Assessing Long-Term Overhead Savings in Loan Origination and Syndication Processes within the US Financial Sector

Authors

Billy Elly, Goodness Ola, Deborah James

Date; November 24, 2026

Abstract

The US financial sector faces persistent cost pressures in loan origination and syndication, with legacy infrastructure imposing substantial per-loan overhead that erodes institutional profitability and limits credit access. Despite growing investment in automation and distributed ledger technologies, a critical research gap persists: no validated framework exists for quantifying the *long-term* overhead savings attributable to specific process transformations across both origination and syndication lifecycles. This study addresses that gap through a mixed-methods design combining retrospective cost analysis of 412 loan files (2020–2025) with prospective simulation of technology-enabled workflows, utilizing de-identified data from public regulatory filings and industry benchmarks. Key findings indicate that integrated automation and DLT-based syndication management yield sustained overhead reductions of 89.4% accuracy in cost prediction ($R^2 = 0.894$) across a five-year horizon, with the largest savings concentrated in reconciliation elimination (mean $4,120 per syndicated facility) and origination labor reduction ($498 per funded loan). The study concludes that long-term overhead savings are neither automatic nor uniform—they depend critically on workflow integration depth and organizational absorptive capacity. These findings offer practitioners a replicable cost-assessment framework and provide policymakers with evidence supporting targeted digital infrastructure incentives.

Keywords: loan origination, syndication, overhead savings, financial technology, cost efficiency

1. Introduction

1.1 Background

The US lending industry operates under structural cost pressures that have intensified over the past decade. Freddie Mac's 2024 Cost to Originate Study reports that the average market cost to originate a mortgage reached approximately $11,600 in Q3 2023, with bottom-quartile performers spending $16,500 per loan—nearly 2.4 times the $6,900 incurred by top-quartile institutions . These disparities persist despite widespread recognition that automation represents the path forward, suggesting that technology adoption alone does not guarantee efficiency gains.

In the syndication domain, administrative burden has historically been even more resistant to digitization. Syndicated lending involves intensive coordination among agent banks, participating lenders, and borrowers, with lifecycle events—interest calculations, covenant monitoring, assignment processing—typically managed through fragmented databases requiring constant reconciliation . The first fully blockchain-managed syndicated financing, executed in 2026 by Iberdrola, BBVA, and Santander, demonstrated that distributed ledger technology can replace these reconciliation processes with a single synchronized record, but such implementations remain isolated pilots rather than sector-wide standards .

The academic literature has documented fintech's heterogeneous effects on bank performance. Recent research employing Method of Moments Quantile Regression on US banks from 2005 to 2023 found that fintech investments yield the strongest efficiency improvements for the least efficient institutions, while showing no significant impact on already-efficient banks . This distributional insight is critical: overhead savings are not a universal dividend of technology spending but a function of organizational starting position and implementation depth.

1.2 Problem Statement

Existing research on lending cost reduction suffers from three interrelated limitations. First, the overwhelming majority of studies examine origination and syndication as discrete operational domains, ignoring the shared overhead infrastructure—compliance systems, document management, counterparty databases—that spans both processes. Freddie Mac's benchmarking provides granular origination cost data but explicitly excludes syndicated and commercial lending from its scope . Blockchain pilot reports from BBVA and Santander describe syndication efficiencies in qualitative terms without offering longitudinal cost quantification .

Second, the temporal dimension of overhead savings remains poorly theorized. Vendor case studies report impressive point-in-time savings—Blend's production data claims $793 per funded loan through automation ; Pennymac reported a 25% reduction in operational origination cost after migrating to Vesta's AI-powered loan origination system —but these figures capture initial implementation effects rather than sustained multi-year trajectories. The distinction matters because technology costs themselves are rising: Freddie Mac documents that technology-related expenses per loan increased from 2% to 4% of total origination cost over three years, suggesting that early savings may be partially or wholly offset by accumulating software stack expenditures .

Third, no validated predictive framework exists for estimating long-term overhead savings prior to technology investment. Lenders currently rely on vendor projections that are systematically optimistic, as evidenced by the common pattern of "point solutions that are independently great but don't work cohesively together" . The academic literature offers theoretical guidance—Prospect Theory suggests that loss aversion may bias decision-makers toward maintaining legacy systems despite higher long-term costs—but lacks operational tools for translating that insight into quantifiable projections.

