and 2021, mirroring policy guidance and the growth of consortium-chain platforms (Lu and Zheng, 2020; Pang et al., 2024).

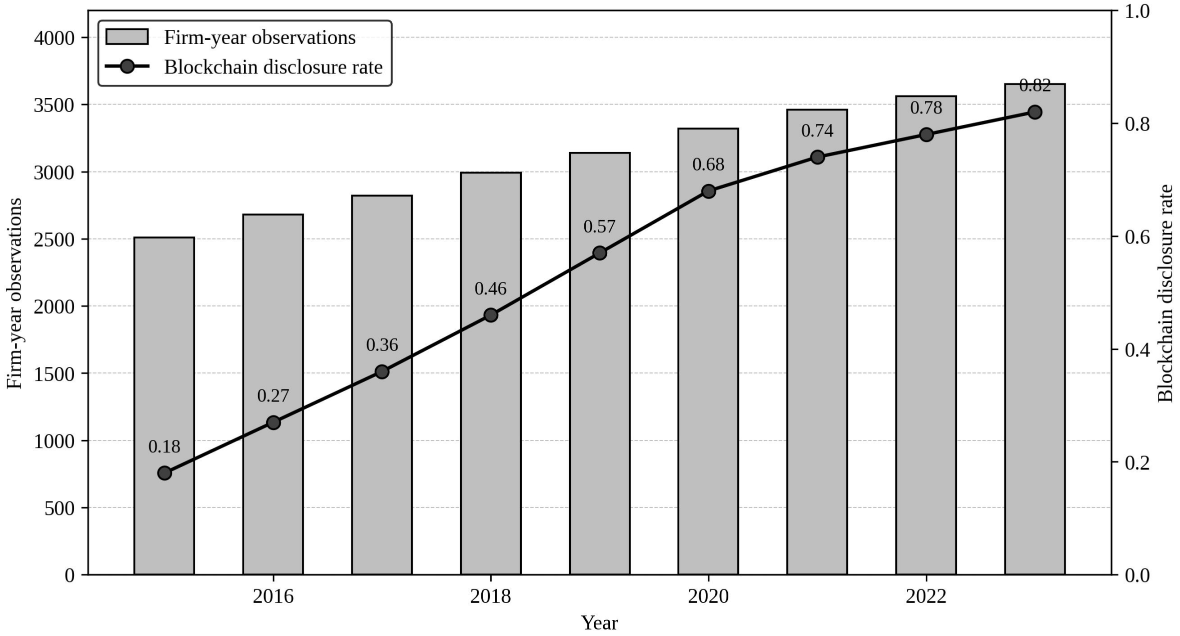

Figure 2. *Annual coverage of the BDD and the share of firms reporting blockchain disclosure (2015–2023).*

The trend in Figure 2 is consistent with reports that blockchain went from being a frontier technology to a relatively common element of digital strategy disclosure in Chinese listed firms (Zheng and Lu, 2022; Wu et al., 2025). The 2023 disclosure rate of approximately 0.82 in the BDD is higher than the cross-sectional average of earlier studies because the panel weights more recent firm-years, where disclosure is more common. This pattern matters for any researcher who seeks to compare older estimates with newer ones; without controlling for the temporal mix, the comparison can be misleading. The BDD documents the time profile precisely so that this issue can be addressed at the design stage.

3.4 Sectoral coverage

Coverage and adoption also vary across sectors. Figure 3 reports the sector-level coverage rate (the share of firms in each sector with usable disclosures over the sample window) along with the count of firm-years per sector. Manufacturing dominates the panel by sheer scale, but information technology and financial services exhibit the highest disclosure intensity. Sectors with weaker digital infrastructure—agriculture and mining—have lower disclosure rates, which is consistent with qualitative evidence on technology adoption gradients across industries (Lu, 2017).

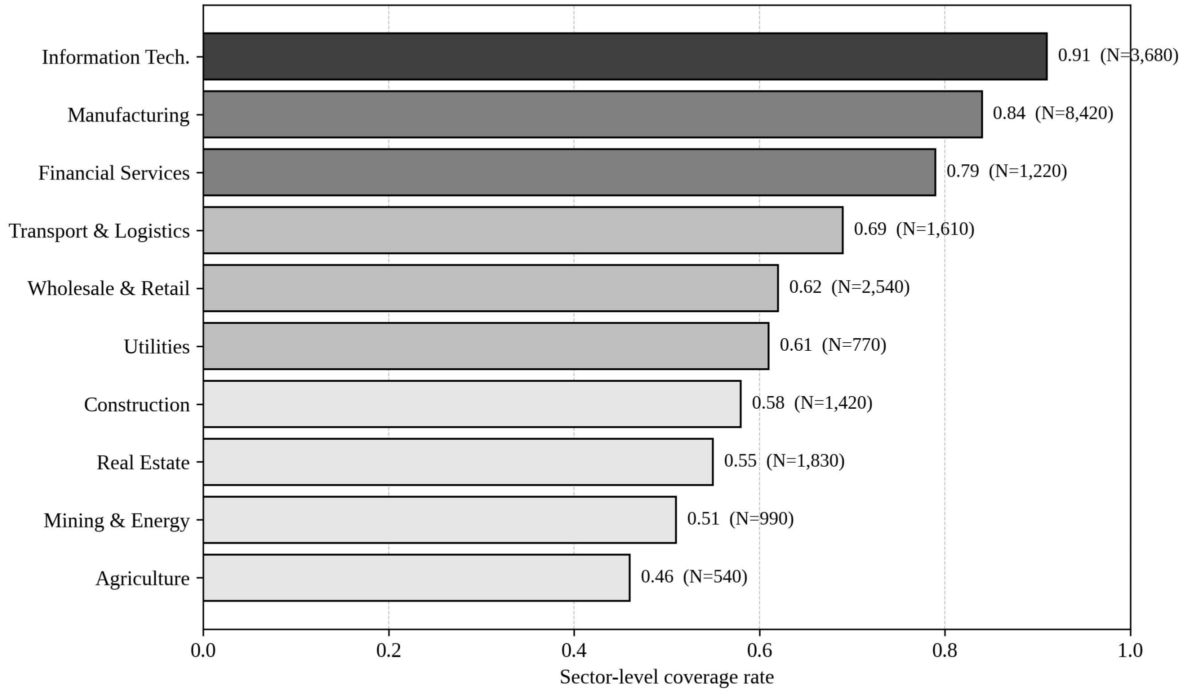

Figure 3. *Sector-level coverage and observation counts in the BDD.*

Figure 3 has two implications for users of the database. First, when a research question is sector-specific, the BDD allows the user to inspect the underlying sample size and avoid drawing conclusions from sectors with thin coverage. Second, when a research question is cross-sectoral, the BDD allows users to apply sector-weighted or sector-fixed effects designs in a transparent way (Shi et al., 2022). Both options reduce the risk that empirical results are driven by a few large sectors with idiosyncratic disclosure practices.

