## ARTICLE

Check for updates

# From insights to impact: leveraging data analytics for data-driven decision-making and productivity in banking sector

Raazia Gul [1] & Mamdouh Abdulaziz Saleh Al-Faryan [2,3]

This research stems from the disturbing phenomenon known as digital transformation and the colossal data creation. A vast amount of data has been produced as a result of increased usage of digital technologies to generate commercial value. However, data does not hold any value, so organizations are motivated to invest in data analytics and make informed decisions to enhance their performance. Against this backdrop, this study investigates the impact of data-driven decision-making (DDDM) on productivity in the presence of data analytics capability of Pakistan's banking sector. We explore this link based on innovation diffusion theory using Instrumental Variable Two-Stage Least Square. A composite index of DDDM was developed based on primary data collected through online survey. This index is then regressed on actual productivity measures for 2016–2020. The findings suggest that banks exploiting analytics and adopting DDDM methods results in an increase in productivity of about 9–10%. It further indicates that DDDM adoption with an investment in data analytics leads to enhanced productivity.

[1] Faculty of Management Sciences, SZABIST University, Karachi, Pakistan. [2] School of Accounting, Economics and Finance, Faculty of Business and Law, University of Portsmouth, Portsmouth, UK. [3] Consultant in Economics and Finance, Riyadh, Saudi Arabia. ✉email: raazigul@gmail.com

## Introduction

Since the pandemic's onset, everyone has become ever more reliant on digital services. Hybrid meetings, virtual learning and digital payments for online purchasing need dependable, fast digital technologies. The efficient application of digital technologies to generate economic value has led to the generation of enormous amounts of data, and harnessing this data for decision-making is crucial. The ever-increasing volume of data and continuous advancement of analytics open up new opportunities and avenues for companies. There are multiple ways of exploiting these opportunities; the most prominent one is to make decisions based on data instead of intuition or expertise, known as data-driven decision-making (DDDM) (Brynjolfsson and McElharan 2019). The main goal of investment in innovative technologies is to embrace decision-making based on data (Erickson and Rothberg 2018; Welbourne 2015; Anderson 2015) and to develop a culture where senior managers and workers make decisions based on data rather than intuition.

The global business environment is changing as a result of big data and the insights generated from it for informed decision-making (Lochy 2017; Conejero et al. 2021). Because of this, industries are changing their business models and procedures to become more competitive, dynamic, and informed. With the use of data analytics, corporate managers may forecast risks, predict future trends, and comprehend the dynamics of their industry. However, if data is not efficiently disseminated to the managers to be incorporated into their decision-making, there is no use in piling data.

Today, many organizations have adopted digitalization and want to be data-driven; however, few could adopt DDDM (Anderson 2015) because either the required skill and expertise are lacking or the traditional procedures and routines are not redefined. This goes in line with Gul et al. (2023), who argue that after introducing innovative technologies to the organization, a more significant challenge would emerge from adapting and redefining processes within the enterprise. This indicates that DDDM adoption is both a technological and management revolution. Therefore, companies must use data as the basis for each decision to remain competitive in today's fast-paced environment (Lochy 2017), learn new skills and move from traditional business models to data-driven practices.

Notably, digital transformation in financial services across the globe has been catalyzed and accelerated (Hussain et al. 2023; IIF, 2020) and as a result, the status quo of businesses has come to an end. Technology is rapidly changing, causing decision-making more automated and data-driven in the financial sector. Fintechs, on the other hand, are the rising force behind the digitalization of financial institutes. If financial institutes fail to cater to digital disruption, Fintech startups can reduce the role and relevance of established financial institutions (Alam et al. 2019).

Pakistan's banking sector is the most prominent sector in the country to contribute to economic development and GDP growth. About 44% of the country's GDP is represented by the total assets of the banking sector (NFIS, 2019). Yet, only 16% of the population had a bank account in 2015, leaving more than 50% of the population without access to official financial services (SBP, 2020). Given the massive amount of data generated by digital transactions, Pakistan's banking industry is the real beneficiary of big data analytics (BDA). The digitalization of the banking sector in Pakistan took place in 2008, which has improved with increased use of mobile banking, online shopping and digital payments (GSMA, 2019). The banking sector is at the digital transformation stage and numerous banks have begun to invest in data analytics, though big data capabilities are still limited. With careful mining of data from mobile phone and internet usage, multiple internet service providers, and enabling

regulators, Pakistan's banking sector meets all the criteria to be explored for its depth of digitalization and big data adoption. This investigation is necessary for developing policies to help the government and firms direct the needed resources and benefit immensely from the growing digital revolution. Thus, this paper aims to investigate the major implications of digital transformation and big data in Pakistan's banking sector, including investment in data analytics and data-driven decision-making practices.