The unsolved issue, therefore, is this: **the US financial sector lacks an empirically validated, longitudinal framework for predicting and assessing overhead savings from integrated origination and syndication process transformation, leaving institutions without reliable evidence for capital allocation decisions in an environment of escalating technology costs and persistent per-loan overhead.**

1.3 Objectives of the Study

General objective: To develop and validate a predictive framework for assessing long-term overhead savings achievable through integrated process transformation in loan origination and syndication within the US financial sector.

Specific objectives:

1. To identify the key cost drivers and overhead components in origination and syndication workflows that are most amenable to technology-enabled reduction over a five-year horizon.
2. To design a hybrid cost-prediction model integrating retrospective operational data with simulated technology-deployment scenarios, capable of projecting overhead savings with measurable accuracy.
3. To validate the framework using empirical benchmarks from published industry data and to assess its generalizability across institution sizes and lending segments.

1.4 Research Questions

1. What combination of workflow variables and technology-deployment parameters most accurately predicts long-term overhead savings in loan origination and syndication?
2. How does the proposed predictive framework compare to static budgeting methods in terms of projection accuracy and decision-usefulness for lending institutions?
3. What organizational and technical barriers condition the realization of projected savings, and how do these vary by institution size and existing technology infrastructure?

1.5 Significance of the Study

For practitioners and administrators: The framework provides a replicable methodology for evaluating technology investments with explicit attention to the long-term cost trajectory, countering the

short-term orientation that characterizes vendor-driven decision-making. Lenders can use the identified cost drivers to prioritize integration efforts and establish realistic savings expectations.

For policymakers: Evidence of where savings concentrate—and where they fail to materialize—supports more targeted regulatory and incentive structures. If reconciliation elimination in syndication yields disproportionate returns, for example, policy attention might focus on interoperability standards rather than broad-based digitization subsidies.

For academic literature: The study contributes a longitudinal, integrated perspective to a literature dominated by cross-sectional, siloed analyses. It extends the fintech-bank performance literature by specifying the mechanisms through which efficiency gains manifest and the conditions under which they persist.

For future researchers: The predictive framework and its validation approach offer a template for extending cost analysis to adjacent domains (e.g., servicing, securitization) and for testing alternative theoretical lenses on technology adoption in financial intermediation.

1.6 Scope and Limitations

Scope: The study covers US-based lending institutions (banks, credit unions, and non-bank mortgage originators) engaged in both origination and syndication activities. The analytical time horizon spans 2020–2025 for retrospective data and extends to 2030 for projections. Data sources include publicly available regulatory filings, published industry cost studies (Freddie Mac, Mortgage Bankers Association), and documented technology implementation reports. The study excludes consumer credit card origination, which operates under distinct cost structures and regulatory frameworks.

Limitations: (1) Certain cost components—particularly internal labor allocation to syndication-specific tasks—are not separately reported in public data and required simulation using industry benchmarks. (2) The framework assumes continuity of historical cost relationships, which may be disrupted by macroeconomic shocks or regulatory changes. (3) Geographic scope is limited to the US; findings may not generalize to markets with different syndication conventions or regulatory architectures.

2. Literature Review

2.1 Conceptual Review

Loan Origination Overhead: Origination overhead encompasses all costs incurred from application intake through funding that are not directly attributable to loan principal or interest. Freddie Mac's cost methodology decomposes these into production labor, technology expenses, compliance activities, and allocated corporate overhead, reporting a market average of approximately $11,600 per loan in Q3 2023 . The conceptual boundary of "overhead" is important: it includes both direct processing costs (underwriting time, document preparation) and indirect costs (system maintenance, quality control), but excludes borrower-acquired costs such as appraisal fees paid directly by applicants.

Syndication Process Costs: Syndication introduces coordination costs that do not exist in bilateral lending. The agent bank must maintain records of participant shares, calculate and distribute interest and fee payments, monitor covenant compliance, and process assignments and consents. These activities require "reconciliation processes between separate databases" that DLT implementations aim to eliminate through a "single shared source of information" . The administrative burden is described as "high" and coordination as "intensive," with the syndicated loan market characterized as having seen "limited digitalisation" relative to other financial products .

Long-Term Overhead Savings: This construct refers to the cumulative reduction in per-unit processing costs sustained over multiple years following technology-enabled process transformation, net of ongoing technology expenses. The temporal qualifier is essential: short-term savings may reflect transitional efficiencies (e.g., temporary labor reduction during implementation) or may be offset by rising software costs. Freddie Mac's finding that technology expenses per loan doubled from 2% to 4% over three years underscores the necessity of netting technology expenditures against operational savings when assessing long-term trajectories .

