



## Article

# Big Data Adoption and Bank Performance in Uzbekistan: Panel Fixed Effects Evidence from Commercial Banks

Shakhnoza Turabova<sup>1</sup>

1. Department of Digital Economics and International Digital Integration, Tashkent State University of Economics, Uzbekistan

\* Correspondence: [sh.turabova@tsue.uz](mailto:sh.turabova@tsue.uz)

**Abstract:** Commercial banks in emerging economies invest in Big Data infrastructure without direct evidence that such investment improves financial performance, largely because banks do not disclose technology spending separately. This study measures Big Data adoption through a proxy based on the ratio of information technology depreciation to total operating expenses and estimates its effect on profitability and credit quality in the Uzbek banking sector. A panel of six commercial banks observed over seven years, from 2018 to 2024, is analysed with fixed effects estimators, with model choice confirmed by Hausman specification tests and inference based on cluster-robust standard errors. The estimated coefficients of the adoption proxy are positive for return on assets and return on equity and positive for the non-performing loan ratio, but none reaches conventional levels of statistical significance, while the 2023 ownership change dummy is strongly significant across all three specifications. The results indicate that over a seven-year window the measurable effect of Big Data adoption on bank performance is not yet distinguishable from zero, a pattern consistent with the delayed returns documented in the information technology investment literature. The study contributes a replicable measurement procedure for settings where technology expenditure is not separately reported.

**Keywords:** bank performance; Big Data adoption; fixed effects; panel data; Uzbekistan

**Citation:** Turabova S. A Big Data Adoption and Bank Performance in Uzbekistan: Panel Fixed Effects Evidence from Commercial Banks. Companies: Problems and Solutions. American Journal of Economics and Business Management 2026, 9(9), 41-48.

Received: 10<sup>th</sup> Jun 2026

Revised: 11<sup>th</sup> Jul 2026

Accepted: 11<sup>th</sup> Aug 2026

Published: 04<sup>th</sup> Sep 2026



**Copyright:** © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

The banking sector produces one of the densest streams of transactional data in any economy. Card payments, mobile sessions, remote transfers and credit applications generate records continuously, and the volume of those records has grown rapidly in economies where digital financial services expanded from a low base. Banks in these markets have responded by investing in data warehouses, analytical platforms and machine learning tools. The expectation behind such investment is that better use of data will lower operating costs, sharpen credit assessment and ultimately improve financial results.[1]

Whether that expectation is met is an empirical question, and one that has proved difficult to answer. The obstacle is measurement. Banks do not report Big Data expenditure as a separate line in their financial statements. Technology spending is distributed across capital expenditure, amortisation of intangible assets, staff costs and outsourced services, and no accounting standard requires it to be disclosed in a way that would let an outside researcher isolate it. Survey instruments have been used to fill this gap, but surveys measure perceived capability rather than realised spending, they depend on the

willingness of banks to respond, and they cannot be reconstructed from public records by a second researcher.[2]

This paper addresses the measurement problem directly and then applies the resulting indicator in a panel setting. The approach is to build a proxy from audited financial statements. Depreciation of information technology assets, scaled by total operating expenses, captures the share of a bank cost base that is tied to technology infrastructure. The measure is imperfect, because depreciation reflects past capital outlays rather than current commitments, and because banks differ in depreciation policy. It has the compensating advantage of being reconstructible by anyone with access to the same published statements, which makes the results verifiable and the procedure transferable to other markets where technology spending is similarly opaque.[3]

The empirical setting is the Uzbek banking system. The sector has undergone a rapid and well documented digital expansion, which means that variation in technology intensity across banks and over time is large enough to be detected. At the same time the sector remains concentrated: according to the financial system stability assessment prepared jointly by the International Monetary Fund and the World Bank, the system comprises thirty-six banks, of which nine are state controlled and together hold about sixty-seven percent of sector assets (International Monetary Fund, ). Concentration limits the number of institutions available for analysis but also means that a small set of banks accounts for most of the system.[4][5]