Due to its association with higher productivity, the digitalization of the banking industry has advanced over time (Mehmood et al. 2015; Gul et al. 2021; Gul and Ellahi 2021; Gul et al. 2023). Although several banks have begun to invest in data analytics, little is known about how these analytics are used in decision-making. Against this backdrop, this study examines the productivity of banks in relation to data analytics and data-driven decision-making.

Our research contributes in a number of ways. First, there has been a scant empirical research on how DDDM affects firms' productivity (Brynjolfsson and McElheran 2019; Gul et al. 2023), despite the fact that DDDM increasingly offers major solutions to the financial sector (Lochy 2017). Second, the adoption of DDDM in the presence of DA may have a different effect on banks' productivity that without investment in DA. Our study would reflect on it by incorporating the moderating role of DA on the relationship between DDDM and productivity. Third, the use of the Instrumental Variable Two-Stage Least Square (2SLS-IV) methodology in the banking sector of Pakistan offers a significant theoretical contribution to the existing literature. This approach addresses endogeneity and omits variable bias, a common concern in studies examining complex relationships within the banking industry. Finally, Pakistan stands relatively high in technological competitiveness in the telecommunication and banking sectors; there is a lag in adopting digitalization at an organizational level for decision-making. It is observed that while there are many players (technology vendors, consulting companies, government, and others) that are actively promoting digital transformation, its adoption by firms remains unexplored at a decision-making level. This study attempts to fill this gap.

## Literature review

Making decisions is the most crucial managerial job, and how well-rounded they are will determine how successful a corporation is since they give them a competitive advantage. BDA offers organizations knowledge-based resources and facilitates automated decision-making, which is quicker, more precise, and more efficient than ever. Algorithm-based decisions extract valuable insight while reducing risks. However, to make efficient and timely decisions, the data must be reliable and accurate and the organizations must be data-driven. Organizations are data-driven if they meet two prerequisites; (i) right data is collected and (ii) accessible to everyone (Anderson 2015). Further, this data should be used in decision-making (Gul et al., 2021, 2023) and the data-driven organization should be forward-looking (Anderson 2015).

Automated decision-making using neural network algorithms and decision trees, which is quicker, more efficient, and possibly more accurate than before, can facilitate more powerful decision-making using data analytics today (Li et al. 2020). As data analyses get increasingly complex, DDDM's potential for organisations increases. DDDM is a process for gathering and analysing data to produce insightful findings and insights, then communicating those insights to interested parties to assist managers in enhancing the performance of their businesses (Schelling and Rubenstein, 2021). Moreover, data-driven decision-making should be adopted at all levels for DDDM practices to be effective

as it calls for an organization that is Data-driven and promotes a culture where analytics and tools guide decision-making (Kiron 2017).

The diffusion of innovation theory (DOI) can shed light on how DA and DDDM can boost banks' productivity. This idea holds that innovation and change can be sparked by accepting and spreading technologies, such as ICT-based financial services, presumably through improved information availability, lower transaction costs, and data-driven decision-making (Ong et al. 2023; Daud and Ahmad 2023).

Brynjolfsson and McElheran (2016) found that DDDM practises have grown dramatically over time in US firms, and performance gain was also a result of significant IT investments. This indicates that for DDDM practices and data and analytics work hand in hand to improve firm performance (Anderson 2015). Similar findings are made by Liberatore et al. (2017), who indicate that businesses utilizing DDDM practices are typically 5–6% more productive than those who do not adopt DDDM. According to Acharya et al. (2018), data can help with knowledge co-creation, which promotes the use of evidence in decision-making and enhances business success. Long (2018) further reveals a strong nexus between DDDM and key profitability ratios and production plants' productivity increases when data is used in decision-making (Brynjolfsson and McElheran 2019).

Embedded on both the innovation diffusion theory and empirical evidence, the following hypotheses are derived to be tested in essay 2.

**H1**: The DDDM significantly affects the productivity of Pakistan's banking sector.

Exhaustive literature suggests that investment in big data and analytics enhances organizational performance (Awan et al. 2022; Gul et al. 2021; Gul and Ellahi 2021; Carbó-Valverde et al. 2020; Shamim et al. 2020; Mikalef et al. 2019; Muller et al. 2018). The investment in DA is valuable if an organization adopt DDDM. However, how investment in DA affects decision-making and the subsequent firm performance is largely unexplored. It is assumed that DDDM will adopt in organizations that have invested in big data analytics or have digitally transformed. Therefore, we aim to investigate how investment in DA affects the relationship between DDDM and banks' productivity. The second hypothesis is as follows:

**H2**: DA has a significant impact on the relationship between DDDM and productivity of banks in Pakistan.

## Materials and methods

**Sample and data collection**. All commercial and microfinance banks registered with the State Bank of Pakistan are included in our sample except foreign and specialist institutions. This left us with 26 commercial and 10 microfinance banks, and a total of 180 firm level observations for 2016 to 2020. A structured survey filled by the chief information officer, data analyst, IT heads, and/or senior bank management of each bank in the sample served as the primary method of gathering data. The survey asks questions on

investing in DA, utilising IT by employees, and adjusting to organisational change brought on by the adoption of DDDM. Secondary data was collected from online publicly accessible sources, including State Bank of Pakistan, Banks' websites and annual reports. Finally, primary and secondary data was combined for econometric analysis.

