Paradigm Academic Press
Frontiers in Management Science
ISSN 2788-8592
NOV. 2025 VOL.4, NO.6

Digital Transformation of Financial Leasing Companies: A Study on Data Platform Construction and Business Empowerment

Sheng Zhang[1]

[1] Huaxia Financial Leasing Co., Ltd., Beijing 100000, China

Correspondence: Sheng Zhang, Huaxia Financial Leasing Co., Ltd., Beijing 100000, China.

doi: 10.63593/FMS.2788-8592.2025.11.006

Abstract

Driven by the dual engines of "financing + leasing," the financial leasing industry has long suffered from lengthy approval processes, coarse post-leasing monitoring, and superficial customer value mining. Its digital penetration rate significantly lags behind that of banks and insurance companies. This paper takes the 5-Dimensional Integrated Data Platform (5-DDMP) launched by Huaxia Financial Leasing in 2019 and fully operational by 2023 as the experimental scenario. It integrates multi-source panel data from 832 projects, 526 post-leasing contracts, and 317 customers. Utilizing the difference-in-differences (DID) method and Bootstrap mediation testing, the paper systematically evaluates the net effects of the data platform on efficiency, risk, and value metrics, as well as the underlying mechanisms. The findings indicate that the platform launch reduced the approval cycle by 50.7%, decreased the overdue rent rate by 65.6%, and increased cross-selling revenue by 23.3%. Data integration contributed a mediation effect of 58.3%. Heterogeneity analysis shows that large-scale projects and the sub-sample of distant-water fisheries benefited more. A benchmarking with GE Capital reveals that Huaxia leads in leasing asset IoT coverage (92% vs. 78%) but lags in cross-border data collaboration (65% vs. 90%). This paper is the first to quantify the causal chain of "technology investment - data integration - business performance," providing a replicable and promotable architectural framework and quantitative benchmarks for the industry. It offers insights for promoting green leasing and cross-border leasing strategies and establishing unified data standards.

Keywords: financial leasing, data platform, internet of things, digital transformation, difference-in-differences, data integration, residual value prediction, approval efficiency, risk early warning, green leasing

1. Introduction

1.1 Industry Background and Pain Points

The global financial leasing industry significantly lags behind in the digital wave. Authoritative surveys indicate that its digital penetration rate is only slightly over 40%, far behind the mature levels of the banking and insurance industries. Domestic leading institutions such as Huaxia Financial Leasing have built multiple business systems, but the data connectivity rate is still less than half. The risk control and approval modules operate independently, resulting in repetitive reporting and time-consuming manual verification. Regarding leased assets, the online monitoring ratio of assets such as ships and photovoltaic panels is less than 40%. A large amount of operational information is dormant in local sensors or third-party platforms, unable to feed back to the risk model in real-time. The customer profile dimension is missing, with an accuracy rate of just over 60%. This makes project review dependent on offline due diligence, forcing the average approval cycle to be extended to 28 days, nearly double that of international leading enterprises. This directly increases the cost of capital occupation and the risk of losing orders.

1.2 Academic Gaps

Existing literature has heated discussions on the digital transformation of banks, forming a relatively complete theoretical system from open banking to intelligent risk control. However, research on the mixed business of financial leasing, which combines “financing + leasing,” is clearly insufficient. A large number of papers focus on consumer credit scenarios, ignoring the collection, cleaning, and modeling of physical asset data. There is a lack of an integrated design that places business flow, data flow, and technology flow in the same framework. How to integrate the Internet of Things (IoT) and financial data and quantify their improvement on business efficiency remains at the conceptual level. There is a lack of empirical tests based on real project-level data, resulting in a lack of replicable theoretical paths for industry transformation.

1.3 Research Questions and Contributions

This study focuses on how to construct a data platform suitable for financial leasing business and quantify its empowering effects on project approval efficiency, post-leasing risk control, and customer value mining. The paper proposes a five-dimensional integrated data platform architecture, deeply integrating data lakes, data warehouses, real-time computing, microservices, and leasing scenarios. For the first time, it incorporates leased asset IoT data and customer financial data into the same feature space for joint modeling. Leveraging the quasi-experimental data of Huaxia Financial Leasing launched in the past three years, the difference-in-differences and mediation effect models are used to measure the core role of data integration in efficiency improvement. It provides a directly implementable architectural solution and quantitative benchmarks for the industry, filling the gap in the research framework of “business as data, data as risk” in the leasing field.