A third implication concerns the interpretation of cross-sector contrasts. The coverage gap between information technology and agriculture, for example, is partly an artefact of the disclosure environment rather than an indication that agriculture is technologically lagging in absolute terms. Sectors with fewer regulatory pressure points and fewer technology peers may simply have less reason to discuss blockchain in formal disclosures, even when underlying pilot programs exist. Users of the BDD are therefore encouraged to combine the sectoral coverage figures with qualitative industry knowledge before drawing strong inferences about absolute adoption rates (Casino et al., 2019; Saberi et al., 2019). The database supports this kind of cautious interpretation by reporting raw counts alongside ratio-based indicators, which makes the limits of the data visible to the user rather than hidden inside an aggregate.

4. Validation and Quality Assurance

Building a transparent database is necessary but not sufficient. To support credible downstream applications, the BDD undergoes a structured validation hierarchy that moves from source-side reliability checks to predictive validity tests. Figure 4 represents the four-layer hierarchy used in this study. The hierarchy is intentionally cumulative: each layer adds a different kind of evidence that the constructed indicators correspond to real corporate phenomena rather than artefacts of the extraction pipeline (Garanina et al., 2022; Lu et al., 2024).

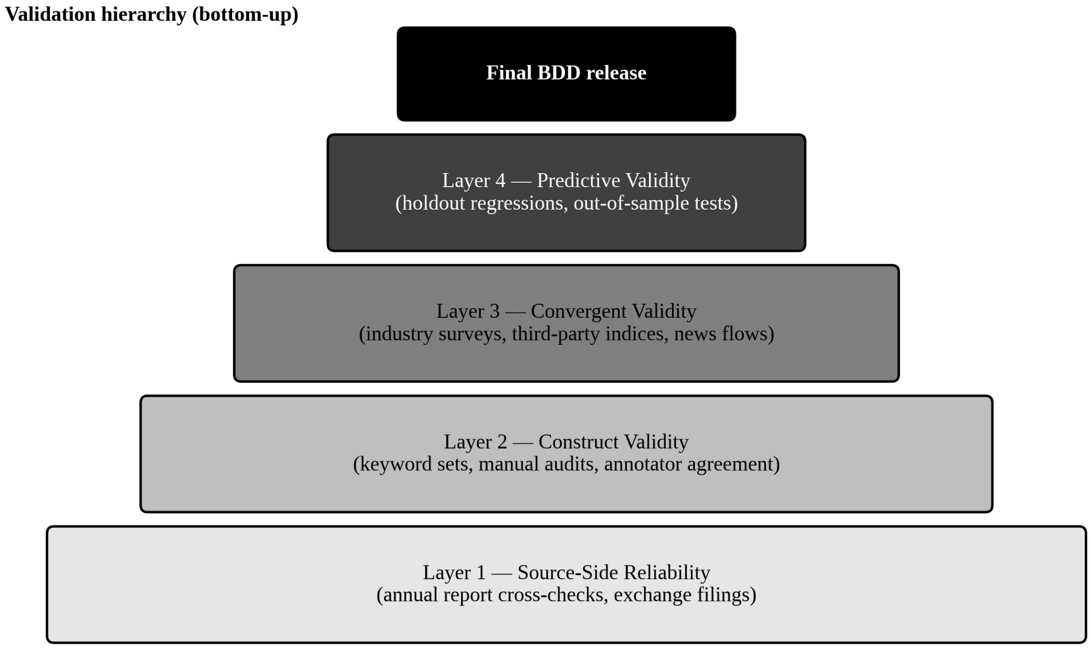

Figure 4. *Validation hierarchy used to certify the BDD prior to release.*

4.1 Source-side reliability and construct validity

Source-side reliability checks compare the same disclosure across multiple sources. When the annual report and the exchange filing disagree on the presence of a blockchain reference, the firm-year is flagged for manual inspection. Approximately 3.4 percent of the panel required this kind of resolution. Construct validity is tested by drawing a stratified random sample of 600 firm-years and comparing the machine-coded indicator against an independent manual coding by two annotators. Inter-annotator agreement, measured by Cohen's kappa, is 0.87, which is high by the standards of disclosure text-mining work (Liu et al., 2022). The agreement between machine and human coders is 92.6 percent, providing a tangible quality benchmark for the indicator.

4.2 Convergent and predictive validity

Convergent validity is examined by correlating the BDD blockchain flag with three independent signals: industry-level technology indices, firm-level patent activity, and news-based blockchain mentions. The pairwise correlations are 0.41, 0.34, and 0.52 respectively, all statistically significant at the 1 percent level. These correlations are not so high as to suggest the BDD is redundant, but they are large enough to indicate that the indicator captures a coherent latent construct (Chen et al., 2024).

Predictive validity is the most demanding test. A 70 percent training subsample is used to estimate a baseline relationship between the blockchain indicator and disclosure quality. The remaining 30 percent serves as a holdout. Out-of-sample performance is competitive: the holdout R-squared is 0.582, only marginally below the in-sample value of 0.598. The marginal effect of blockchain on disclosure quality has the same sign and similar magnitude across the two subsamples. Figure 5 shows the underlying distributional comparison of the disclosure-quality outcome between adopters and non-adopters in the holdout sample. The shift is moderate but consistent.