### Construction of variables

*Primary data.* We contacted IT heads, Data analysts, bank managers, and information executives in all banks included in our sample through LinkedIn, emails, and personal visits. The online survey was shared through google docs, whereas the hard copy was filled in during personal visits. All 36 banks participated in this survey. Since we adopted the questionnaire (Brynjolfsson and McElheran 2019), we did not validate it further. However, we conducted pilot testing to identify any issues with item clarity, response options, and instruction to avoid potential problems and refinement of the questionnaire before the actual data collection began. A summary of all variables based on primary data is presented in Table 1.

*Secondary data.* Output (Y) is the dependent variable that is measured through the Cobb-Douglas production function. The banking industry's output differs from other industries' because of its unique business structure, therefore instead of units sold, we use loans and deposits as the output variable (Martin-Oliver and Salas-Fumas 2008). Production function input variables include capital and employees. We also used control variables, including Z_Score, non-performing loans, type of banks, and listing on the Stock exchange (List). The summary of secondary variables is presented in Table 2.

**Methods**. Previous empirical evidence suggests endogeneity and reverse causality issues exist in studies related to investment in IT and firm performance (Brynjolfsson and McElheran 2016, 2019; Tambe 2014). Given that our research involves DDDM and DA, we want to use such estimation technique that addresses reverse causality and endogeneity. Against this backdrop, we employ 2SLS-IV to estimate the impact of DA and DDDM on banks productivity (Brynjolfsson and McElheran 2019; Muller et al. 2018). Further, a suitable instrument or instruments are necessary for IV estimation that affects the explanatory variable and assist explaining the causal relationship between the independent and dependent variables. The instrument must be correlated with the endogenous variable but uncorrelated with the error term in the explanatory equation in order to estimate the IV; otherwise, it will experience the previous endogeneity issue once more. When an instrument meets these requirements, the general IV model can be calculated using the two-stage least squares estimator. The instruments for DDDM include lagged output variable and adjustment cost for the change in banks since a minimum of one instrumental variable must be related to an endogenous variable (such as DDDM) but unrelated to the dependent variable (such as

| Table 1 Construction of variable- primary data. | | |
|---|---|---|
| **Variables (Notations)** | **Definitions** | **Sources** |
| Data analytics (DA) | DA is a dummy variable, '1' for banks that have invested in DA and '0' otherwise. | Gul and Ellahi (2021); Muller et al. (2018) |
| DDDM | A composite index created by PCA based on 5 survey questions. | Gul et al. (2023); Brynjolfsson and McElheran (2019) |
| Adjustment cost (AdjC) | A composite measure created via PCA using six survey items | Brynjolfsson et al. (2011) |
| Exploration (EXPL) | A composite indicator created using PCA using six survey questions. | Brynjolfsson et al. (2011) |
| Human capital (HC) | A composite measure produced via PCA using five survey items | Brynjolfsson and McElheran (2016) |

**Table 2 Construction of variable- primary data.**

| Variables (Notations) | Definitions | Sources |
|---|---|---|
| Output (Y) | Sum of deposits and non-bank Loans | Martin-Oliver and Salas-Fumas (2008) |
| Capital (K) | Net fixed assets | Gul et al. (2023) Octrina and Setiawati (2019) |
| Employees (L) | Full-time employees | Gul et al. (2023); Brynjolfsson and McElheran (2019) |
| Age of investment in data analytics (DA-Age) | Number of years of investment in data analytics | Brynjolfsson and McElheran (2019) |
| Z-Score | The ratio of sum of return on assets and capital asset ratio to ROA's standard deviation. | Boyd and Crawford (2012) |
| Non-performing loans (NPL) | A ratio of NPL to total loans | Octrina and Setiawati (2019) |
| Type (Type) | Dummy variable, using '0' for microfinance banks and 'I' for the commercial banks | Gul et al. (2021) |
| Listing (LIST) | Dummy variable 1 if a bank is listed on the Pakistan Stock Exchange and to "0" if not listed. | Gul et al. (2021) |

