HONG KONG INSTITUTE FOR MONETARY AND FINANCIAL RESEARCH

ASSESSING THE IMPACT OF DIGITALISATION ADOPTION ON BANK'S FINANCIAL PERFORMANCE: NEW EVIDENCE BASED ON TEXTUAL ANALYSIS

Andrew Wong, Stephen Lam and Kelvin Ho

HKIMR Working Paper No.07/2024

July 2024

HONG KONG ACADEMY OF FINANCE
金融學院

Assessing the impacts of digitalisation adoption on banks' financial performance: New evidence based on textual analysis[1]

Andrew WONG[2], Stephen LAM[3] and Kelvin HO[4]

July 2024

Abstract

This paper introduces a novel textual measure of digital technology exposure for a large sample of listed banking institutions, using a topic modelling algorithm called *Top2Vec* on their earnings call transcripts. Our novel approach reveals a strong correlation between this measure and banks' actual technological advancements, validating its usefulness as a gauge for the progress of banks' digital transformation. By examining the potential effects of digitalisation on banks' performance using this textual measure, our findings provide fresh evidence to confirm the benefits of adopting digital technologies on banks' business and operations. Specifically, banks with higher digitalisation adoption exposure tend to exhibit improved cost-efficiency, better asset quality, and a stronger capital and liquidity position compared to other banks. In addition, our results show that higher adoption of digitalisation by banks can have positive signalling effects on their future profitability, which in turn improves their market valuation as perceived by market participants. Given that there is significant variation in the extent of digitalisation adoption among banks globally, these findings underscore the importance to policymakers of promoting digitalisation in the banking sector to capitalise on its benefits and to maintain competitiveness.

[1] This paper represents the views of the authors, which are not necessarily the views of the Hong Kong Monetary Authority, Hong Kong Academy of Finance Limited, or Hong Kong Institute for Monetary and Financial Research. The above-mentioned entities except the authors take no responsibility for any inaccuracies or omissions contained in the paper.

[2] Wong (Corresponding author): acmwong@hkma.gov.hk
[3] Lam: slhlam@hkma.gov.hk
[4] Ho: kkwho@hkma.gov.hk

1. Introduction

The rapid advancement of digital technologies, such as artificial intelligence (AI), has significantly impacted various industries, including the banking sector. The integration of these technological innovations into traditional banking practices has the potential to significantly enhance banking performance. The potential benefits have prompted banks worldwide to embrace digital transformation, aiming to modernize their business models and generate new value-creation opportunities. For instance, Hornuf et al. (2021) reveals that around 60% of sampled banks from four advanced economies in Europe and North America have already adopted a digital strategy. This indicates the widespread adoption of digital transformation in the banking sector. Apart from survey results, the growing interests and adoption of digital innovations by banks can be observed from the increasing references to the "tech" word in banks' earnings call and at investor conference events (see the news reported by the Business Insider on 5 Jan 2023[5]).

In view of the significant resources and efforts devoted globally in adopting digital and technological innovations, it is useful to evaluate whether, and to what extent, these digital innovations have impacted bank performance so far. While some recent studies examine these issues for the banking industry as a whole, most rely on samples from a single jurisdiction only and so evidence covering a broader context remains scant. A major challenge in assessing impacts is the measurement of banks' adoption of digital technologies, which is not readily available in conventional financial data. Existing studies attempt to gauge the extent of banks' digitalisation adoption through surveys, but this approach is resource intensive and challenging to conduct over time for a wide set of international banks.

To tackle this data challenge, this paper proposes an unsupervised natural language processing (NLP) method for analysing banks' earnings call transcripts to gauge the degree of banks' digitalisation. In brief, we employ a novel clustering-based topic model Top2vec algorithm (Angelov, 2020) to identify topics, content, and associated words within earnings call transcripts using an unsupervised learning approach. The Top2vec algorithm offers several advantages over other topic modelling

[5] Link to news article: https://www.businessinsider.com/why-banks-are-losing-the-war-against-fintechs-2023-1.

methods, allowing for a more accurate measurement of the targeted concept, in this case, banks' digital exposure. This makes it easier to apply and repeat the process. Our constructed proxy measure has a positive association with technological-related product announcements or business expansion events in the same calendar year. This supports the usefulness of the textual variable in gauging banks' digitalisation adoption, which can be used for empirical analyses. More detailed discussion on the construction of relevant textual measures from banks' earnings call transcripts is provided in Section III. Using our developed textual measure of banks' digitalisation adoption, this study examines the potential impact on banks' performance and fundamentals across a wide sample of global banks.