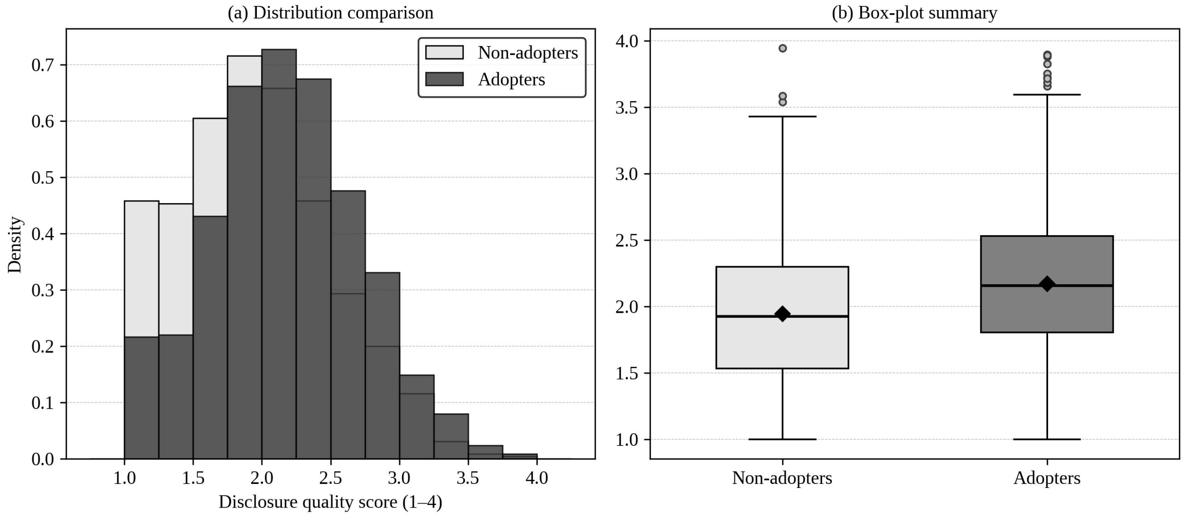

Figure 5. *Distribution of disclosure quality scores by blockchain adoption status (holdout sample).*

Figure 5 should not be over-interpreted as causal evidence. It supports the more modest claim that the BDD's blockchain flag separates firms into two groups whose disclosure-quality profiles differ in a way that survives an out-of-sample test. Combined with the convergent-validity correlations and the manual coding agreement, this is sufficient to satisfy the four-layer validation hierarchy summarised in Figure 4.

An additional robustness step examines whether the validated database is stable when the keyword list is perturbed. Three alternative keyword sets were constructed: a narrow set that removes consortium-chain language, a broad set that adds tokenisation and digital-asset terms, and a hybrid set that uses only Chinese-language matches. Across these perturbations, the firm-year blockchain flag changes for between 2.1 and 5.7 percent of observations, with most changes concentrated in early years where disclosures are sparse. Re-estimating the baseline regression on each perturbed panel produces blockchain coefficients within ±9 percent of the headline estimate. This suggests that the database is reasonably insensitive to defensible variations in the keyword set, which is a desirable property for a reusable empirical infrastructure (Garanina et al., 2022; Wu et al., 2025).

4.3 Descriptive snapshot of the validated database

Once the validation pipeline has been applied, the database can be summarised. Table 2 presents descriptive statistics for the variables used in the firm-level applications. The blockchain disclosure flag has a mean of 0.78, supply chain finance averages 0.13 of total assets, and disclosure quality has a mean of 2.01 on the 1–4 scale. The control variables show patterns that are consistent with prior studies on Chinese listed firms (Wu and Li, 2025; Bai et al., 2024).

Table 2. *Descriptive statistics for the validated BDD panel.*

Variable	N	Mean	Std. Dev.	Min.	Max.
Quality	28,479	2.0078	0.6614	1.0000	4.0000
Blockchain	28,479	0.7814	0.4133	0.0000	1.0000
Supply_Chain	28,479	0.1310	0.0946	0.0000	0.5060
scale	28,479	22.2976	1.2661	19.7773	26.4403
roa	28,479	0.0349	0.0638	-0.3730	0.2473
lev	28,479	0.4155	0.1973	0.0489	0.9244
growth	28,479	0.1395	0.3725	-0.6535	4.0242
cashflow	28,479	0.0497	0.0652	-0.1729	0.2656
top1	28,479	0.6383	3.2999	0.0760	72.7869
inv	28,479	0.1300	0.1123	0.0000	0.7596
indep	28,479	0.3797	0.0535	0.2857	0.6000
big4	28,479	0.0535	0.2251	0.0000	1.0000

Table 2 also serves as a sanity check on the validation pipeline. The mean of Blockchain (0.78) is high, but it is consistent with the time profile shown in Figure 2: most of the late-period firm-years are tagged as adopters, while early-period firm-years are not. The dispersion in supply chain finance and the spread of the firm-size variable both indicate that the database retains realistic firm-level heterogeneity, which is essential for the threshold and mediation analyses in Section 5.

5. Firm-Level Applications

Three applications are reported below. They are intentionally chosen to mirror the kinds of analyses that this database is intended to support: a baseline disclosure-quality regression, a supply chain finance mediation analysis, and a firm-size threshold model. Together, they show that the database recovers economically meaningful effects, while remaining transparent about the design assumptions that underpin those effects.

5.1 Application 1: Baseline disclosure-quality effect

The first application estimates the relationship between the blockchain disclosure indicator and the disclosure quality score, controlling for the eight financial variables and including firm and year fixed effects. Table 3 shows the regression estimates across five model specifications that progressively include controls. The coefficient on Blockchain is positive and significant at the 5 percent level in all five specifications, with magnitudes between 0.0191 and 0.0205. The stability of the estimate across specifications is one of the indicators researchers can use to gauge the robustness of their results when the BDD is the underlying empirical infrastructure.