**Table 3 Diagnostic test for PCA.**

| | Bartlett test of sphericity | Kaiser-Meyer-Olkin |
|---|---|---|
| DDDM | Chi-square = 25.746 (0.004) | 0.751 |
| Human capital (HC) | Chi-square = 24.062 (0.007) | 0.683 |
| Exploration (EXPL) | Chi-square = 108.425 (0.000) | 0.754 |
| Adjustment cost (AdjC) | Chi-square = 180.555 (0.000) | 0.784 |

Barlett test of spehricity H0: variables are not intercorrelated. The $p$ values are in parenthesis.

banks' performance) in IV regression (Brynjolfsson and McElheran 2019). However, for our second model, the instruments include lagged output variable and exploration. Exploration is a measure of innovation and banks' tendency to explore new markets or invest in new technology (DA in our case). Lagged output variable will control the endogeneity issue in the model and give us consistent DDDM estimates. These two instruments (exploration and adjustment cost to change) meet the criteria discussed above.

PCA was used to build the DDDM, Exploration, Adjustment Cost to Change, and Human Capital indexes. PCA is a statistical technique for lowering the dimensionality of data while preserving many variations (Jalil et al. 2010). The best method for creating indices for DDDM and other control variables is to use PCA, which shrinks the amount of the data while capturing data variation (Jalil et al. 2010). Before the development of indices, the results of preliminary tests, such as the Bartlett test of sphericity and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy, are reported in Table 3 below. PCA can be employed for these variables because all of the variables' values for the Kaiser-Meyer-Olkin test are larger than 0.6 and the items are correlated (Kaiser and Rice 1974). For analysis, only the components with eigenvalues greater than one are maintained. Tables 4 presents the explained variance for component 1 of the DDDM and associated factors.

**Model specification**. The following model was used to determine the effect of DDDM and DA on banks' productivity.

$$
\begin{aligned}
\log(Y_{it}) = {} & \beta_0 + \beta_1 DDDM_{it} + \beta_3 \log(K_{it}) \\
& + \beta_3 \log(L_{it}) + \beta_2 \log(ITE_{it}) \\
& + \sum_{i=0}^{t} \gamma_i X_{it} + \sum_{i=0}^{t} \beta_i X_{it} + U_{it}.
\end{aligned}
\tag{1}
$$

**Table 4 Principal component analysis.**

| | Eigenvalue (Component 1) | Variable (%age) |
|---|---|---|
| DDDM | 1.622 | 0.678 |
| Human capital (HC) | 2.2064 | 0.6374 |
| Exploration (EXPL) | 3.668 | 0.6103 |
| Adjustment cost (AdjC) | 2.984 | 0.751 |

The following model will check the impact of DA on DDDM and banks' productivity.

$$
\begin{aligned}
\log(Y_{it}) = {} & \beta_0 + \beta_1 DDDM_{it} + \beta_2 \log(K_{it}) \\
& + \beta_3 \log(L_{it}) + \beta_4 DA_{it} + \beta_4 DA_{it}*DDDM_{it} \\
& + \sum_{i=0}^{t} \gamma_i X_{it} + \sum_{i=0}^{t} \beta_i X_{it} + U_{it}.
\end{aligned}
\tag{2}
$$

For above equations, $Y_{it}$, calculated as the log of the total of the loans and investments, is the result of the extended Cobb-Douglas production function. Since the banking industry in Pakistan serves as a middleman between borrowers and lenders (Gul et al. 2023), this study's measurement of bank productivity has been done using the intermediate approach. According to the asset approach's intermediate method, banks act as the go-betweens for borrowers and suppliers (Kovner et al. 2014), hence labor and capital are the main inputs in addition to IT investment. The value of the index $DDDM_{it}$ can range from 0 to 1. $L_{it}$ is the quantity of full-time employees, while Kit is the fixed assets. $DA_{it}$ is the investment in DA. $\sum \gamma_i X_{it}$ represents DDDM-related control variables, including human capital. $\sum \beta_i X_{it}$ includes control variables whereas $U_{it}$ represents the error term.

## Discussion of results

**Data and descriptive statistics.** Tables 5 and 6 show the descriptive statistics for each primary and secondary variable, respectively. Primary data was collected to measure four variables, written in bold in Table 5. The value of the DDDM's Cronbach's alpha, along with other control variables on a 5-point Likert scale linked to exploration, adjustment costs, and human capital, are shown in Table 5. The value of Cronbach's alpha is 0.691, 0.893, 0.900 and 0.755 for DDDM, Adjustment costs, exploration and human capital, respectively. These values are consistent with previous literature.

The descriptive statistics of secondary variables are given in Table 5. There are 178 firm-level observations. The mean non-performing loan ratio is less than one, indicating banks' productivity in terms of loans collection. The mean Z_Score value is 9.75, indicating that banks are stable and robust. With an annual output of Rs. 131 billion, banks are typically quite large in Pakistan. Type and listing are dummy variables with a range of 0 to 1. Tables 6 and 7 below shows the pairwise Pearson correlation analysis, which shows no collinearity among independent variables.