The period covered by the study coincides with the most intensive phase of banking sector reform in the country. Over these years the number of payment cards in circulation, the number of remote banking users and the volume of terminal payments all expanded several times over, and banks moved substantial parts of their customer interaction onto digital channels. The period therefore offers a setting in which an effect, if it exists, would have had reason to appear.[6]

The research question is whether higher Big Data adoption, measured by the proxy, is associated with better bank performance. Performance is captured by three standard indicators: return on assets and return on equity for profitability, and the non-performing loan ratio for credit quality. The first two are expected to rise with adoption if analytical capability improves operational efficiency and pricing. The third is expected to fall if better data improves borrower screening and early warning.[7]

The study makes three contributions. First, it sets out a measurement procedure for Big Data adoption that relies only on published financial statements. Second, it estimates the association between adoption and performance using a fixed effects panel specification, with the choice between fixed and random effects settled by formal specification testing rather than assumed. Third, it reports the results honestly, including the finding that the estimated effects are not statistically significant, and interprets that outcome in the light of the established literature on delayed returns to technology investment in banking.[8]

The relationship between technology investment and bank performance has been studied for more than two decades, and the accumulated evidence is more nuanced than the intuition that technology raises profits. Berger surveyed the economic effects of technological progress in the banking industry and documented that productivity gains from technology appear with substantial delay, because implementation absorbs resources before it releases them, and because organisational adjustment lags the installation of systems. Beccalli examined European banks directly and found that information technology expenditure affects profitability with a lag, and that the short run association can be negative even where the long run association is positive. The pattern is often described as a J-curve: costs are incurred immediately, benefits accrue later.[9]

This literature carries a methodological implication that is central to the present study. If the effect of technology investment on performance materialises over a horizon longer

than the observation window, then a short panel will not detect it, and a statistically insignificant coefficient is not evidence that the effect is absent. It is evidence that the design lacks the power or the time span to identify it.[10]

More recent work has extended the question to digital transformation understood more broadly. He, Jin, Hong, Ma, Jiang and Chen analysed listed Chinese commercial banks and reported that digital transformation improves operational performance, with the magnitude of the effect conditional on bank size and ownership structure. Their finding that larger institutions extract more value from the same transformation is consistent with the presence of fixed costs in building analytical capability.[11]

A parallel strand concerns what Big Data capability is and how it should be measured. Chen, Chiang and Storey positioned business intelligence and analytics as a driver of measurable impact in finance. Gandomi and Haider (2015) clarified the volume, velocity and variety dimensions that determine infrastructure requirements. Gupta and George developed a survey based instrument for measuring big data analytics capability as a firm level resource. Mikalef, Pappas, Krogstie and Giannakos reviewed this measurement literature and observed that most instruments rely on self-reported perceptions, which limits comparability across studies and prevents replication from public sources.[12]

The survey based approach and the accounting based approach adopted here answer different questions. A survey captures what managers believe their organisation can do; a financial ratio captures what the organisation has actually committed to technology assets. The accounting measure is narrower, because it excludes capability embodied in staff and processes, but it is observable, comparable across institutions reporting under the same standards, and reproducible without the cooperation of the banks studied.

Applications of Big Data within banking have been mapped in bibliometric terms by Nobanee, Dilshad, Al Dhanhani, Al Neyadi, Al Qubaisi and Al Shamsi, who identified credit risk assessment, fraud detection and customer segmentation as the dominant themes of the field. Work in the credit scoring stream, such as Bahri, Herdiyeni and Suprehatin, demonstrates that machine learning classifiers built on alternative data can achieve discriminatory power above conventional scorecards. These application studies establish that analytical tools work at the level of the individual decision. They do not establish that adopting such tools moves institution-level financial indicators, which is the question addressed here. For the Uzbek banking system specifically, institution level panel evidence linking technology intensity to financial outcomes has not been established.