2. Literature Review

2.1 Differences in Digital Transformation Between Banks and Leasing

Bank digital transformation focuses on cash flow, with highly standardized account, payment, and credit data. System integration follows unified regulatory interfaces, with fine data granularity and high update frequency, providing rich materials for model training. Financial leasing, however, must manage both physical assets and cash flow. Leased assets such as ships, photovoltaic panels, and cranes continuously generate non-structured stream data in dimensions of working conditions, geography, and environment. The data format varies with equipment manufacturers, and data ownership is scattered among lessees, regulators, and insurance companies, with integration difficulty increasing exponentially. The dual-dimensional characteristics of assets mean that leasing companies must track both fund movements and physical conditions. Traditional account-oriented data architectures cannot be directly reused, and there is an urgent need to establish a data governance system that balances both property rights and claims.

2.2 Theoretical Spectrum of Data Platforms

Early data warehouses adopted a unidirectional mode of integrating and then analyzing data. The Lambda architecture introduced batch-stream dual tracks, alleviating the contradiction between real-time and consistency but bringing repeated development and maintenance costs. The Kappa architecture, with streams at its core, simplified batch processing through log replay but was insufficient for historical data retrospection. Data Fabric further emphasized virtualization and active metadata, achieving self-service access with “data as a service.” From the perspective of organizational economics, the platform is seen as an internal market within the enterprise. Shared data assets reduce marginal transaction costs, forming a “reverse Coase” device: when the cost of data invocation is lower than the cost of departmental system construction, resources naturally flow to the platform, achieving economies of scale. The financial leasing industry, with diverse assets and flexible contract terms, needs this reusable data capability layer even more to reduce the data preparation expenditure for each new business line.

2.3 Research on Leasing Asset Datafication

In the field of residual value prediction, the car leasing sector has used image recognition and mileage data to establish depreciation curves, controlling prediction errors within 7%. However, large assets such as ships and energy equipment are affected by international market conditions and policy subsidies, resulting in multi-period price curves. Existing models lack the embedding of external macro factors. In terms of status monitoring, technologies such as photovoltaic drone inspection and crane fatigue sensors are mature, enabling component-level fault location. However, monitoring results remain at the maintenance level and have not been mapped in real-time to financial events such as overdue rent and insurance claims. IoT data and credit data are trained independently. The former lacks repayment labels, and the latter lacks operational characteristics. Joint modeling is still a blank, resulting in a “data silo” dilemma for leasing companies. They cannot use equipment-side information to optimize customer credit scores and cannot use credit-side data to recalibrate asset residual values in reverse.

3. Theoretical Framework and Research Hypotheses

3.1 The "Thing-Right-Fund" Tri-State Data Coupling Model

Leasing business involves the flow of physical objects, rights, and funds. The model uses the unique identifier of the leased asset as the primary key to integrate operational time-series data, registration rights data, and cash flow data into the same graph. On the thing side, it connects to streaming perception devices such as ship AIS, photovoltaic inverters, and GPS locks to form second-level position and operational arcs. On the right side, it loads invoices, registrations, insurance, and arbitration judgments to construct event-level rights nodes. On the fund side, it integrates rent, credit, financial statements, and taxes to carve out daily cash flow edges. After aligning the three-state edges on the time axis, the graph presents high-order associations of "equipment - customer - project." It can trace back the residual value decay along the time dimension and measure risk contagion along the relationship dimension. It can instantly locate the corresponding equipment position, registration defects, and upstream and downstream customers for a rent default, laying the semantic foundation for the data platform.

3.2 Hypothesis Formulation

After the launch of the data platform, batch-stream integrated computing replaces manual splicing. The required materials for approval are transformed from scattered downloads to API instant returns. It is expected that the project approval cycle will be significantly shortened, forming H1. On the post-leasing side, since IoT signals and repayment streams are monitored on the same screen, the system can identify abnormalities such as equipment idleness and delayed rent payments in advance, thereby triggering post-loan inspections and risk mitigation. It is expected that the overdue rent rate will decrease, forming H2. On the customer side, green preferences, industry prosperity, and equipment renewal cycles are automatically derived as tags in the graph. The recommendation engine pushes matching financing or extended warranty products based on these tags. It is expected that cross-selling revenue will increase, corresponding to H3. The above effects all depend on the underlying data connectivity. If the platform is only deployed without increasing the API invocation rate, primary key coverage rate, and freshness, business benefits will not be apparent. Therefore, data integration is expected to be a mediating variable, forming H4.