**Empirical findings.** The initial findings to gauge the effect of DDDM practises on banks' productivity and the moderating effect of DA on the relationship between DDDM and productivity are presented in Table 8 using IV-2SLS/IV estimation technique. The results regarding the impact of DDDM on banks' output exhibit suggest that DDDM causes banks' output to be increased by 10.5%. The coefficient estimate is statistically significant at 1%. We used lagged output as an instrument for DDDM to confirm that our estimates are robust. As DDDM is an endogenous variable, we also used another instrument, 'adjustment cost to change,' which is related to DDDM but unrelated to banks' output. Thus, the coefficient estimate of DDDM does not suffer endogeneity and omitted variable bias.

Additionally, Sargan examines the null hypothesis of over-identification, which was rejected at 0.12, indicating that the choice of instruments did not affect our findings. Two of our control variables, Z_Score and Type, are significant, showing that banks stability increases if DDDM is adopted; however, commercial banks perform better than microfinance banks regarding the DDDM adoption. It seems that commercial banks are large in size, they have kept data-driven decision-making intact and can translate into improved productivity. Other control variables, including HC, Non-performing loans and listing on the stock exchange are insignificant. Thus, hypothesis 1 cannot be rejected and it is concluded that DDDM has a statistically significant and positive impact on banks' output.

We also incorporated data analytics to assess its moderating impact on the relationship between DDDM and banks' output. DA has a significant impact on Bank's output and the coefficient estimate of DA of 0.067 shows that investment in DA causes an increase of 6.7% in banks' output. It is important to note that the coefficient estimate of DDDM remains almost the same 0.0966. The moderating impact of DA is also positively significant, indicating that the impact of DDDM on banks' output is enhanced by 4.43% if banks invest in data analytics. We used two instruments: exploration, adjustment cost and lagged dependent variable for DDDM to ensure robustness and consistent estimates of DDDM. Since we introduced DA in our model, we used exploration as an additional instrument as it measures the measure innovation and banks' tendency to explore new markets or invest in new technology such as DA. Our instruments are valid and passed the weak instrument tests at Wald Chi-square of 14.583 with p-value = 0.00. Thus, we fail to reject the null hypothesis and conclude that DA significantly and positively impacts the relationship between DDDM and banks' output. If banks invest in DA, the impact of DDDM on banks' output is strengthened by 4.43%.

## Discussion of results and conclusion

This study is one of the pioneer studies investigating the primary role of big data, translating it into decision-making to get the best use of investment in analytics. We employed IV-2SLS to investigate big data impacts on banks' productivity. We performed the analysis in two steps: first, we measured the effect of DDDM on banks' output and then we incorporated DA to check if it affects the relationship between DDDM and banks' output. Our findings suggest a positive and significant impact of DDDM on banks' output (10.5%), which is strengthened by 4.43% if banks invest in DA. Our results are robust and give consistent estimates of

### Table 5 Descriptive statistics of survey items.

| | Cronbach's Alpha |
|---|---|
| **Measure 1: data driven decision making (DDDM)**<br>We rely on data for the creation of a new product or service<br>We depend on data to support our decision making (the work practices and environment of the entire company)<br>We have the data we need to make decisions<br>feedback based on data is integrated into existing system<br>data is preferred over management's' experience and intuition. | 0.6916 |
| **Measure 2: Adjustment Cost (AdjC)**<br>Skill mix of existing staff<br>Employment contracts<br>Work Rules<br>Organizational culture<br>customer relationships<br>senior management involvement | 0.8933 |
| **Measure 3: Exploration (EXPL) IT facilitates ----**<br>to offer new products<br>to explore new markets<br>growth opportunities that are competitive<br>transforming our business<br>new business capabilities supported by technology<br>substantial changes to traditional processes | 0.9003 |
| **Measure 4: General Human Capital (HC)**<br>The importance of education for hiring<br>% of employees using PC/terminals/workstations<br>% of employees using e-mails<br>% of non-IT employees using DA in the office | 0.7551 |

### Table 6 Descriptive statistics of production function variable.

| Variable | N | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Y | 178 | 8.120745 | 0.9040256 | 5.684095 | 9.48715 |
| DDDM | 178 | 0.0215304 | 1.425276 | −3.235694 | 2.89548 |
| DA | 178 | 0.4438202 | 0.04982353 | 0 | 1 |
| Log(K) | 178 | 6.500109 | 0.8407309 | 3.869232 | 7.878187 |
| Log(L) | 178 | 3.462727 | 0.5087566 | 1.973128 | 4.324468 |
| NPL ratio | 178 | 0.0784761 | 0.0707019 | 0 | 0.442291 |
| Z_Score | 178 | 9.753443 | 16.2418 | −31.2393 | 74.3001 |