## 2. Materials and Methods

### Instrument

Big Data adoption is measured by an indicator constructed from audited financial statements. For bank  $i$  in year  $t$  the indicator is defined as the ratio of information technology asset depreciation to total operating expenses, expressed as a percentage. The logic follows the treatment of technology intensity in Berger (2003) and Beccalli, where the resource commitment to technology is inferred from cost structure rather than from disclosed project budgets.[13]

Constructing the indicator requires one intermediate step, because banks reporting under international financial reporting standards typically disclose property and equipment in aggregated categories. Where information technology equipment is reported within a combined furniture and equipment line, the technology component is separated using the proportion disclosed in the notes to the financial statements. In this study the separation coefficient was taken from the note on property and equipment in one bank statement that reports the breakdown explicitly, and applied consistently to institutions that do not disaggregate. This is the least satisfactory element of the procedure, and it is stated openly so that any replication can vary the assumption.[14]

Three performance indicators serve as dependent variables. Return on assets and return on equity measure profitability. The ratio of non-performing loans to the gross loan portfolio measures credit quality. Three control variables are included: the capital adequacy ratio, the net interest margin, and the natural logarithm of total assets, the last capturing scale effects of the kind reported by He et al. A dummy variable marks 2023, the year in which one of the banks in the sample underwent a change of ownership.

**Table 1.** Definition of variables and data sources

| Symbol | Variable                                 | Unit       | Source                               |
|--------|------------------------------------------|------------|--------------------------------------|
| ROA    | Return on assets (dependent)             | %          | IFRS statements of banks             |
| ROE    | Return on equity (dependent)             | %          | IFRS statements of banks             |
| NPL    | Non-performing loan ratio (dependent)    | %          | IFRS statements; Central Bank        |
| BD     | Big Data adoption proxy (main regressor) | %          | IT depreciation / operating expenses |
| CAR    | Capital adequacy ratio (control)         | %          | IFRS statements of banks             |
| NIM    | Net interest margin (control)            | %          | IFRS statements of banks             |
| ln(A)  | Logarithm of total assets (control)      | ln(bn UZS) | Balance sheets of banks              |
| D2023  | Ownership change dummy (control)         | 0 / 1      | Public disclosures                   |

#### Data collection

The panel covers six commercial banks over the seven years from 2018 to 2024, giving forty-two potential bank-year observations. The banks included are Xalq Bank, the National Bank of Uzbekistan, Sanoat Qurilish Bank, Kapitalbank, Agrobank and Ipoteka Bank. Selection was determined by data availability: these are the institutions that published audited statements under international reporting standards continuously across the whole period, with notes detailed enough to permit construction of the adoption proxy.

All financial variables were collected from the audited annual reports of the banks concerned. Sector level context and the classification of non-performing exposures were cross-checked against the statistical bulletins of the Central Bank of the Republic of Uzbekistan (2026) and against the joint financial system stability assessment (International Monetary Fund, ). Where a bank restated a prior year figure in a later report, the restated value was used.

The effective number of observations is smaller than forty-two because some banks did not disclose every variable in every year. The non-performing loan ratio and the capital adequacy ratio are the most frequently missing items, with thirty-six observations each, while the adoption proxy is available for forty observations. Estimation samples therefore range from thirty-two to thirty-three observations. Missing values were left as missing rather than imputed, since imputation across a panel this small would impose structure that the data cannot support.[15]

#### Data analysis

The estimation framework is a static linear panel model in which a performance indicator for bank  $i$  in year  $t$  is explained by the adoption proxy, the vector of controls, a bank-specific intercept and a disturbance term. The bank-specific intercept absorbs time-invariant characteristics such as ownership form, franchise position and management culture, which are plausibly correlated with both technology intensity and performance.

The choice between fixed and random effects was made using the Hausman specification test, which compares the two estimators under the null hypothesis that the bank-specific effects are uncorrelated with the regressors. The test was conducted separately for each dependent variable, and the results are reported below rather than assumed.

Inference is based on standard errors clustered at the bank level, following the practitioner guidance of Cameron and Miller. The number of clusters is six, which is small, and cluster-robust inference with few clusters is known to under-reject or over-reject

depending on the design. This limitation is acknowledged directly: the standard errors reported here should be read as indicative rather than exact. The panel structure and estimation follow the standard treatment in Baltagi. Computations were carried out in Python using the linearmodels package.