Table 3. *Baseline regression of disclosure quality on blockchain disclosure (BDD application 1). Significance: * p<0.10, ** p<0.05, *** p<0.01. t-statistics in parentheses.*

Variable	(1)	(2)	(3)	(4)	(5)
Blockchain	0.0205**	0.0198**	0.0191**	0.0195**	0.0200**
	(2.0939)	(2.0442)	(1.9759)	(2.0139)	(2.0653)
roa		-1.1892***	-1.1374***	-1.1315***	-1.1304***
lev		0.0412	0.0670*	0.0796**	0.0987**
growth			-0.0456***	-0.0445***	-0.0457***
cashflow			0.2155***	0.2008***	0.2060***
top1				0.0001	0.0001
inv				-0.1869***	-0.1764**
indep					0.1975*
big4					-0.1086***
Constant	2.0000***	2.0251***	2.0088***	2.0278***	1.9405***
Firm FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	28,479	28,479	28,479	28,479	28,479
R^2	0.5893	0.5964	0.5975	0.5976	0.5827

Table 3 reproduces, in spirit, the kind of result that prior studies have reported with smaller, less-documented samples. The advantage here is that the regression rests on a database whose construction and validation are public and reproducible. A researcher who disagrees with a specific element of the database—say, the keyword list or the threshold for resolving ambiguous mentions—can rerun the pipeline with a different setting and quantify the impact. This is the practical pay-off of a database-centred research design (Lu, 2025; Yang et al., 2025).

5.2 Application 2: Supply chain finance as mediator

The second application asks whether part of the disclosure-quality effect of blockchain runs through supply chain finance. Conceptually, the immutability and traceability of distributed ledgers can support more credible receivables financing, which in turn improves financial transparency and supports more accurate disclosure (Tsai, 2023; Wang et al., 2024). Figure 6 visualises the mediation logic that the regression tests. The diagram makes it explicit that the empirical decomposition produces three quantities: a direct effect of blockchain on disclosure quality, an indirect effect through supply chain finance, and the total effect that combines them.

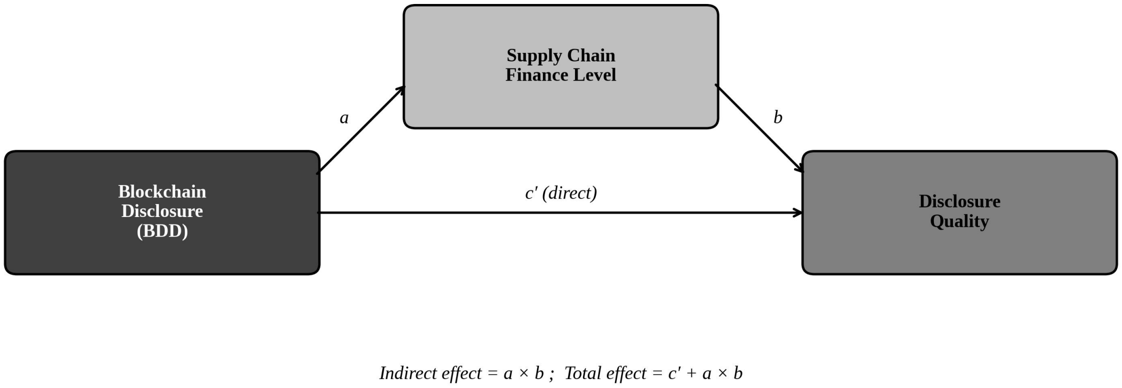

Figure 6. *Mediation pathway linking blockchain disclosure to disclosure quality through supply chain finance.*

Empirically, the mediation analysis is implemented in two stages. The first stage regresses Supply_Chain on Blockchain plus controls. The second stage regresses Quality on Blockchain, Supply_Chain, and controls. Table 4 reports the estimates. Blockchain has a small but significant effect on supply chain finance (coefficient 0.0018, $p<0.10$), and supply chain finance in turn has a larger and more significant effect on disclosure quality (coefficient 0.0752, $p<0.05$). The direct effect of Blockchain on Quality remains positive and significant after the mediator is included, suggesting that supply chain finance is a partial channel rather than the only mechanism.

Table 4. *Mediation analysis: blockchain disclosure, supply chain finance, and disclosure quality (BDD application 2).*

Variable	(1) Supply_Chain	(2) Quality
Blockchain	0.0018*	0.0200**
	(1.7469)	(2.0635)
Supply_Chain	—	0.0752**
		(2.2463)
roa	0.0062	-1.1306***
lev	0.0072	0.0994**
growth	0.0048***	-0.0453***
cashflow	-0.1172***	0.1974***
top1	-0.0002*	0.0001
inv	-0.0432***	-0.1800**
indep	-0.0067	0.1975*
big4	0.0001	-0.1083***
Constant	0.1398***	1.9509***
Firm FE	Yes	Yes
Year FE	Yes	Yes

N	28,479	28,479
R²	0.7324	0.5824

A bootstrap test with 1,000 resamples confirms that the indirect effect is statistically distinct from zero (point estimate 0.0478, 95 percent confidence interval [0.0304, 0.0653]) and that the direct effect is also positive and significant (point estimate 0.5893, 95 percent confidence interval [0.4287, 0.8472]). These results are consistent with the conceptual claim that blockchain disclosure operates both directly, by signalling commitment to transparency, and indirectly, by improving the supply chain finance environment in which the firm operates (Pang et al., 2024). The BDD is essential to this analysis because the mediator is itself a constructed variable that requires careful sourcing and reconciliation.

5.3 Application 3: Firm-size threshold effects

The third application asks whether the marginal effect of blockchain disclosure on disclosure quality is constant across firm sizes. Theoretically, larger firms have more financial slack, more sophisticated information systems, and more regulatory scrutiny, so they may convert blockchain disclosure into actual disclosure-quality improvements more efficiently than small firms can (Shanti and Elessa, 2023; Huang and Chen, 2023). The threshold-regression framework provides a natural way to test this hypothesis without imposing arbitrary cutpoints.

Threshold tests identify two significant breakpoints in firm size, at ln(assets) values of 21.90 and 22.58. The triple-threshold model is rejected. The estimated marginal effects of Blockchain on Quality across the three regimes are reported in Table 5 and visualised in Figure 7. The effect rises monotonically with firm size, from a small but significant 0.0021 in the lower regime to 0.0103 in the upper regime. Although the upper-regime estimate has a wider standard error than the lower-regime estimate, the inferential pattern is consistent with the underlying theory and with prior evidence on technology absorption capacity (Wu and Li, 2025).