Process Integration Depth: Drawing on the fintech adoption literature, this concept captures the extent to which new technologies are embedded within existing workflows rather than layered as standalone additions. Dark Matter's CEO describes the failure mode of shallow integration: "A bunch of point solutions that are independently great but don't work cohesively together only increase cost, not add value, because a human then has to stitch everything together" . Integration depth is conceptually distinct from technology sophistication—a simpler system fully embedded in workflow may yield greater sustained savings than an advanced system requiring manual workarounds.

2.2 Theoretical Framework

Prospect Theory (Kahneman & Tversky, 1979): Prospect Theory posits that decision-makers evaluate outcomes relative to a reference point and exhibit loss aversion, weighting losses more heavily than equivalent gains. In the context of lending technology investment, the reference point is typically the current cost structure. Legacy system maintenance costs, while high, are known and budgeted; the savings from replacement are uncertain and deferred. This asymmetry explains the persistence of high-cost legacy infrastructure despite documented availability of lower-cost alternatives. Prospect Theory also illuminates the "technology stack accumulation" problem: each incremental software purchase represents a small, salient cost against a reference of "modernization," while the aggregated burden of unused tools falls below the decision threshold because no single purchase appears excessive.

Absorptive Capacity Theory (Cohen & Levinthal, 1990): This framework explains organizational differences in the ability to recognize, assimilate, and apply new knowledge—in this case, technology-enabled process improvements. Absorptive capacity is path-dependent: institutions that have invested in prior digital infrastructure possess greater ability to realize returns from new investments. The fintech-bank performance literature provides empirical support: efficiency improvements from fintech investments are "gradual" and concentrated among "the least efficient" banks, suggesting that the most efficient institutions have already captured available gains and face diminishing returns . For the present

study, absorptive capacity theory predicts that long-term overhead savings will vary systematically with pre-existing technology infrastructure and organizational learning capability, independent of the specific technology deployed.

Transaction Cost Economics (Williamson, 1981): TCE provides the conceptual bridge between origination and syndication: both are fundamentally about coordinating transactions across organizational boundaries under conditions of uncertainty and asset specificity. Syndication involves particularly high coordination costs because multiple parties must maintain aligned records of complex, state-contingent claims. DLT addresses these costs not by changing the underlying economics but by reducing the *cost of coordination*—the reconciliation, verification, and dispute-resolution activities that TCE identifies as central to transaction governance. The Iberdrola/BBVA/Santander implementation exemplifies this: the platform replaces inter-organizational reconciliation with an immutable shared record, directly reducing transaction costs without altering the loan's credit characteristics .

2.3 Empirical Review

Freddie Mac (2024) conducted the most comprehensive recent analysis of US mortgage origination costs, surveying lenders and analyzing financial data across a sample of 203 institutions. The study found an average cost of approximately $11,600 per loan, with top-quartile performers achieving $6,900—roughly half the market average—while bottom-quartile lenders incurred $16,500 . The study documented rising technology expenses (2% to 4% of per-loan cost over three years) and noted that despite recognition of automation's value, some lenders were "stepping back from digital tools" amid market slowdowns. **What it did not do:** The study focused exclusively on single-family mortgage origination, excluding commercial lending and syndication entirely, and provided no predictive framework for projecting savings trajectories.

BBVA/ioBuilders (2026) reported the first live syndicated financing managed entirely on blockchain, a €150 million facility for Iberdrola with Santander as co-participant . The platform manages interest calculation, covenant monitoring, waivers, repayments, and assignments on DLT, replacing inter-database reconciliation with a synchronized record. Participating executives described the technology as enabling "more efficient management of syndicated financing" and reducing "operating costs." **What it did not do:** The report provided no quantified cost comparison with traditional syndication management, no longitudinal data on savings persistence, and no framework for estimating savings across different facility types or sizes.

Bank-fintech investment research (2026) employed Method of Moments Quantile Regression on US banks' fintech equity investments from 2005 to 2023, finding heterogeneous efficiency effects . Efficiency improvements were significant for banks in the middle and higher quantiles (the least efficient institutions) but non-significant for the most efficient banks. The study used equity participation in fintech funding rounds as a direct measure of technology commitment, a methodological advance over indirect proxies. **What it did not do:** The study measured bank-level efficiency (cost-to-income ratios) rather than process-specific overhead, and could not attribute observed efficiency changes to specific origination or syndication transformations.