### 3. Result and Discussion

#### Descriptive statistics

Table 2 reports the distribution of each variable across the pooled sample. Profitability is highly dispersed and strongly left-skewed: return on assets ranges from a loss of 7.97 percent to a gain of 4.77 percent, and return on equity from a loss of 68.71 percent to a gain of 50.77 percent, with skewness coefficients of minus 2.10 and minus 1.49 respectively. These extremes are driven by a small number of loss-making bank-years and they warn against relying on mean comparisons alone.

**Table 2.** Descriptive statistics of the panel variables

| Variable | N  | Mean  | SD    | Min    | Median | Max   | Skew  |
|----------|----|-------|-------|--------|--------|-------|-------|
| ROA, %   | 42 | 1.10  | 2.17  | -7.97  | 1.44   | 4.77  | -2.10 |
| ROE, %   | 42 | 9.78  | 19.19 | -68.71 | 11.94  | 50.77 | -1.49 |
| NPL, %   | 36 | 6.94  | 6.01  | 1.00   | 5.47   | 26.82 | 2.05  |
| CAR, %   | 36 | 17.49 | 4.06  | 10.60  | 17.00  | 28.00 | 0.60  |
| NIM, %   | 40 | 5.48  | 1.85  | 2.41   | 5.08   | 10.10 | 0.80  |
| BD, %    | 40 | 5.16  | 1.83  | 2.31   | 5.22   | 11.21 | 1.01  |

The adoption proxy has a mean of 5.16 percent and a standard deviation of 1.83, with values ranging from 2.31 to 11.21 percent. This dispersion is the variation the model exploits: the most technology-intensive bank-year in the sample devotes almost five times the share of its cost base to information technology depreciation as the least intensive.

#### Specification tests

Table 3 reports the Hausman test statistics. For the non-performing loan ratio and for return on equity the null hypothesis is rejected at the one percent level, indicating that random effects would be inconsistent. For return on assets the test statistic of 9.7 yields a p-value of 0.14, which does not reject the null. Fixed effects were nonetheless adopted for all three specifications, so that the estimator is consistent under both hypotheses and the three models remain directly comparable.

**Table 3.** Hausman specification test results

| Model | Chi-square | p-value | Pooled OLS coefficient | Decision            |
|-------|------------|---------|------------------------|---------------------|
| ROA   | 9.7        | 0.14    | +0.120                 | RE or FE admissible |
| NPL   | 26.9       | < 0.01  | +0.111                 | FE required         |
| ROE   | 20.6       | < 0.01  | +0.434                 | FE required         |

*Note.* The null hypothesis is that bank-specific effects are uncorrelated with the regressors.

The comparison between the pooled and the within estimates is itself informative. The pooled coefficients on the adoption proxy are uniformly smaller than the fixed effects coefficients reported below. Pooled estimation attributes to technology intensity part of what is in fact a stable difference between institutions, and once those differences are absorbed the within-bank association is larger.

#### Regression results

Table 4 reports the fixed effects estimates. The coefficient on the adoption proxy is positive in all three specifications. For return on assets it is 0.1796 with a standard error of 0.1855, for the non-performing loan ratio 0.2485 with a standard error of 0.1936, and for

return on equity 0.9647 with a standard error of 1.5853. None is statistically significant at conventional levels: the p-values are 0.347, 0.218 and 0.551 respectively.

**Table 4.** Fixed effects estimates of the effect of Big Data adoption on bank performance

| Statistic            | Model 1: ROA | Model 2: NPL | Model 3: ROE |
|----------------------|--------------|--------------|--------------|
| Coefficient on BD    | +0.1796      | +0.2485      | +0.9647      |
| Standard error       | (0.1855)     | (0.1936)     | (1.5853)     |
| t-statistic          | 0.97         | 1.28         | 0.61         |
| p-value              | 0.347        | 0.218        | 0.551        |
| Coefficient on D2023 | -7.309***    | +4.857**     | -61.921***   |
| Observations         | 33           | 32           | 33           |
| R-squared (within)   | 0.582        | 0.531        | 0.588        |
| F-statistic          | 2.371*       | 1.809        | 2.423*       |

*Note.* Standard errors clustered at the bank level in parentheses. Asterisks denote significance at the ten (\*), five (\*\*) and one (\*\*\*) percent levels. All specifications include bank fixed effects and the controls listed in Table 1.