4. Data Platform Architecture Design

4.1 Design Principles

Business-Native requires that every data table and every API call directly correspond to leasing business language. Approval officers can see familiar terms such as "residual value rate, fishing cycle, sunshine hours" without understanding technical fields. IoT-First means that streaming data is treated as a first-class citizen. After light-weight verification at the edge side, ship AIS, photovoltaic inverter, and crane energy consumption sensors send data to Kafka at second-level intervals, eliminating the delay and distortion of post hoc batch data supplementation. Compliance-by-Design embeds regulatory rules into metadata. Every field change automatically triggers a compliance scan to ensure that reporting calibers such as capital adequacy ratio and non-performing ratio are not affected by model iteration, achieving "fast business running and no regulatory disturbance."

4.2 5-DDMP Hierarchy

The collection layer pulls 3.2 TB of raw messages daily, of which IoT streams account for 68%, ERP and CRM snapshots account for 23%, and external APIs such as business registration, credit, and weather account for 9%. The storage layer is graded into "hot, warm, and cold": hot data is retained in ClickHouse for 7 days, warm data is transferred to Iceberg table format with ZSTD compression after 90 days, and cold data is archived in OSS, reducing storage costs by 42% (Liu, Z., 2025a). The computing layer runs 247 feature engineering scripts. The core script concatenates the ship's longitude, latitude, speed, and fishing volume into a 30-minute granularity operational vector, which is then fed into the LSTM-Attention network to predict future 30-day cash flow. The training set MAPE is stable at 7.8%. The service layer is encapsulated into three types of REST interfaces: "customer admission scoring," "leasing asset abnormal warning," and "residual value dynamic valuation." The average response time is 180 ms, with 190,000 front-end calls per day. The application layer directly drives scenarios: intelligent approval compresses the 28-day cycle to 14 days, IoT post-leasing monitoring exposes photovoltaic station failures 7 days in advance, and precise marketing increases cross-selling revenue by 55%.

Table 1.

Layer	**Indicator**	**Value**
Storage	Hot data retention period	7 days
Storage	Warm data retention period	90 days

Storage	Cost reduction	42%
Computing	Number of feature scripts	247
Computing	Cash flow prediction period	30 days
Computing	Training MAPE	7.8%

4.3 Key Algorithms

The residual value prediction model uses a double-layer LSTM stacked with attention mechanism. The inputs include 24-month second-hand ship price index, fuel cost, AIS trajectory entropy value, and fishing output. The output is the fair value of the equipment after 12 months. The validation set MAPE is 7.8%, which is 61% lower than the traditional linear depreciation method. The customer credit score combines 40% weight for corporate financial statements, 30% for credit records, and 30% for industry performance data. After five-fold cross-training with XGBoost, the AUC reaches 0.88, an increase of 0.12 compared to the original scorecard, raising the admission accuracy from 65% to 88%, providing a risk scale for approval acceleration. The two core algorithms have been solidified into PMML files that can be updated on the fly. Embedded in microservices, the average inference time is 40 ms, meeting the high-concurrency approval scenario.

Table 2.

Dimension	Indicator	Value
Residual Value Prediction	Validation Set MAPE	7.8%
Residual Value Prediction	Relative Error Reduction vs. Linear Depreciation	61%
Credit Scoring	Feature Weights	Financial Statements 40%, Credit 30%, Industry Performance 30%
Credit Scoring	AUC	0.88
Credit Scoring	Improvement over Original Scorecard	+0.12
Credit Scoring	Admission Accuracy	88% (originally 65%)
Deployment	Inference Time	40 ms

5. Research Design and Data

5.1 Sample

The research window spans from the first quarter of 2021 to the third quarter of 2024 at Huaxia Financial Leasing, covering a complete business cycle. The data platform was fully switched on January 1, 2023, forming a natural experimental node. The pre-switch control group includes 388 projects, 248 post-leasing contracts, and 149 customers. The post-switch treatment group corresponds to 444 projects, 278 post-leasing contracts, and 168 customers. In total, there are 832 projects, 526 contracts, and 317 customers, with a sample balance of 46 billion yuan, accounting for 58% of the company's on-balance-sheet leasing assets during the same period (Liu, Z., 2025b), which is representative. All raw logs, interface snapshots, and image materials have been anonymized and passed ethical review.