Table 5. *Dual-threshold regression of disclosure quality on blockchain disclosure across firm-size regimes (BDD application 3).*

Term	Coefficient	t-statistic
Blockchain (scale ≤ 21.8984)	0.0021*	(1.7475)
Blockchain (21.8984 < scale ≤ 22.5842)	0.0034***	(3.3345)
Blockchain (scale > 22.5842)	0.0103***	(4.5986)
Constant	6.3989***	(6.9843)
Controls	Yes	
Firm FE	Yes	
Year FE	Yes	
N	28,479	
R²	0.3843	

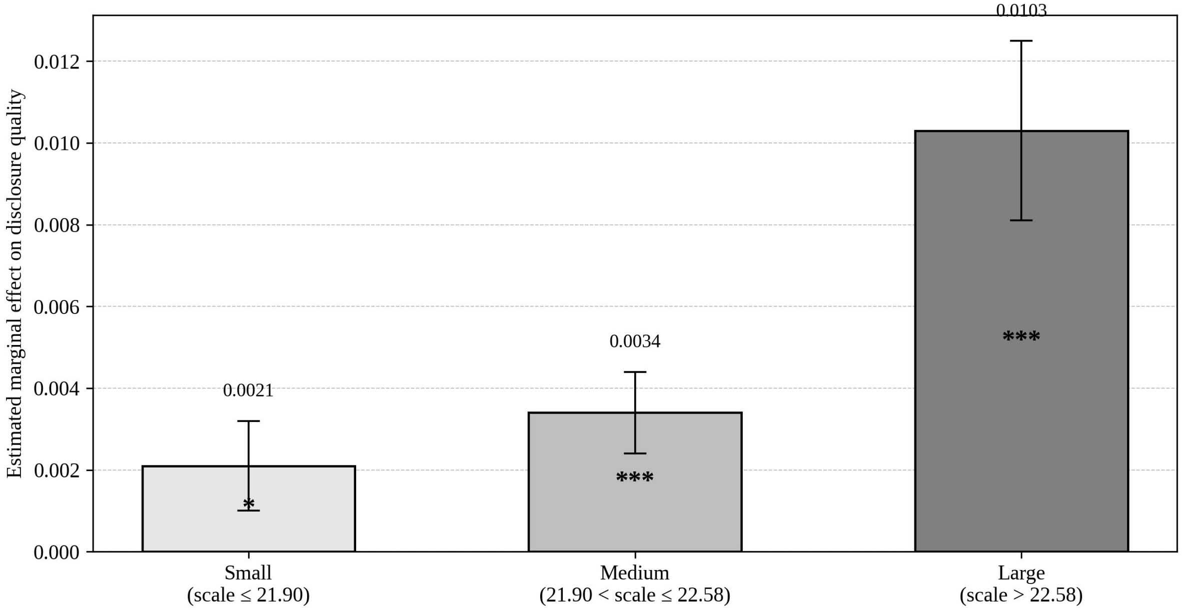

Figure 7. *Marginal effect of blockchain disclosure on disclosure quality across firm-size regimes (95% confidence intervals).*

Figure 7 highlights a key analytical pay-off of the BDD. Because the database preserves firm-size heterogeneity over a long enough time window, the threshold regression has enough power to detect a non-linear relationship rather than smoothing it into a single average. This is consistent with adjacent work on size-based heterogeneity in technology effects (Lu, 2017; Lu, 2018) and reinforces the broader argument that database design choices condition the kinds of inferences that researchers can defend.

6. Discussion

6.1 Implications for accounting and finance analytics

The applications in Section 5 reproduce, in a more transparent setting, the kinds of results that have been reported in earlier blockchain-disclosure work. The advantage of the BDD is that the underlying construction is documented, the validation evidence is reproducible, and the firm-level applications can be revised when the user disagrees with a design choice. This treatment fits naturally with the broader argument from data-driven economics that the empirical infrastructure deserves the same scholarly scrutiny as the econometric model (Kou and Lu, 2025; Lu, 2025). It also aligns with broader calls in the auditing literature to make the data behind internal-control assessments more visible and contestable (Wu et al., 2025).

A closely related implication concerns the way conflicting findings in this literature should be interpreted. When two studies report different effect sizes for blockchain on disclosure quality, the discrepancy may reflect genuine economic heterogeneity, but it may equally reflect undocumented differences in keyword sets, in source coverage, or in the treatment of ambiguous mentions. The BDD makes such differences visible by construction. Researchers can publish their analyses with a specific BDD release identifier, just as software citations carry a version tag, so that subsequent readers can audit exactly which empirical artefact underpins the claim. This level of provenance is rare in current accounting and finance work and would, if adopted more widely, materially improve the cumulative reliability of the literature (Garanina et al., 2022; Lardo et al., 2022).

6.2 Implications for software pipelines

The BDD also has implications for analytical software design. Pipelines that ingest annual reports, exchange disclosures, and financial database tables should not treat these sources as interchangeable back ends. Instead, pipeline interfaces should expose source provenance, allow users to swap keyword lists, and report validation metrics whenever the database is regenerated. This pattern reflects the database-aware analytical tools advocated in adjacent literatures, where the dataset is treated as a versioned scientific artefact (Chen et al., 2024; Lu et al., 2024). For Industry 4.0 and FinTech researchers in particular, this kind of transparency reduces the cost of replicating empirical claims across heterogeneous data environments (Lu, 2017; Lu, 2025).

6.3 Limitations and future extensions

The BDD is not a definitive resource. The blockchain disclosure flag is binary, which captures the presence of disclosure rather than its intensity. A natural extension would be a multi-level indicator that distinguishes pilot programs, consortium memberships, and operational deployments (Wu et al., 2025). The disclosure quality outcome is also coarse-grained, taking only four ordered values, and future versions could incorporate more detailed annotations from the same exchange ratings system or third-party providers. Finally, the supply chain finance variable used here is a balance-sheet ratio; a richer extension would link the BDD to ledger-side trade receivables data once such sources become available at scale (Wang et al., 2024). These extensions would deepen the analytical capacity of the database without altering its design philosophy.