The within R-squared values, between 0.531 and 0.588, indicate that the models account for a little more than half of the variation in performance that occurs within banks over time. The F-statistics for the profitability equations are significant at the ten percent level, so the specifications as a whole carry explanatory content even though the coefficient of primary interest does not attain significance individually.

The dummy for 2023 is the most strongly determined coefficient in every specification. It reduces return on assets by 7.31 percentage points and return on equity by 61.92 percentage points, and raises the non-performing loan ratio by 4.86 percentage points, all at conventional significance levels. This is the statistical signature of the ownership change that affected one bank in the sample. It also demonstrates that the specification is capable of detecting a genuine effect when one is present, which strengthens rather than weakens the interpretation of the insignificant adoption coefficients.

#### Robustness

Re-estimating the three equations as pooled ordinary least squares, without bank intercepts, yields coefficients on the adoption proxy of 0.120 for return on assets, 0.111 for the non-performing loan ratio and 0.434 for return on equity. All three remain positive and all three remain insignificant, so the sign pattern is not an artefact of the within transformation. Table 5 collects the coefficient, standard error, t-statistic, within R-squared and sample size for each equation, which makes the relationship between effect size and precision explicit. In each case the t-statistic lies between 0.6 and 1.3 and the standard error is of the same order of magnitude as the point estimate.

**Table 5.** Summary of specifications and estimated effects

| Specification | Coefficient<br>t | Std. error | t-statistic | p-value | R-sq.<br>(within) | N  |
|---------------|------------------|------------|-------------|---------|-------------------|----|
| (1) ROA       | +0.180           | (0.186)    | 0.97        | 0.347   | 0.58              | 33 |
| (2) NPL       | +0.249           | (0.194)    | 1.28        | 0.218   | 0.53              | 32 |
| (3) ROE       | +0.965           | (1.585)    | 0.61        | 0.551   | 0.59              | 33 |

*Note.* All specifications include bank fixed effects, the controls listed in Table 1 and the 2023 structural dummy.

Because the sample contains a small number of loss-making bank-years with extreme values of return on equity, the profitability equations were re-examined with those years excluded. The coefficient on the adoption proxy remained positive and insignificant, and

the within R-squared changed only marginally, indicating that the central result is not produced by outliers.

#### Discussion

Three points follow from these results. The first concerns the direction of the estimated associations. The positive coefficients on the profitability measures are consistent with the hypothesis that Big Data adoption supports financial performance, but the coefficient on the non-performing loan ratio is also positive, which runs contrary to the expectation that better analytics improve credit screening. Because none of the three coefficients is significant, no weight should be placed on the signs. Reporting them is nevertheless necessary, since selectively presenting only the coefficients that match prior expectations would misrepresent the evidence.

The second point concerns the interpretation of insignificance. The literature establishes that returns to technology investment in banking accrue with a lag, and that the short run association may even be adverse while implementation costs are being absorbed (Berger, 2003; Beccalli, 2007). A panel spanning seven years contains too few time periods to separate the initial cost phase from the later benefit phase. The appropriate conclusion is that this design cannot detect an effect, not that no effect exists.

The third point concerns statistical power. With six clusters and roughly thirty-two usable observations, the design has limited capacity to detect associations of moderate size. The standard error on the return on assets coefficient, 0.1855, is of the same order as the coefficient itself, which means that effects smaller than about a third of a percentage point per unit of the proxy would be indistinguishable from zero. Any strengthening of the evidence requires either more institutions, which the structure of the sector limits, or more periods, which will accumulate with time.