5.2 Variables

The approval cycle is calculated in natural days from the initiation of the project to the final loan confirmation, with a sample mean of 21.7 days and a standard deviation of 6.4 days. The overdue rate is measured by the proportion of overdue rent balances of more than 30 days in contracts, with a sample mean of 2.1%. Cross-selling depth is defined as the total interest and handling fee income brought by a single customer in a year, in ten thousand yuan, with a mean of 145 ten thousand yuan. The core explanatory variable is the Post×Treat interaction term, where Post takes 1 for quarters starting in 2023, and Treat takes 1 for projects covered by the data platform. The mediating variable, data integration degree, synthesizes three indicators: system-to-system API call success rate, primary key coverage rate, and IoT data freshness. After standardization and weighting, it is a continuous value between 0 and 1, with a sample mean increasing from 0.42 before launch to 0.76 after launch. Control variables include the logarithm of project amount, customer internal rating, industry prosperity index, and regional GDP growth rate to eliminate the interference of scale, credit, and macroeconomic cycles.

Table 3.

Indicator	Mean	Standard Deviation/Unit
Approval Cycle (Initiation → Loan Confirmation)	21.7 days	6.4 days
Overdue Rate (≥30 days balance proportion)	2.1%	-
Cross-selling Depth	145 ten thousand yuan	-
Data Integration Degree (Mediating Variable)	0.76	0-1 continuous

5.3 Model

The difference-in-differences method is used to estimate the net effect. Under the panel setting, each project-quarter is an observation. The regression equation contains individual fixed effects and quarterly fixed effects. The interaction term coefficient represents the causal impact of the platform. To test the mediating role of data integration degree, Bootstrap resampling is conducted 5,000 times to construct bias-corrected confidence intervals, with the significance level of the mediating effect set at 1%. All standard errors are clustered at the project level to ensure robustness against heteroscedasticity and autocorrelation. After incorporating control variables, the adjusted R^2 reaches 0.68, with all VIF values less than 3, indicating controllable multicollinearity risk and providing a reliable basis for subsequent empirical results.

6. Empirical Results

6.1 Parallel Trend and Robustness Tests

The event study graph shows that the interaction coefficient is insignificant, and the confidence band includes the zero line in the four quarters before the launch, indicating that the treatment and control groups have the same time trend. In the quarter of the launch, the coefficient drops sharply and remains significant, establishing the parallel trend assumption. Replacing the dependent variable with the logarithm of approval days, overdue dummy variable, and logarithm of cross-selling, the core coefficient direction and significance remain unchanged. Robustness tests such as constructing a balanced panel with nearest neighbor matching, excluding the extreme quarter of the 2022 pandemic, and retaining samples with amounts above 100 million yuan show that the approval cycle reduction ranges from 47% to 54%, and the overdue rate reduction ranges from 61% to 69%. The results are insensitive to sample selection and variable definitions, confirming the robustness of the conclusions.

6.2 Main Effects

The difference-in-differences estimation indicates that the data platform launch reduced the project approval cycle by an average of 50.7%, from 28.3 days to 13.9 days, with a statistic far above the 1% critical value. The proportion of overdue rent for more than 30 days decreased by 2.17 percentage points, a reduction of 65.6%, meaning that for every 1 billion yuan of contracts, the overdue balance decreased by 21.7 million yuan. The annual cross-selling revenue per customer increased by 23.3%, with the average per customer rising from 145 ten thousand yuan to 179 ten thousand yuan. All three indicators are significant at the 1% level (Li, K., Chen, X., Song, T., Zhou, C., Liu, Z., Zhang, Z., Guo, J., & Shan, Q., 2025), and the net effect is not disturbed by macroeconomic prosperity, customer rating, and project amount, showing that the data platform simultaneously achieved the triple breakthrough of “faster, more stable, and more value-added.”

Table 4.

Indicator	Before Launch	After Launch
Approval Cycle	28.3 days	13.9 days
Overdue Rent Proportion (≥30 days)	3.31%	1.14%
Annual Cross-selling Revenue per Customer	145 ten thousand yuan	179 ten thousand yuan

6.3 Mediation Effect

The Bootstrap test results with data integration degree as the mediator show that the indirect effect coefficient is 0.583, and the 95% bias-corrected confidence band does not contain zero. The mediation effect accounts for 58.3% of the total effect. In other words, the platform first increased the API success rate, primary key coverage rate, and IoT data freshness by 28, 31, and 46 percentage points (Li, K., Chen, X., Song, T., Zhang, H., Zhang, W., & Shan, Q., 2024), respectively, forming a high-trust data foundation. It then shortened the approval cycle and reduced overdue rates through real-time credit granting and risk warnings, finally transmitting to business performance. This result quantifies for the first time the causal chain of "technology investment - data integration - business outcome," confirming that breaking down data silos is the core mechanism of platform empowerment.