Several key findings are uncovered in this analysis. Firstly, our textual measure reveals that there has been a rapid pace of digitalisation adoption by banks in recent years. However, there is significant heterogeneity among banks, suggesting that some institutions may be lagging behind in the digital transformation trend which may undermine their long-term competitiveness. Second, there is empirical associative evidence to suggest that banks have benefitted from adopting digital technologies. Consistent with the potential benefits of digital innovations, banks with a higher degree of digitalisation adoption tend to exhibit greater cost efficiency, better asset quality, and a stronger capital and liquidity position compared with other banks. Furthermore, the extent of banks' digitalisation adoption may have positive signalling effects to market participants regarding their future profitability prospects which can help to improve their market valuation. A key policy implication from these results is that it is crucial for policymakers to encourage banks to continue enhancing the digital transformation of their business to maintain their competitiveness.

The structure of this article is as follows. Our main research questions and related literature review are laid out in Section II. Section III describes the textual analytic algorithms and the source of the data employed in this study. Section IV presents empirical questions and results on the relationship between the degree of a bank's digital technology adoption exposure and various dimensions of banking sector performance. The final section concludes.

2. Related literature reviews

This study delves into various streams of literature. Firstly, it explores the application of Natural Language Processing (NLP) techniques to gauge the digitalisation of banks. As digital technology adoption in the banking sector continues to rise, numerous studies have employed a textual analytic approach to measure digital technological exposure among banks. This is because conventional financial statement information, such as IT expenditure, often fails to provide an accurate measurement of the digitalisation of banks. For example, Cheng and Qu (2020), Qi and Cai (2020) and Wu et al. (2021) measure banks' digitalisation exposure using the frequency of Fintech-relevant keywords in banks' annual reports or news articles and study their financial performance implications. Chen et al. (2019), Fang et al. (2023) and Hasan et al. (2023) utilise patent filings by Chinese banks to measure bank's technological adoption progress.

Each of these approaches has its merits and limitations in measuring the targeted concepts. For instance, fintech patent filings provide clear evidence of actual innovations by banks, but the number of filings may be skewed towards larger banks, limiting the comparability of measures across economies. Xie and Wang (2023) survey existing approaches for measuring the digital transformation of commercial banks in China and discuss the pros and cons of each method. On the whole, there is little consensus in the literature on which approach provides a comprehensive measure of digitalisation.

This study differs from existing studies on measuring bank digital technology exposure by utilising earnings call transcripts to construct a measure of digitalisation. Earnings conference calls provide financial market participants and banks with regular venues to discuss recent performance, current business strategies, and future prospects from a bank-specific perspective. Given the rich information content in these calls, various academic studies have applied textual analysis to measure non-financial firm-specific concepts across different dimensions, such as political risk (Hassan et al. 2019), corporate culture (Li et al. 2021b), climate exposure (Sautner et al. 2023; Li et al. 2024), output gap and inflation (Gosselin and Taskin, 2023), bank business sentiment (Soto, 2021), and green innovation premia Leippold and Yu (2024). The literature supports the usefulness of textual data to identify firm-specific exposure to new and non-financial concepts.[6] Such textual information may also be relevant to investor decisions Heinrichs et al. (2019) find that institutional investors are key stakeholders in earnings conference calls, and their participation can potentially contribute to increasing their shareholdings. Therefore, we have applied a state-of-the-art unsupervised topic modelling algorithm (Top2vec) to extract relevant information content from earnings call transcripts, and we then study whether such textual measures can proxy banks' digital technology adoption.

Using a novel measure of the banks' digitalisation adoption, this study aims to contribute to the literature by assessing whether banks benefit from greater digitalisation in their performance. Recent studies provided evidence of potential benefits from "digital" and "fintech" innovations on banks' accounting-based financial performances, such as capital, asset quality, cost efficiency, liquidity, and earnings. While Zhao et al. (2022) find that fintech development in the Chinese economy can lower Chinese bank profitability and asset quality in aggregate due to greater competition. However, other studies suggest that banks may benefit from actively embracing these trends. Bian et al. (2023) find that higher fintech adoption is associated with higher profitability and lower cost-to-income ratios for 181 sampled Chinese banks, as new technology adoption can expand banks' distribution channels, enable automatic business processes, and reduce dependence on human and branch networks. Cheng and Qu (2020) and Zhang et al. (2023) find that Fintech adoption reduces the credit risk faced by banks, which in turn benefits asset quality and banks' capital ratios. Fang et al. (2022) and Guo and Zhang (2023) find that Chinese banks with greater technological exposure achieve higher liquidity creation, potentially through changing banking business structure, deposit inflows, and risk management channels. Overall, these studies suggest a positive role for fintech and digitalisation adoption within banks' business models.