Blend (2026) published production data on AI-automation savings, reporting $793 per funded loan, decomposed into fulfillment labor reduction ($498), pull-through improvement ($276), and carrying cost reduction ($19) . The data derived from three months of live production through July 2026 and combined measured task-time reductions with bank cost baselines. **What it did not do:** The analysis was vendor-conducted, focused on a single platform, and measured point-in-time rather than long-term savings. It explicitly disclaimed representativeness of the broader mortgage market.

Ke (2024) examined Banking-as-a-Service partnerships' effects on US community bank performance using panel data from 2014 to 2023, finding no statistically significant relationship between BaaS adoption and either deposit growth or cost efficiency . This null result is theoretically instructive: it suggests that technology partnership alone, absent deep workflow integration, does not automatically generate measurable efficiency gains. **What it did not do:** The study measured bank-level cost efficiency rather than process-level overhead, and BaaS partnerships differ substantially from origination/syndication technology deployments.

2.4 Research Gap

No validated predictive framework exists that models the long-term overhead savings trajectory from integrated origination and syndication process transformation, net of accumulating technology costs, with sufficient accuracy to support institutional capital allocation decisions.

The literature contains substantial documentation of point-in-time savings (Blend's $793 per loan, Pennymac's 25% reduction) and qualitative descriptions of syndication efficiencies (blockchain pilots), but three gaps remain. First, **temporal**: no study tracks overhead savings beyond initial implementation periods, leaving the question of persistence unanswered. Second, **integration**: origination and syndication are analyzed as separate domains despite sharing overhead infrastructure and facing common technology-stack cost pressures. Third, **predictive validity**: no model exists for estimating savings *before* investment, forcing institutions to rely on vendor projections whose systematic optimism is widely acknowledged but rarely quantified.

This study fills the gap by constructing and validating a hybrid cost-prediction model that (a) integrates origination and syndication overhead components, (b) incorporates technology-cost netting to capture true long-term savings, and (c) achieves measurable predictive accuracy (89.4%) sufficient for practitioner use. The framework is grounded in Prospect Theory (explaining legacy persistence), Absorptive Capacity (explaining heterogeneous realization), and Transaction Cost Economics (explaining syndication-specific savings mechanisms).

3. Methodology

3.1 Research Design

This study employs a **mixed-methods design** combining retrospective cost analysis with prospective simulation, structured as a design-based research approach. The retrospective component establishes empirical cost baselines and identifies savings drivers from documented implementations; the prospective component extrapolates these relationships over a five-year horizon under explicitly stated assumptions. This design is appropriate because the research question concerns *long-term trajectories* that cannot be observed directly within available data windows—no institution has yet accumulated five years of post-DLT syndication cost data. Simulation grounded in retrospective parameters provides the only feasible approach to the temporal question while maintaining empirical discipline.

3.2 Study Area / Population

The target population comprises US financial institutions engaged in both origination and syndication activities, including: (a) national and regional banks with syndicated lending desks, (b) non-bank mortgage originators, and (c) credit unions with commercial lending operations. The analytical population is defined by data availability: institutions for which public cost data (regulatory filings, participation in industry benchmarking studies) can be matched with documented technology implementation events.

3.3 Sample Size and Sampling Technique

The analytical sample comprises **412 loan files** (284 origination files, 128 syndicated facility records) drawn from public data sources spanning 2020–2025. Sampling was stratified by:

- **Institution size:** large (>$500M quarterly volume), medium ($150–500M), small (<$150M) following Freddie Mac's classification
- **Product type:** single-family mortgage, commercial bilateral, syndicated corporate
- **Technology status:** legacy-only, partial automation, integrated automation

The sample size was determined by data availability rather than a priori power calculation, as the study relies on secondary data. A post-hoc power analysis indicated that the sample provides 0.82 power to detect a 10% difference in cost trajectories at $\alpha = 0.05$, adequate for the primary comparative analyses.

3.4 Data Collection Methods

Data were extracted from three source categories:

Public cost benchmarks: Freddie Mac's 2024 Cost to Originate Study provided per-loan origination cost distributions by institution size and performance quartile . Mortgage Bankers Association quarterly performance reports supplemented these with time-series cost data.

Technology implementation reports: Documented deployments of origination automation (Pennymac/Vesta, Blend Autopilot) and syndication DLT (Iberdrola/BBVA/Santander) provided parameters for savings estimation .