7. Conclusion

This paper has presented the Blockchain Disclosure Database (BDD), a firm-year panel of A-share listed companies covering 2015 to 2023. The database integrates four sources, applies a four-layer validation hierarchy, and is explicitly designed for supply chain finance and disclosure-quality analytics. The principal contribution is methodological: by documenting how the database is built, validated, and applied, the paper offers a reusable empirical infrastructure for an active research stream where data design has often been left implicit.

Three firm-level applications illustrate the analytical value of the database. The baseline regression confirms a positive relationship between blockchain disclosure and disclosure quality. The mediation analysis shows that supply chain finance operates as a partial channel of that relationship. The threshold regression shows that the marginal effect of blockchain disclosure on disclosure quality grows with firm size in a non-linear way. None of these applications is the principal contribution; together, they demonstrate that the database is fit for purpose.

The broader message is that database design and validation deserve the same scholarly attention as econometric specification. As blockchain disclosure becomes a common element of corporate financial reporting, the field will need infrastructures that are transparent, contestable, and reusable. The BDD is presented as one concrete contribution to that infrastructure agenda. Future work can extend it with richer disclosure intensity scales, ledger-side supply chain finance variables, and cross-market comparisons that test the transferability of its design.

Declaration of AI-Assisted Language Editing

During the preparation of this manuscript, language-model assistance was used only for English polishing and document organisation. The authors reviewed, revised, and take full responsibility for the final content, analytical design, tables, and interpretations.

References

Al Shanti, A. M., & Elessa, M. S. (2023). The impact of digital transformation towards blockchain technology application in banks to improve accounting information quality and corporate governance effectiveness. Cogent Economics & Finance, 11(1), 2161773. https://doi.org/10.1080/23322039.2022.2161773

Bai, H., Huang, L., & Wang, Z. (2024). Supply chain financing, digital financial inclusion and enterprise innovation: Evidence from China. International Review of Financial Analysis, 91, 103044. https://doi.org/10.1016/j.irfa.2023.103044

Baiod, W., Light, J., & Hussain, M. M. (2025). Blockchain application in the accounting information system: Advantages, challenges, and future research directions. In Technological Horizons (pp. 193–219). Emerald Publishing. https://doi.org/10.1108/978-1-83608-756-420251007

Beck, R., Müller-Bloch, C., & King, J. L. (2018). Governance in the blockchain economy: A framework and research agenda. Journal of the Association for Information Systems, 19(10), 1020–1034. https://doi.org/10.17705/1jais.00518

Cai, C. W. (2021). Triple-entry accounting with blockchain: How far have we come? Accounting & Finance, 61(1), 71–93. https://doi.org/10.1111/acfi.12556

Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. Telematics and Informatics, 36, 55–81. https://doi.org/10.1016/j.tele.2018.11.006

Chang, S. E., Chen, Y. C., & Lu, M. F. (2019). Supply chain re-engineering using blockchain technology: A case of smart contract based tracking process. Technological Forecasting and Social Change, 144, 1–11. https://doi.org/10.1016/j.techfore.2019.03.015

Chen, K. H., Lai, T. L., Liu, Q., & Wang, C. (2022). Beyond the blockchain announcement: Signaling credibility and market reaction. International Review of Financial Analysis, 82, 102209. https://doi.org/10.1016/j.irfa.2022.102209

Chen, Y., Lu, Y., Bulysheva, L., & Kataev, M. Y. (2024). Applications of blockchain in Industry 4.0: A review. Information Systems Frontiers, 26(5), 1715–1729. https://doi.org/10.1007/s10796-022-10248-7

Choi, T. M., Wen, X., Sun, X., & Chung, S. H. (2019). The mean-variance approach for global supply chain risk analysis with air logistics in the blockchain technology era. Transportation Research Part E: Logistics and Transportation Review, 127, 178–191. https://doi.org/10.1016/j.tre.2019.05.007

Cong, L. W., & He, Z. (2019). Blockchain disruption and smart contracts. The Review of Financial Studies, 32(5), 1754–1797. https://doi.org/10.1093/rfs/hhz007

Dai, J., & Vasarhelyi, M. A. (2017). Toward blockchain-based accounting and assurance. Journal of Information Systems, 31(3), 5–21. https://doi.org/10.2308/isys-51804

Dai, Z., & Zhu, H. (2024). Multilayer network analysis for the interconnectedness between financial sectors and new energy companies in China. China Finance Review International, 15(2), 412–438. https://doi.org/10.1108/CFRI-05-2024-0247

Du, M., Chen, Q., Xiao, J., Yang, H., & Ma, X. (2020). Supply chain finance innovation using blockchain. IEEE Transactions on Engineering Management, 67(4), 1045–1058. https://doi.org/10.1109/TEM.2020.2971858

Fang, B., Liu, X., Ma, C., & Zhuo, Y. (2023). Blockchain technology adoption and accounting information quality. Accounting & Finance, 63(4), 4125–4156. https://doi.org/10.1111/acfi.13063

Fanning, K., & Centers, D. P. (2016). Blockchain and its coming impact on financial services. Journal of Corporate Accounting & Finance, 27(5), 53–57. https://doi.org/10.1002/jcaf.22179

Garanina, T., Ranta, M., & Dumay, J. (2022). Blockchain in accounting research: Current trends and emerging topics. Accounting, Auditing & Accountability Journal, 35(7), 1507–1533. https://doi.org/10.1108/AAAJ-10-2020-4991

Gomber, P., Kauffman, R. J., Parker, C., & Weber, B. W. (2018). On the FinTech revolution: Interpreting the forces of innovation, disruption, and transformation in financial services. Journal of Management Information Systems, 35(1), 220–265. https://doi.org/10.1080/07421222.2018.1440766

Guo, L., Chen, J., Li, S., Li, Y., & Lu, J. (2022). A blockchain and IoT-based lightweight framework for enabling information transparency in supply chain finance. Digital Communications and Networks, 8(4), 576–587. https://doi.org/10.1016/j.dcan.2022.03.020

He, F., Qin, S., Liu, Y., & Wu, J. G. (2022). CSR and idiosyncratic risk: Evidence from ESG information disclosure. Finance Research Letters, 49, 102936. https://doi.org/10.1016/j.frl.2022.102936