However, the above findings are largely based on a specific country cases. It remains unclear whether the results can be generalised across different regions as banks

[6] While unsupervised natural language processing method for analyzing earnings calls can mainly refer to what banks claim to be doing without necessarily referring to actual actions taken to support these claims, we consider the presence of shareholders and non-affiliated research analysts in the call conference can lower the tendency for bank managements to merely walk the talk on related issues. For instance, investors and research analysts can question any uncorroborated subjects by the managements in Q&A or follow-up sections, in turn limiting the management's capacity to "walk the talk" on digitalisation subjects during the conference call events.

are subject to different operating rules and business environments. To address this gap in the literature, our study expands on previous research by examining a sample of globally listed banks between the period of 2013 and 2022. This allows us to provide results based on a global perspective and assess whether there is a positive impact of digitalisation on banking sector performance that is consistent across different regions.

In addition to the impact of digitalisation on banks' accounting indicators, our study also examines valuation implications of increased exposure to digital technology by banks. This aspect is less well studied but can provide insights into the future performance of banks. Intuitively, while the fintech revolution could pose a threat to bank business models by intensifying competition among financial institutions, banks that are able to adopt the new technology faster can potentially increase their valuation. The underlying intuition is that market participants may believe that banks with higher exposure to digital technology adoption might have a more sustainable future business outlook than peer banks as technology continues to advance. As a result, these banks may attract more investment from institutional investors, who value their digital exposure and are willing to invest because of brighter business prospects. This can lead to a higher price-to-book ratio relative to peer competitors.

Previous research, such as Chen and Srinivasan (2023) and Fritzsch et al. (2021), find that digital activities among non-tech firms and insurance firms is associated with a higher price-to-book ratio than their industry peers. Similarly, Kueschnig and Schertler (2023) find that traditional financial institutions announcing new fintech merger and acquisition deals experience higher abnormal equity returns compared with non-fintech deals. However, there is limited empirical research so far on the impact of digital technology exposure on banks' price-to-book ratios and institutional investor holdings. Our study aims to fill this gap in the literature. Details of the empirical specifications and relevant data sources for these research questions will be discussed in later sections.

3. Data and description of textual analysis algorithm

In this section, we provide a brief overview on our datasets, including financial data and unstructured textual data, as well as the topic modelling algorithm (Top2vec) used to construct the text-based measure of banks' digitalisation adoption. Further detail on the topic modelling algorithm is given in the Appendix.

3.1 Data descriptions

3.1.1 Financial data of banks

Several types of bank-specific information are used in this study. Various bank accounting and market-based data, including return on asset, capital ratios, non-performing loan ratios, and price-to-book ratios, are obtained from S&P Capital IQ Pro. Institutional investor holding-level data for our sample of publicly listed banks are obtained from S&P Capital IQ, which will be used to assess whether and how far institutional investors' holdings of these banks' shares is affected by the degree of banks' digitalisation.[7]

Banks' key development events are also collected, which will be used to examine the usefulness of our text-based measure in gauging banks' digitalisation adoption. Such information is obtained from S&P Capital IQ, and these events are categorised as either "products-related announcement" or "business expansion". In total, 5019 key development events are obtained for our sampled banks (including events related to their subsidiaries) over the period between 2012 and 2022. To further segment those that are related to tech events (denoted as *TechEvent*), we manually review the headlines of these events to identify those that have tech-relevant wordings.[8] Among these 5019 events, we identify 663 *TechEvents* that are directly related to banks' technological adoption: these include events as either announcing a launch of new digital products or expanding new business segments with a technological focus. The

[7] The availability of granular details at banks' individual institutional investors level, such as the percentage of common share outstanding reported at annual frequency held by individual investors, the geographical locations of these institutional investors, and their investment styles and orientations, etc, are conducive to more robust empirical analysis in Section 4.3

[8] Such as "digit", "fintech", "cloud", "blockchain", "App", "Mobile", "virtual", "innova", "crypto", "payment solution", "contactless", etc.

extra information on key development events facilitates an assessment of the usefulness of our text-based indicator in gauging banks' digitalisation exposures which we do in Section 4.1.[9]

3.1.2 *Banks' earnings call transcripts*

As mentioned, we mainly draw on banks' earnings call transcripts to construct a bank-level measure of digitalisation adoption. These transcripts are sourced from S&P Capital IQ database, capturing details such as dates, company names, speakers, and call content for publicly listed banks between 2010 and 2022.[10] Standard pre-processing and data cleansing procedures are applied. Following the literature (for instance, Li et al., 2021a; Li et al., 2021b), we segment each transcript into multiple paragraphs by different speakers, and each of these paragraphs is considered as a separate "document" for our topic modelling algorithm. A total of 1,001,739 paragraphs from 11,070 transcripts of 361 banks are used to 'train' our topic modelling algorithm.