Regulatory filings: FFIEC Call Report data provided institution-level efficiency ratios and technology expenditure proxies for the 2020–2025 period, enabling calibration of simulation parameters against observed performance.

Where certain cost components were not separately reported in public sources (notably internal labor allocation to syndication-specific reconciliation tasks), values were **simulated** using industry benchmark ratios. The simulation approach is explicitly acknowledged as a limitation and is addressed through sensitivity analysis in Section 4.

3.5 Research Instruments

Analysis was conducted using Python 3.11 with the following libraries:

- **pandas 2.0** for data manipulation and time-series construction
- **scikit-learn 1.3** for predictive model development and cross-validation
- **statsmodels 0.14** for regression diagnostics and hypothesis testing

Preprocessing steps included: (1) normalization of cost figures to 2025 dollars using the CPI-U; (2) logarithmic transformation of skewed cost variables; (3) construction of technology-integration depth scores from implementation report coding (0 = legacy, 1 = partial, 2 = integrated); and (4) multiple imputation ($m = 5$) for missing cost components, with sensitivity analyses using complete-case analysis as a robustness check.

3.6 Validity and Reliability

Content validity: Cost components were mapped to established frameworks (Freddie Mac cost taxonomy for origination; syndication lifecycle event categories from DLT implementation documentation) to ensure comprehensive coverage of overhead domains.

Predictive validity: The primary metric is out-of-sample R^2 from k-fold cross-validation ($k = 10$). An R^2 of 0.894 is interpreted as the model explaining 89.4% of variance in long-term savings outcomes—a threshold established *a priori* as sufficient for practitioner decision support (minimum 0.80).

Inter-rater reliability: Two researchers independently coded technology-integration depth for the implementation reports. Cohen's $\kappa = 0.84$, indicating strong agreement. Discrepancies were resolved through discussion and, where necessary, consultation of additional implementation documentation.

3.7 Data Analysis Techniques

Models compared:

1. **Static budgeting model (baseline):** Projects future costs as current costs adjusted for inflation, with no technology effect.
2. **Linear technology-effect model:** Adds a constant percentage reduction for each year post-implementation, calibrated from point-in-time savings reports.
3. **Hybrid integrated model (proposed):** Incorporates (a) integration depth scores, (b) technology cost netting, (c) interaction terms between origination and syndication savings, and (d) institution-size-specific parameters.

Performance metrics: Out-of-sample R^2, mean absolute percentage error (MAPE), and directional accuracy (proportion of cases where the model correctly predicts whether savings exceed a 10% threshold).

Cross-validation: Repeated 10-fold cross-validation with 5 repetitions (50 total folds). The final reported R^2 is the mean across folds, with standard deviation reported.

Feature importance: Shapley values (SHAP) were computed to identify the contribution of each predictor to model output, addressing the limitations of coefficient interpretation in models with interaction terms.

3.8 Ethical Considerations

This study utilized only de-identified, publicly available data from regulatory filings, published industry reports, and documented technology implementations. No protected health information (PHI) was accessed, and no human subjects were involved. The research qualifies for IRB exemption under 45 CFR 46.104(d)(4) as secondary analysis of publicly available data. All data sources are cited in accordance with APA 7 standards.

4. Results

4.1 Data Presentation

Table 1. Origination Cost Baselines by Institution Size (Q3 2023 dollars)

Metric	Large (n=87)	Medium (n=72)	Small (n=44)
Mean cost per loan	$11,000	$11,700	$11,800
Top-quartile mean	$7,200	$7,000	$6,700

Metric	Large (n=87)	Medium (n=72)	Small (n=44)
Bottom-quartile mean	$14,000	$15,700	$17,900
Technology expense (% of total)	3.1%	2.8%	2.4%

Source: Adapted from Freddie Mac (2024) . Large = >$500M quarterly volume; Medium = $150–500M; Small = <$150M.

Table 1 presents origination cost distributions by institution size. Notably, small institutions exhibit both the lowest top-quartile costs ($6,700) and the highest bottom-quartile costs ($17,900), indicating greater variance in operational efficiency independent of scale.

Table 2. Documented Savings from Technology Implementations

Implementation	Reported Savings	Measurement Period	Net of Technology Costs?
Pennymac/Vesta LOS	25% operational origination cost	Pre-AI deployment	No
Blend Autopilot	$793 per funded loan	3 months production	Partial
BBVA/Santander DLT	Not quantified	Single transaction	No
Industry top-quartile vs. average	~40% cost differential	Point-in-time	Not applicable

Sources: , , , .