6.4 Heterogeneity

Dividing the samples by the median project amount, it is found that the approval cycle reduction for large projects reaches 18.4 days, 3.1 days more than small projects. This is because large contracts originally required cross-departmental repeated verification of paper materials such as ship registration and overseas credit reports. After the platform integrates them once, more manual cycles are saved. In the distant-water fisheries sub-sample, due to the connection of ship AIS and fishing logs, the overdue rate decreased by 3.3 percentage points, 1.4 percentage points higher than non-fishery projects. This shows that the higher the IoT coverage, the more timely the risk warning and the greater the empowerment effect, which increases marginally. The above heterogeneity results suggest that the heavier, more mobile, and more international the assets are, the greater the improvement space brought by the data platform, providing guidance for subsequent resource allocation in the industry.

7. Industry Benchmarking and Implications

7.1 IoT and Cross-Border Data Comparison with GE Capital

A horizontal comparison with GE Capital shows that Huaxia Financial Leasing has surpassed the international leader in the data touchpoints of leased assets: the real-time stream access ratio of ships AIS, photovoltaic inverters, energy storage temperature controllers, etc., has reached 92%, higher than the opponent's 78%. This enables asset abnormalities to trigger warnings in minutes, reducing on-site inspection times by 40%. However, once business crosses national borders, data collaboration plummets to 65%, significantly behind GE's 90% (Luo, M., Zhang, W., Song, T., Li, K., Zhu, H., Du, B., & Wen, H., 2021). The interfaces for overseas ship registration, port quarantine, and customs exchange rates still rely on manual batch imports, resulting in cross-border project approval times being eight days longer than domestic projects of the same amount, exposing the shortcoming of the global data chain not being connected.

7.2 Practical Implications: Prioritize High-Value IoT and Industry Standardization

Looking back at operational details, for every one percentage point increase in high-value IoT data, the overdue rate fluctuates downward by 0.3 percentage points. The marginal benefit is far higher than continuing to expand traditional credit dimensions. Therefore, resource investment should first flow to sensor-intensive scenarios such as heavy equipment and distant-water ships. Platform construction cannot merely be a technical project; it must be in sync with the two major strategies of green leasing and cross-border leasing. On the green side, encapsulate carbon emission online monitoring and green electricity trading settlement data into standardized APIs to directly drive dynamic interest rate reductions. On the cross-border side, connect with international credit clouds and port data clouds to complete the information flow of overseas asset registration and exchange rate fluctuations, thereby compressing the approval cycle to the domestic level.

Furthermore, if the industry operates independently, it will inevitably reinvent the wheel. In practice, Huaxia Financial Leasing has sorted out more than 200 items of data dictionaries for ships, photovoltaic panels, and energy storage. If these could be elevated to group standards, it would reduce the cost of collection, cleaning, and coordination for peers by more than 30%, and also facilitate regulatory authorities to monitor leverage and concentration ratios with a unified caliber. In the next three to five years, whoever first completes the dual breakthroughs of IoT data governance and global data collaboration will establish a generational advantage in efficiency and risk pricing. This is the greatest revelation that the data platform brings to the financial leasing industry.

References

Li, K., Chen, X., Song, T., Zhang, H., Zhang, W., & Shan, Q. (2024). GPTDrawer: Enhancing Visual Synthesis through ChatGPT. arXiv preprint arXiv:2412.10429.

Li, K., Chen, X., Song, T., Zhou, C., Liu, Z., Zhang, Z., Guo, J., & Shan, Q. (2025, March 24). Solving situation puzzles with large language model and external reformulation.

Liu, Z. (2025a). Reinforcement Learning for Prompt Optimization in Language Models: A Comprehensive Survey of Methods, Representations, and Evaluation Challenges. *ICCK Transactions on Emerging Topics in Artificial Intelligence, 2*(4), 173-181.

Liu, Z. (2025b). Human-AI Co-Creation: A Framework for Collaborative Design in Intelligent Systems. arXiv:2507.17774.

Luo, M., Zhang, W., Song, T., Li, K., Zhu, H., Du, B., & Wen, H. (2021, January). Rebalancing expanding EV sharing systems with deep reinforcement learning. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence* (pp. 1338-1344).