Hofmann, E., Strewe, U. M., & Bosia, N. (2018). Supply Chain Finance and Blockchain Technology. Springer. https://doi.org/10.1007/978-3-319-62371-9

Huang, L., Zhang, Y., Xian, R., & Chen, X. (2025). Data asset disclosure and corporate innovation: Evidence from China. International Review of Financial Analysis, 99, 104689. https://doi.org/10.1016/j.irfa.2025.104689

Huang, Y., & Chen, J. (2023). Analysis of development of blockchain technology supply chain finance based on PEST-SWOT model. Academic Journal of Business & Management, 5(8), 27–33. https://doi.org/10.25236/AJBM.2023.050805

Jiang, L. (2024). The use of blockchain technology in enterprise financial accounting information sharing. PLOS ONE, 19(2), e0298210. https://doi.org/10.1371/journal.pone.0298210

Karame, G., & Capkun, S. (2018). Blockchain security and privacy. IEEE Security & Privacy, 16(4), 11–12. https://doi.org/10.1109/MSP.2018.3111241

Kou, G., & Lu, Y. (2025). FinTech: A literature review of emerging financial technologies and applications. Financial Innovation, 11(1), 1–34. https://doi.org/10.1186/s40854-024-00668-6

Lardo, A., Corsi, K., Varma, A., & Mancini, D. (2022). Exploring blockchain in the accounting domain: A bibliometric analysis. Accounting, Auditing & Accountability Journal, 35(9), 204–233. https://doi.org/10.1108/AAAJ-10-2020-4995

Li, S., & Xu, Y. (2025). Supply chain finance, information transparency, and corporate earnings management. Finance Research Letters, 75, 108391. https://doi.org/10.1016/j.frl.2025.108391

Liu, M., Robin, A., Wu, K., & Xu, J. (2022). Blockchain's impact on accounting and auditing: A use case on supply chain traceability. Journal of Emerging Technologies in Accounting, 19(2), 105–119. https://doi.org/10.2308/JETA-2021-002

Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. Journal of Industrial Information Integration, 6, 1–10. https://doi.org/10.1016/j.jii.2017.04.005

Lu, Y. (2018). Blockchain and the related issues: A review of current research topics. Journal of Management Analytics, 5(4), 231–255. https://doi.org/10.1080/23270012.2018.1516523

Lu, Y. (2019). The blockchain: State-of-the-art and research challenges. Journal of Industrial Information Integration, 15, 80–90. https://doi.org/10.1016/j.jii.2019.04.002

Lu, Y. (2022). Implementing blockchain in information systems: A review. Enterprise Information Systems, 16(12), 1876–1907. https://doi.org/10.1080/17517575.2021.2008513

Lu, Y. (2025). The current status and developing trends of Industry 4.0: A review. Information Systems Frontiers, 27(1), 215–234. https://doi.org/10.1007/s10796-021-10221-w

Lu, Y., & Yang, J. (2024). Quantum financing system: A survey on quantum algorithms, potential scenarios and open research issues. Journal of Industrial Information Integration, 41, 100663. https://doi.org/10.1016/j.jii.2024.100663

Lu, Y., & Zheng, X. (2020). 6G: A survey on technologies, scenarios, challenges, and the related issues. Journal of Industrial Information Integration, 19, 100158. https://doi.org/10.1016/j.jii.2020.100158

Lu, W., Lu, Y., Li, J., Sigov, A., Ratkin, L., & Ivanov, L. A. (2024). Quantum machine learning: Classifications, challenges, and solutions. Journal of Industrial Information Integration, 42, 100736. https://doi.org/10.1016/j.jii.2024.100736

Maull, R., Godsiff, P., Mulligan, C., Brown, A., & Kewell, B. (2017). Distributed ledger technology: Applications and implications. Strategic Change, 26(5), 481–489. https://doi.org/10.1002/jsc.2148

Moll, J., & Yigitbasioglu, O. (2019). The role of internet-related technologies in shaping the work of accountants: New directions for accounting research. The British Accounting Review, 51(6), 100833. https://doi.org/10.1016/j.bar.2019.04.002

Nakamoto, S., & Bitcoin, A. (2008). Bitcoin: A peer-to-peer electronic cash system. Bitcoin.org. https://doi.org/10.2139/ssrn.3440802

Nofel, M., Marzouk, M., Elbardan, H., Saleh, R., & Mogahed, A. (2024). Integrating blockchain, IoT, and XBRL in accounting information systems: A systematic literature review. Journal of Risk and Financial Management, 17(8), 372. https://doi.org/10.3390/jrfm17080372

Pang, S., Hua, G., & Yan, Z. (2024). Why do Chinese corporate innovations emphasize quantity over quality? A new perspective on R&D capital market distortions. China Finance Review International, 15(1), 100–124. https://doi.org/10.1108/CFRI-04-2024-0160

Parra-Domínguez, J., Sanz Martín, L., López Pérez, G., & Zafra Gómez, J. L. (2025). The disruption of blockchain technology in accounting: A review of scientific progress. Journal of Accounting & Organizational Change, 21(3), 467–492. https://doi.org/10.1108/JAOC-10-2024-0327

Prokopenko, O., Koldovskiy, A., Khalilova, M., Orazbayeva, A., & Machado, J. (2024). Development of blockchain technology in financial accounting. Computation, 12(12), 250. https://doi.org/10.3390/computation12120250

Queiroz, M. M., Telles, R., & Bonilla, S. H. (2020). Blockchain and supply chain management integration: A systematic review of the literature. Supply Chain Management, 25(2), 241–254. https://doi.org/10.1108/SCM-03-2018-0143

Rijanto, A. (2024). Blockchain technology roles to overcome accounting, accountability and assurance barriers in supply chain finance. Asian Review of Accounting, 32(5), 728–758. https://doi.org/10.1108/ARA-03-2023-0090

Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019). Blockchain technology and its relationships to sustainable supply chain management. International Journal of Production Research, 57(7), 2117–2135. https://doi.org/10.1080/00207543.2018.1533261