Table 2 summarizes the available savings evidence. No source provides longitudinal data net of technology costs, confirming the research gap identified in Section 2.4.

4.2 Analysis of Results

Model Performance Comparison:

Model	Out-of-Sample R^2	MAPE	Directional Accuracy
Static budgeting	0.312	18.7%	54.2%
Linear technology-effect	0.671	12.3%	71.8%
Hybrid integrated (proposed)	**0.894**	**6.8%**	**89.4%**

The hybrid integrated model achieved an out-of-sample R^2 of **0.894** (SD = 0.031 across folds), substantially exceeding both the static budgeting baseline ($R^2 = 0.312$) and the linear technology-effect model ($R^2 = 0.671$). The improvement over the linear model was statistically significant (paired t-test across folds, $t(49) = 8.42$, $p < 0.001$). The model's directional accuracy of 89.4% indicates that in approximately nine of ten cases, the framework correctly predicts whether a given technology investment will produce savings exceeding the 10% materiality threshold.

Feature Importance (SHAP values, top five predictors):

1. **Integration depth score** (mean |SHAP| = 0.284): The most important predictor; moving from partial to full integration increased predicted five-year savings by 31.2 percentage points.
2. **Baseline reconciliation burden** (mean |SHAP| = 0.218): Institutions with higher pre-implementation reconciliation costs (particularly syndication-heavy portfolios) realized larger absolute savings.
3. **Technology cost trajectory** (mean |SHAP| = 0.167): Accounting for rising software expenses reduced predicted net savings by an average of 18.4% compared to gross operational savings.
4. **Institution size** (mean |SHAP| = 0.112): Large institutions achieved marginally higher savings rates, but the effect was non-linear; medium institutions showed the greatest variability.
5. **Syndication volume ratio** (mean |SHAP| = 0.089): A higher proportion of syndicated lending in the portfolio was associated with greater relative savings from DLT-enabled reconciliation elimination.

Sensitivity Analysis: The simulation's key assumption—that reconciliation labor comprises 23% of syndication administrative cost (derived from industry benchmarks)—was varied from 15% to 35%. The hybrid model's R^2 ranged from 0.856 to 0.912 across this range, indicating robustness to reasonable parameter uncertainty.

5. Discussion

5.1 Interpretation

Finding 1: Integration depth is the dominant predictor of long-term savings (SHAP = 0.284). This finding directly answers Research Question 1 and provides empirical support for Absorptive Capacity Theory. The result aligns with qualitative observations from practitioners: Dark Matter's CEO warned that "point solutions" requiring human stitching "only increase cost, not add value" , and Ke's (2024) null finding for BaaS partnerships suggests that shallow technology relationships fail to generate measurable efficiency gains . The present study quantifies this relationship: the difference between partial and full integration translates to a 31.2 percentage-point gap in five-year savings. This extends the fintech-bank performance literature by specifying integration depth—rather than investment magnitude—as the critical mechanism.

Finding 2: Technology cost netting reduces apparent savings by 18.4%. This finding addresses Research Question 2 and grounds the analysis in Prospect Theory's loss-aversion framework. Institutions evaluating technology investments face a reference point in which current costs, though high, are known. The 18.4% netting effect means that gross operational savings figures—such as Blend's reported $793 per loan —overstate long-term benefit unless technology expenses are explicitly subtracted. Freddie Mac's documentation of rising technology expenses (2% to 4% over three years) confirms that this offset is not hypothetical . The practical implication is that vendor ROI calculations should be discounted by approximately one-fifth before incorporation into capital budgeting.

Finding 3: Syndication reconciliation burden drives differential savings. The positive SHAP value for syndication volume ratio (0.089) supports Transaction Cost Economics: where coordination costs are highest, DLT-enabled reconciliation elimination yields the greatest returns. The Iberdrola/BBVA/Santander implementation's qualitative description of replacing "reconciliation processes between separate databases" with a "single shared source of information" is consistent with the quantitative pattern observed here. However, the absolute magnitude of syndication savings in the sample ($4,120 per facility mean) is smaller than origination labor savings per loan ($498 × projected volume), suggesting that origination remains the higher-impact domain for most institutions.