**Table 7 Correlation between DDDM and IT expense and employees.**

|  | Y | DDDM | DA | K | L | NPL ratio | Z_Score |
|---|---|---|---|---|---|---|---|
| Y | 1.0000 |  |  |  |  |  |  |
| DDDM | 0.1992 | 1.0000 |  |  |  |  |  |
| DA | 0.2820 | −0.2027 | 1.0000 |  |  |  |  |
| K | 0.9502 | 0.2520 | 0.3471 | 1.0000 |  |  |  |
| L | 0.9041 | 0.2557 | 0.2761 | 0.9170 | 1.0000 |  |  |
| NPL ratio | 0.1550 | −0.1087 | 0.0024 | 0.1835 | 0.0590 | 1.0000 |  |
| Z-Score | 0.4128 | 0.0193 | 0.1823 | 0.3296 | 0.4050 | −0.1316 | 1.0000 |

The variables in Table 7 are presented in bold to emphasize their significance and importance within the context of the survey items. These measures are essential for assessing various aspects of the organization's DDDM, adjustment costs, exploration capabilities, and general human capital. The bolded values are indicative of the overall scores (Cronbach's Alpha) associated with each respective measure, providing a summary of their significance within the study.

**Table 8 IV-regressions of DA and DDDM on productivity measures.**

| DV = Log(output) | 2SLS/IV (2016–2020) | 2SLS/IV (2016–2020) |
|---|---|---|
| DDDM | 0.105*** (0.299) | 0.0996*** 0.033 |
| DA |  | 0.0676* (0.041) |
| DA*DDDM |  | 0.0443 (0.165) |
| HC | −0.0254* (0.013) |  |
| Log (K) | 0.288*** (0.134) | 0.2301*** (0.0764) |
| Log (L) | 0.662*** (0.0847) | 0.7156*** (0.082) |
| Type | 0.791*** (0.056) | 0.8127*** (0.068) |
| List | −0.002 (0.0436) | −0.0190 (0.038) |
| NPL | −0.397 (0.2319) | −0.234 (0.2319) |
| Z_Score | 0.006*** (0.001) | 0.005*** (0.009) |
| Observations | 142 | 142 |
| Adj. R-squared | 0.637 | 0.536 |
| OverID test (Sargan Test) | 4.23 | 3.98 |
| Chi-sq (p value) | 0.134 | (0.157) |
| Weak instrument: | 12.35 | 14.583 |
| Wald Chi-square (p value) | 0.0004 | (0.000) |
| Endogeneity Chi-sq (P value) | 7.989 (0.0023) | 9.202 (0.000) |

***$p < 0.001$, **$p < 0.05$, *$p < 0.01$.

DDDM and DA. Our findings align with diffusion innovation theory and previous literature (Gul et al. 2023; Kim et al. 2021; Shamim et al. 2020; Rialti et al. 2019).

Our findings suggest that banks should ensure better quality and that audited data is available to the right individuals in real time. Banks would be encouraged to make decisions in real-time and continue to be productive if DA is exploited and data is available for decision-making. As a result, investing in analytics, implementing DDDM, and picking up new skills would boost banks' productivity. Given the financial sector's increased investment in IT and analytics, DDDM adoption and bank productivity will likely rise. Moreover, implementing DDDM can also lead to better risk management and cost reduction for banks. With the help of data-driven insights, banks can identify potential risks and take proactive measures to mitigate them, resulting in a more secure and stable financial system. Additionally, banks can lower operational costs and increase profitability by leveraging analytics to optimize processes and reduce inefficiencies.

**Theoretical contributions**. Applying the Instrumental Variable Two-Stage Least Square (2SLS-IV) methodology in the banking sector in Pakistan offers a notable theoretical contribution to the existing literature. By employing this approach, we address the endogeneity issue, a common concern in studies examining complex relationships within the banking industry. Incorporating the 2SLS-IV methodology enhances the empirical validity of our findings and strengthens the causal interpretation of the relationships under investigation.

**Managerial implications and policy relevance**. The findings of the present study have many implications for Pakistani shareholders, investors, bank managers, legislators, and regulators. With the help of well-organized data-driven decision-making practices, businesses can leverage their resources to provide relevant insights when and where needed. It will help incumbents develop real expectations before investing in analytics and adopting DDDM and learn the benefits of automated decision-making compared to manual decision-making for various banking functions, including lending, investment, and many others. Third, it is already established that a bank's bankruptcy might have a domino effect, and that one bank's riskiness could endanger the stability of the entire banking industry. Banks can therefore balance too much risk exposure and too little risk presence while reducing risk using DDDM. *Fourth*, in developing countries like Pakistan, banks must efficiently channel the economy's resources. Suppose banks fail to direct the country's financial resources to the masses, which is impossible without the support of big data and DDDM. In that case, it might have disastrous consequences for the whole economy.