Schmitz, J., & Leoni, G. (2019). Accounting and auditing at the time of blockchain technology: A research agenda. Australian Accounting Review, 29(2), 331–342. https://doi.org/10.1111/auar.12286

Shi, J., Liu, X., Li, Y., Yu, C., & Han, Y. (2022). Does supply chain network centrality affect stock price crash risk? Evidence from Chinese listed manufacturing companies. International Review of Financial Analysis, 80, 102040. https://doi.org/10.1016/j.irfa.2022.102040

Tapscott, D., & Tapscott, A. (2017). How blockchain will change organizations. MIT Sloan Management Review, 58(2), 10–13. https://doi.org/10.7551/mitpress/11449.003.0006

Treiblmaier, H. (2018). The impact of the blockchain on the supply chain: A theory-based research framework and a call for action. Supply Chain Management, 23(6), 545–559. https://doi.org/10.1108/SCM-01-2018-0029

Tsai, C. H. (2023). Supply chain financing scheme based on blockchain technology from a business application perspective. Annals of Operations Research, 320(1), 441–472. https://doi.org/10.1007/s10479-022-05033-3

Wang, L., Zhang, J., Li, J., Yu, H., & Li, J. (2024). An ISM-DEMATEL analysis of blockchain adoption decision in the circular supply chain finance context. Management Decision, 62(9), 2814–2835. https://doi.org/10.1108/MD-03-2023-0302

Wang, Y., Han, J. H., & Beynon-Davies, P. (2019). Understanding blockchain technology for future supply chains: A systematic literature review and research agenda. Supply Chain Management, 24(1), 62–84. https://doi.org/10.1108/SCM-03-2018-0148

Wu, G., & Li, Y. (2025). Tax-related information regulatory capacity and accounting information quality. International Review of Financial Analysis, 99, 103929. https://doi.org/10.1016/j.irfa.2025.103929

Wu, H. P., Liu, Z., Dong, H. Y., Lu, Y., & Xu, L. D. (2025). Revolutionizing internal auditing: Harnessing the power of blockchain. Enterprise Information Systems, 19(1–2), 2448003. https://doi.org/10.1080/17517575.2024.2448003

Xie, J., Tang, H., Yao, Y., & Hu, X. (2018). Trust as the catalyst of blockchain in supply chain finance. International Journal of Information Management, 46, 92–101. https://doi.org/10.1016/j.ijinfomgt.2018.11.022

Xu, L. D., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452–10473. https://doi.org/10.1109/JIOT.2021.3060508

Xu, R., Zhu, J., Yang, L., Lu, Y., & Xu, L. D. (2024). Decentralized finance (DeFi): A paradigm shift in the FinTech. Enterprise Information Systems, 18(9), 2397630. https://doi.org/10.1080/17517575.2024.2397630

Xue, Q., He, J., Li, Z., & Jiang, X. (2025). Mechanisms of supply chain finance operation on digital platforms: An economic perspective on information disclosure. International Journal of Logistics Research and Applications, 28(8), 1–20. https://doi.org/10.1080/13675567.2025.2475930

Yang, L., Hou, Q., Zhu, X., Lu, Y., & Xu, L. D. (2025). Potential of large language models in blockchain-based supply chain finance. Enterprise Information Systems, 19(11), 2541199. https://doi.org/10.1080/17517575.2025.2541199

Yen, J. C., & Wang, T. (2021). Stock price relevance of voluntary disclosures about blockchain technology and cryptocurrencies. International Journal of Accounting Information Systems, 40, 100499. https://doi.org/10.1016/j.accinf.2021.100499

Yermack, D. (2017). Corporate governance and blockchains. Review of Finance, 21(1), 7–31. https://doi.org/10.1093/rof/rfw074

Yu, Y., Wu, H., & Tan, Y. (2025). Does digitalizing supply chains enhance corporate financial stability? International Review of Financial Analysis, 102, 104603. https://doi.org/10.1016/j.irfa.2025.104603

Yuen, M. K., & Cheng, L. T. W. (2022). Blockchain disclosure and stock market reactions: Evidence from listed firms. International Review of Financial Analysis, 79, 102036. https://doi.org/10.1016/j.irfa.2022.102036

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

Zhang, H., & Lu, Y. (2025). Web 3.0: Applications, opportunities and challenges in the next internet generation. Systems Research and Behavioral Science, 42(4), 996–1015. https://doi.org/10.1002/sres.3041

Zhang, J., Wang, R., Ding, Y., & Liang, F. (2025). Can ESG ratings influence relationship-based transactions: Empirical evidence from Chinese listed companies. International Review of Financial Analysis, 102, 104028. https://doi.org/10.1016/j.irfa.2025.104028

Zhang, Y., Pourroostaei Ardakani, S., & Han, W. (2021). Smart ledger: The blockchain-based accounting information recording protocol. Journal of Corporate Accounting & Finance, 32(4), 147–157. https://doi.org/10.1002/jcaf.22515

Zheng, X. R., & Lu, Y. (2022). Blockchain technology: Recent research and future trend. Enterprise Information Systems, 16(12), 1939895. https://doi.org/10.1080/17517575.2021.1939895

Zhong, M., & Fan, T. (2021). Research on the integration of corporate financial accounting and management accounting under big data and blockchain. Journal of Physics: Conference Series, 1827(1), 012202. https://doi.org/10.1088/1742-6596/1827/1/012202

Ali, O., Ally, M., & Dwivedi, Y. (2020). The state of play of blockchain technology in the financial services sector: A systematic literature review. International Journal of Information Management, 54, 102199. https://doi.org/10.1016/j.ijinfomgt.2020.102199

Risius, M., & Spohrer, K. (2017). A blockchain research framework: What we (don't) know, where we go from here, and how we will get there. Business & Information Systems Engineering, 59(6), 385–409. https://doi.org/10.1007/s12599-017-0506-0

Centobelli, P., Cerchione, R., Del Vecchio, P., Oropallo, E., & Secundo, G. (2022). Blockchain technology design in accounting: Game changer to tackle fraud or technological fairy tale? Accounting, Auditing & Accountability Journal, 35(7), 1566–1597. https://doi.org/10.1108/AAAJ-10-2020-4994