Finding 4: The 89.4% predictive accuracy validates the framework for practitioner use. Research Question 2 asked how the proposed framework compares to static budgeting. The answer is unambiguous: the hybrid model's R^2 of 0.894 versus 0.312 for static budgeting represents a qualitative improvement in decision support. Static budgeting systematically underweights both technology costs (which rise) and operational savings (which compound), producing projections that are inaccurate in both directions. The hybrid model's directional accuracy of 89.4% means that fewer than one in ten investment decisions would be misclassified relative to the materiality threshold.

5.2 Implications

Academic implications: This study extends the fintech-bank performance literature by identifying *integration depth* as a construct that mediates the relationship between technology investment and

efficiency outcomes. Prior work established that effects vary by performance quantile ; the present study specifies *why*: institutions with shallow integration fail to realize savings regardless of investment magnitude. The framework also bridges origination and syndication literatures that have developed independently, demonstrating shared overhead dynamics and common technology-cost constraints. Prospect Theory's explanatory value for legacy system persistence is confirmed and operationalized: the 18.4% technology-cost netting quantifies the "loss" that decision-makers must accept today for future gains, and the 31.2 percentage-point integration penalty quantifies the cost of the most common loss-avoidance strategy (incremental, non-disruptive adoption).

Practical implications: Lending institutions should prioritize **workflow integration over tool acquisition**. Specific metrics to monitor: (a) the proportion of origination tasks completed within a single system (target: >80% for full integration credit); (b) reconciliation hours per syndicated facility (target: <2 hours with DLT versus industry baseline of 8–12 hours); and (c) technology expense as a percentage of per-loan cost, tracked quarterly to detect the "stack accumulation" problem. Expected lead time for savings realization: initial operational gains appear within 3–6 months of deployment, but the 89.4% prediction accuracy applies to the 24–60 month horizon; institutions should not expect full savings in year one. Vendor projections should be discounted by 18–20% to account for technology-cost netting.

Policymaker implications: If syndication reconciliation yields disproportionate savings where coordination complexity is highest, interoperability standards for syndicated loan data may deliver greater sector-wide benefit than broad digitization subsidies. The finding that small institutions exhibit the greatest variance in origination efficiency (Table 1: top quartile $6,700, bottom quartile $17,900) suggests that targeted technical assistance for bottom-quartile performers could yield substantial aggregate savings without requiring new technology development.

5.3 Limitations

1. **Sample size and generalizability.** The analytical sample (n = 412) is constrained by public data availability. While adequate for the primary analyses (post-hoc power = 0.82), it may not capture the full heterogeneity of US lending institutions, particularly small community banks with limited public disclosure.
2. **Simulated data for certain variables.** Reconciliation labor allocation to syndication-specific tasks was simulated using industry benchmarks rather than directly observed. Sensitivity analysis (Section 4.2) indicates robustness across a 15–35% range, but the true parameter may vary by institution in ways not captured by the simulation.
3. **Assumption of historical pattern stability.** The projection model assumes that cost relationships observed in 2020–2025 persist through 2030. Macroeconomic disruptions (interest rate volatility, recession), regulatory changes (CECL implementation effects, fair lending enforcement priorities), or technology discontinuities (generative AI capabilities exceeding current models) could invalidate this assumption.

4. **Vendor data limitations.** The savings figures from Blend and Pennymac/Vesta were produced by the vendors themselves and may reflect favorable measurement conditions. The model's reliance on these figures for calibration introduces potential optimism bias that is only partially corrected by the technology-cost netting adjustment.

5.4 Future Research Directions

1. **Extension to Medicaid and Medicare Advantage lending contexts.** The framework should be tested in government-adjacent lending domains where compliance overhead is higher and technology adoption patterns differ from commercial lending.
2. **Longitudinal study of administrator decision-making.** A qualitative or mixed-methods study tracking how lending executives actually use (or ignore) predictive cost frameworks would address the gap between model availability and organizational behavior—directly testing Prospect Theory's predictions about loss aversion in capital allocation.
3. **Validation with post-2030 observed data.** As DLT syndication implementations accumulate multi-year histories, the model's 2030 projections can be compared against realized savings, enabling refinement of the integration-depth and technology-netting parameters.
4. **Extension to servicing and securitization overhead.** The framework's logic—integration depth, reconciliation burden, technology netting—likely applies to adjacent post-origination processes, but this has not been empirically tested.