**Limitations of study and future work**. It is important to note that because our study focuses on the banking industry in Pakistan, the conclusions may not be immediately relevant to other businesses or developed countries with different contextual factors. In addition, we focused on productivity as a performance measure only, whereas other potential outcomes, such as risk exposure, financial performance and others, could be relevant to DDDM and DA. Future research should also look at other industries, such as healthcare, telecommunications, and other service industries, as this study only looked at the financial sector. Additionally, it is critical to consider the difficulties that organizations and decision-makers may encounter when utilizing DDDM and to devise solutions for them. Future studies should investigate the impact of DDDM on various performance metrics, including profitability, credit and risk management. Moreover, it would be beneficial to explore the potential of combining DDDM with other emerging technologies, such as artificial intelligence, to enhance decision-making processes in the financial sector. As the financial industry continues to evolve, understanding the

implications and benefits of DDDM will be crucial for staying competitive and successful.

## Data availability

## References

Acharya A, Singh S, Pereira V, Singh P (2018) Big data, knowledge co-creation and decision-making in fashion industry. Int J Inf Manag 42:90–101. https://doi.org/10.1016/j.ijinfomgt.2018.06.008

Alam N, Gupta L, Zameni A (2019) Digitalization and disruption in the financial sector. In Fintech and Islamic Finance (pp. 1-9). Palgrave Macmillan, Cham

Anderson C (2015) Creating a data-driven organization: Practical advice from the trenches. Data-Driven Healthcare, 55–65. https://doi.org/10.1002/9781119205012.ch5

Awan U, Bhatti SH, Shamim S, Khan Z, Akhtar P, Balta ME (2022) The role of big data analytics in manufacturing agility and performance: moderation–mediation analysis of organizational creativity and of the involvement of customers as data analysts. Br J Manag 33(3):1200–1220

Boyd D, Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Inf Commun Soc 15(5):662–679. https://doi.org/10.1080/1369118x.2012.678878

Brynjolfsson E, Hitt LM, & Kim HH (2011) Strength in numbers: How does data-driven decision making affect firm performance? Available at SSRN 1819486

Brynjolfsson E, McElheran K (2016) The rapid adoption of data-driven decision-making. Am Econ Rev 106(5):133–139. https://doi.org/10.1257/aer.p20161016

Brynjolfsson E, McElheran K (2019) Data in action: data-driven decision-making and predictive analytics in US manufacturing. Rotman School of Management Working Paper. https://doi.org/10.2139/ssrn.3422397

Carbó-Valverde S, Cuadros-Solas PJ, Rodríguez-Fernández F, EY (2020) The effect of banks' IT investments on the digitalization of their customers. Global Policy 11:9–17

Conejero JM, Preciado JC, Prieto AE, Bas MC, Bolós VJ (2021) Applying data-driven decision-making to rank vocational and educational training programs with TOPSIS. Dec Support Syst 142:113470

Daud SNM, Ahmad AB (2023) Financial inclusion, economic growth and the role of digital technology. Finance Res Lett 53:103602

Erickson GS, Rothberg HN (2018) Intangible dynamics: Knowledge assets in the context of big data and business intelligence. Anal Knowl Manag 325–354. https://doi.org/10.1201/9781315209555-11

GSMA (2019) The power of mobile to accelerate digital transformation in Pakistan. GSMA. https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2020/06/The-Power-of-Mobile-to-Accelerate-Digital-Transformation-in-Pakistan-2019.pdf

Gul R, Ellahi N (2021) The nexus between data analytics and firm performance. Cogent Business & Management 8(1):1923360

Gul R, Ellahi N, Leong K, Malik QA (2021) The complementarities of digitalisation and productivity: redefining boundaries for financial sector. Technol Anal Strat Manag 1–13

Gul R, Leong K, Mubashar A, Al-Faryan MAS, Sung A (2023) The empirical nexus between data-driven decision-making and productivity: Evidence from Pakistan's Banking Sector. Cogent Bus Manag 10(1):2178290

Hussain S, Gul R, Ullah S (2023) Role of financial inclusion and ICT for sustainable economic development in developing countries. Technol Forecast Soc Change 194:122725

IIF (2020) Realizing the digital promise. Institue of International Finance. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Financial-Services/gxfsi-realizing-the-digitalpromise-covid-19-catalyzes-and-accelerates-transformation.pdf

Jalil A, Feridun M, Ma Y (2010) Finance-growth nexus in China revisited: New evidence from principal components and ARDL bounds tests. Int Rev Econ Finance 19(2):189–195. https://doi.org/10.1016/j.iref.2009.10.005