Set against these limitations, the measurement contribution stands independently of the regression results. The adoption proxy can be constructed for any bank that publishes statements under international reporting standards, it requires no cooperation from the institution, and it can be recomputed by another researcher from the same sources. Extending the time dimension would allow dynamic specifications in which lagged adoption enters alongside current adoption, which is the natural way to test the delayed-returns hypothesis directly, and where longer series are available at higher frequency the bounds testing approach of Pesaran, Shin and Smith (2001) offers a route to separating short run from long run relationships.

#### 4. Conclusion

This study measured Big Data adoption in the Uzbek banking sector through a proxy constructed from audited financial statements, defined as the ratio of information technology depreciation to total operating expenses, and estimated its association with profitability and credit quality using a fixed effects panel of six commercial banks observed from 2018 to 2024, with the specification confirmed by Hausman tests and inference based on cluster-robust standard errors; the estimated coefficients were positive for return on assets and return on equity and positive for the non-performing loan ratio, but none reached conventional significance, while the structural dummy marking the 2023 ownership change was strongly significant throughout, indicating that the specification detects genuine effects when they are present and that the absence of a detectable adoption effect over a seven-year window is consistent with the delayed returns established in the technology investment literature rather than with the absence of any effect; the contribution of the study lies primarily in the measurement procedure, which is reproducible from public sources and transferable to other markets where banks do not disclose technology spending separately.

## ACKNOWLEDGEMENT

The author thanks the academic supervisor for guidance during the preparation of this study. No external funding was received for this research. The author declares no conflict of interest.

## REFERENCES

- [1] K. A. Bahri, Y. Herdiyeni, and Suprehatin, "Credit scoring model for farmers using random forest," *Jurnal RESTI*, vol. 7, no. 4, pp. 800–807, 2023.
- [2] B. H. Baltagi, *Econometric Analysis of Panel Data*, 3rd ed. Chichester, U.K.: John Wiley & Sons, 2005.
- [3] E. Beccalli, "Does IT investment improve bank performance? Evidence from Europe," *Journal of Banking & Finance*, vol. 31, no. 7, pp. 2205–2230, 2007.
- [4] A. N. Berger, "The economic effects of technological progress: Evidence from the banking industry," *Journal of Money, Credit and Banking*, vol. 35, no. 2, pp. 141–176, 2003.
- [5] A. C. Cameron and D. L. Miller, "A practitioner's guide to cluster-robust inference," *Journal of Human Resources*, vol. 50, no. 2, pp. 317–372, 2015.
- [6] Central Bank of the Republic of Uzbekistan, *Statistical Bulletin*. Tashkent, Uzbekistan: Central Bank of the Republic of Uzbekistan, 2026.
- [7] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165–1188, 2012.
- [8] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
- [9] M. Gupta and J. F. George, "Toward the development of a big data analytics capability," *Information & Management*, vol. 53, no. 8, pp. 1049–1064, 2016.
- [10] J. A. Hausman, "Specification tests in econometrics," *Econometrica*, vol. 46, no. 6, pp. 1251–1271, 1978.
- [11] S. He, Z. Jin, Y. Hong, L. Ma, Y. Jiang, and Y. Chen, "The impact of digital transformation on commercial banks' business performance," *SAGE Open*, vol. 15, no. 2, pp. 1–18, 2025.
- [12] International Monetary Fund, *Republic of Uzbekistan: Financial System Stability Assessment*, IMF Country Report No. 25/227. Washington, DC, USA: International Monetary Fund, 2025.
- [13] P. Mikalef, I. O. Pappas, J. Krogstie, and M. Giannakos, "Big data analytics capabilities: A systematic literature review and research agenda," *Information Systems and e-Business Management*, vol. 16, pp. 547–578, 2018.
- [14] H. Nobanee, M. N. Dilshad, M. Al Dhanhani, M. Al Neyadi, S. Al Qubaisi, and S. Al Shamsi, "Big data applications in the banking sector: A bibliometric analysis approach," *SAGE Open*, vol. 11, no. 4, pp. 1–17, 2021.
- [15] M. H. Pesaran, Y. Shin, and R. J. Smith, "Bounds testing approaches to the analysis of level relationships," *Journal of Applied Econometrics*, vol. 16, no. 3, pp. 289–326, 2001.