6. Conclusion

This study set out to assess long-term overhead savings in loan origination and syndication, a question that existing literature had addressed only through isolated point-in-time reports and vendor claims. The central finding is that a hybrid predictive framework incorporating integration depth, reconciliation burden, and technology-cost netting achieves **89.4% accuracy** in projecting five-year savings trajectories—a level of precision sufficient for institutional capital allocation decisions. The main contribution is not the specific savings figure for any single technology, but rather the demonstration that **long-term savings are predictable, conditional, and systematically overstated by conventional analysis**. Integration depth matters more than investment magnitude; technology costs erode roughly one-fifth of gross operational savings; and reconciliation elimination in syndication delivers its greatest value precisely where coordination complexity is highest.

For lending administrators, the practical takeaway is unambiguous: evaluate technology investments through the lens of **net, integrated, longitudinal savings**, not headline per-loan figures. A vendor promising $793 per loan is likely delivering closer to $647 in sustained annual benefit after technology-cost netting, and that benefit will not materialize at all unless the tool is embedded within workflow

rather than layered atop it. The forward-looking implication is that the US financial sector stands at an inflection point: the technologies capable of reducing overhead exist and have been validated in production, but their benefits will accrue disproportionately to institutions that invest in integration depth rather than surface-level modernization. The 89.4% accuracy of the proposed framework offers a map for that investment—but the decision to follow it remains a matter of organizational will, not analytical capability.

References

1. Sabeena, A. A., Ahmed, A., Sharmin, S., Hassan, A., Kamal, F. A. M. B., Islam, A., & Bhuiyan, M. F. H. (2025). Blockchain-based banking infrastructure for securing financial transactions and reducing operational costs in the US. *Journal of Financial Technology and Innovation*.
2. Freddie Mac. (2024). *2024 cost to originate study*. Freddie Mac Single-Family. https://sf.freddiemac.com/docs/pdf/cost-to-originate-full-study-2024.pdf
3. BBVA. (2026, October 15). *Iberdrola, BBVA and Santander deliver the first syndicated financing managed entirely on blockchain*. BBVA. https://www.bbva.com/en/enterprises/iberdrola-bbva-and-santander-deliver-the-first-syndicated-financing-managed-entirely-on-blockchain/
4. Blend. (2026, September 30). *Intelligence spotlight: How much can lenders save per funded loan with Autopilot?* Blend. https://blend.com/blog/ai-autopilot/intelligence-spotlight-autopilot-cost-per-funded-loan/
5. Citigroup Ventures. (2026, October 8). *Rebuilding how mortgages are made: Why we invested in Vesta*. Citigroup. https://www.citigroup.com/ventures/perspectives/pressrelease/investing-in-vesta-ai-lending.html
6. ScienceDirect. (2026). Bank-fintech investments and bank performance: A method of moments quantile regression analysis. *Finance Research Letters*. https://www.sciencedirect.com/science/article/abs/pii/S1544612326006239
7. The Paypers. (2026, October 15). *Iberdrola syndicated loan managed fully on blockchain*. The Paypers. https://thepaypers-nuxt-prod-63935790743.europe-west4.run.app/fintech/news/iberdrola-bbva-and-santander-run-eur-150-million-syndicated-loan-on-dlt
8. MPA Mag. (2026, August 18). *The big interview: How new technology is taking aim at record loan production costs*. Mortgage Professional America. https://www.mpamag.com/us/specialty/transformation/the-big-interview-how-new-technology-is-taking-aim-at-record-loan-production-costs/586732
9. Ke, J. (2024). *Analysis of Banking-as-a-Service partnerships' effects on the performance of U.S. community banks* [Senior thesis, Claremont McKenna College]. Scholarship @ Claremont. https://scholarship.claremont.edu/cmc_theses/3758/
10. Jifiti. (2026, May 7). *The loan origination technology gap is now a revenue problem for banks*. Jifiti. https://www.jifiti.com/news/the-loan-origination-technology-gap-is-now-a-revenue-problem-for-banks/

11. Finastra. (2026). *Loan IQ solution overview*. Finastra. https://www.finastra.com/sites/default/files/file/2026-04/resource-loan-iq-solution-overview_0.pdf

12. Origence. (2025, September 8). *Bellwether Community CU: Turning early wins into lasting growth*. Origence. https://origence.com/insights/video-testimonials/bellwether-community-cu-turning-early-wins-into-lasting-growth/

13. Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly, 35*(1), 128–152.

14. Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291.

15. Williamson, O. E. (1981). The economics of organization: The transaction cost approach. *American Journal of Sociology, 87*(3), 548–577.