Kaiser H. F, Rice J (1974) Little jiffy, mark IV. Educ. Psychol. Meas 34(1):111–117

Kim J, Dibrell C, Kraft E, Marshall D (2021) Data analytics and performance: The moderating role of intuition-based HR management in major league baseball. J Bus Res 122:204–216

Kiron D. (2017) Lessons from becoming a data-driven organization. MIT Sloan Management Review 58(2). https://sloanreview.mit.edu/case-study/lessons-from-becoming-a-data-driven-organization/

Kovner A, Vickery J. I, Zhou L (2014) Do big banks have lower operating costs? Eco Policy Rev 20(2):1–27

Li T, Ma L, Liu Z, Liang K (2020) Economic granularity interval in decision tree algorithm standardization from an open innovation perspective: Towards a platform for sustainable matching. J Open Innov 6(4):149

Liberatore MJ, Pollack-Johnson B, Clain SH (2017) Analytics capabilities and the decision to invest in analytics. J Comput Inf Syst 57(4):364–373

Lochy J (2017) Big data in the financial services industry—from data to insights, Finextra. https://www.finextra.com/blogposting/17847/big-data-in-the-financial-services-industry---from-data-to-insights

Long Q (2018) Data-driven decision making for supply chain networks with agent-based computational experiment. Knowl.-Based Syst. 141:55–66

Martin-Oliver A, et Salas-Fumás S. -F. V (2008) The output and profit contribution of information technology and advertising investments in banks. J. Financ. Intermediation 17(2):229–255. https://doi.org/10.1016/j.jfi.2007.10.001

Mehmood B, Nisar A, et Ur Rehman H (2015) Technology matters: Evidence from Pakistani banking sector using flexible transcendental logarithmic production function. Pakistan Econ. Soc. Rev 53(2):203–224. www.jstor.org/stable/26153257

Mikalef P, Boura M, Lekakos G, Krogstie J (2019) Big data analytics and firm performance: Findings from a mixed-method approach. J Bus Res 98:261–276

Muller O, Fay M, vom Brocke J (2018) The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. J Manag Inf Syst 35(2):488–509. https://doi.org/10.1080/07421222.2018.1451955

NFIS (2019) Digital Financial Services (DFS) – Innovation Challenge Facility. SBP. https://www.sbp.org.pk/Guidelines/ACMFD/2018/DFS-Guidelines.pdf

Octrina F, Setiawati R (2019) Competitiveness of Indonesian banking industry based on commercial bank business group: Panzar Rosse Model. Jurnal Perspektif Pembiayaan dan Pembangunan Daerah 7(1):37–48. https://doi.org/10.22437/ppd.v7i1.7475

Ong HB, Wasiuzzaman S, Chong LL, Choon SW (2023) Digitalisation and financial inclusion of lower middle-income ASEAN. Heliyon 9(2):e3347

Rialti R, Zollo L, Ferraris A, Alon I (2019) Big data analytics capabilities and performance: Evidence from a moderated multi-mediation model. Technol Forecast Soc Change 149:119781

SBP (2020) Digitalization of services in Pakistan. State Bank of Pakistan. Available from: https://www.sbp.org.pk/reports/annual/arFY18/Chapter-07.pdf

Schelling N, Rubenstein L. D (2021) Elementary teachers perceptions of data-driven decision-making. Educ. Assess. Evaluation Account. 33:317–344

Shamim S, Zeng J, Khan Z, Zia NU (2020) Big data analytics capability and decision making performance in emerging market firms: The role of contractual and relational governance mechanisms. Technol Forecast Soc Change 161:120315

Tambe P (2014) Big data investment, skills, and firm value. Manag Sci 60(6):1452–1469. https://doi.org/10.1287/mnsc.2014.1899

Welbounre TM (2015) Data-driven storytelling: The missing link in HR data analytics. Employment RelatToday 41(4):27–33. https://doi.org/10.1002/ert.21471

## Competing interests

The authors declare no competing interests.

## Ethical approval

This survey was conducted during my PhD (Management Sciences) from Foundation University Rawalpindi Campus, Pakistan in 2020 and this particular survey like other similar surveys of not harming the anonymity of organizations, disclosing sensitive information, etc. are exempted from ethical approval by the university. Moreover, this data was related to the use of big data, DDDM and DA, no written ethical approval was required by the university. However, the anonymity of participants and their organizations was maintained.

## Informed consent

Throughout the course of this study, participants were explicitly informed about our survey's scope. They were guaranteed complete anonymity, reaffirming our commitment to ethical research practices and data confidentiality.

## Additional information

**Correspondence** and requests for materials should be addressed to Raazia Gul.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.