3.2 Key steps in constructing a text-based measure of banks' digitalisation adoption

3.2.1 *Identification of fintech/digital-related topics in banks' earning call transcript based on Top2Vec algorithm*

In this study, we use an unsupervised clustering-based topic modelling algorithm, called Top2Vec, to identify fintech-related topics from banks' earnings call transcripts which are then used to generate a bank-time-specific exposure variable for banks' fintech adoptions. Top2Vec model is a clustering-based model[11] proposed by Angelov (2020) that extends the framework of the widely used Word2Vec and

---

[9] Selected illustrative examples of identified "TechEvents" are provided in Table A2.

[10] Specifically, the raw transcripts are available in pdf format and therefore we follow the approach adopted by Li et al. (2021a) to extract the relevant contents from pdf files into csv and txt formats. The relevant Python codes for extracting contents are available for download in Github by the authors (https://github.com/ssrn3632395/The-Role-of-Corporate-Culture-in-Bad-Times). We express gratitude to the authors of Li et al. (2021a) for their generosity in uploading open source codes online for other's reference.

[11] In a clustering-based model, documents are organised into topics based on their similarity, with the most relevant words for each cluster serving as the topic's representative elements (Xie and Xing, 2013).

Doc2Vec algorithms (Mikolov et al., 2013; Le and Mikolov, 2014).[12] This model employs word and document embeddings, that are learned vector representations of their meanings (Mikolov et al., 2013; Le & Mikolov, 2014), to cluster semantically and contextually similar words. The centroid of the document embeddings in every cluster, which is denoted as "topic vector", thus represents a topic. Specific key words with the most similar embeddings to a topic vector are therefore associated with that topic. This functionality conveniently enables relevant textual analytical tasks in this study, such as classifying a paragraph from the transcripts to its most related topic vector associated with a particular keyword concept (i.e. "Digital", "Fintech").[13]

To identify fintech-related topics in banks' earning call transcripts, the following steps are performed. First, the Top2Vec model is trained on a dataset consisting of 1,001,739 paragraphs from 11,070 transcripts from 361 banks.[14] The model identifies 2,622 fine-grained centroids topic vectors. As these topic vectors span a wide range of different topics, we extract a set of topic vectors that is contextually similar to the two fundamental keywords "digital" and "fintech". In our analysis, we initially choose the top 20 topic vectors that are most contextually similar to the two keywords of "digital" and "fintech". After validating the accuracy of these topic vectors by examining the word clouds corresponding to these topic vectors, a total of 15 topic vectors are obtained that are considered to be most relevant to our target topics (i.e. fintech/digital adoption). Panel A and Panel B of Figure 3.1 provide two examples of word clouds for the most relevant topic vector identified by Top2vec algorithm with the keywords "digital" and "Fintech". The application step here is conceptually similar to the approach adopted in Tavakkolnia and Smeulders (2023) in generating fine-grain risk factor clusters with a high level of detailed information.

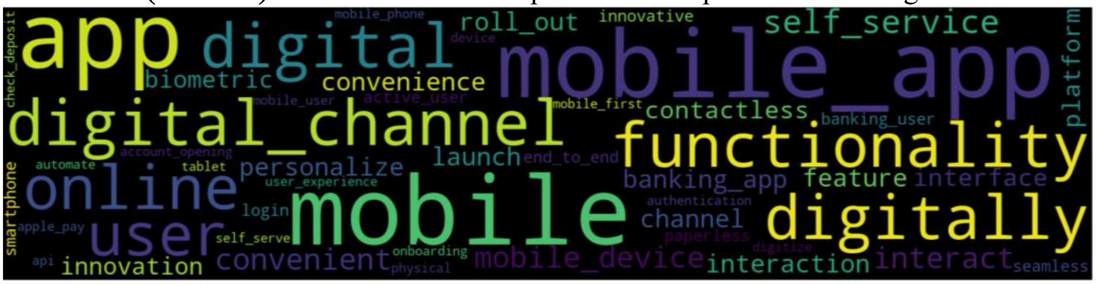
[12] Top2vec offers several advantages over the conventionally adopted Latent Dirichlet allocation (LDA) method (Blei, et al., 2003), including learning the number of topic vectors within the algorithm instead of pre-defining it, capturing context information in paragraphs for each word, and generating more interpretable word topic vectors by considering the similarity between words (Angelov 2020; Dieng et al. 2020).

[13] In light of this enhanced capability, Top2vec algorithm and other alternative clustering-based topic modelling models have gained increased traction in the finance literature (for instance, Tavakkolnia and Smeulders, 2023; Dangl et al. 2023; Alexopoulos et al. 2023).

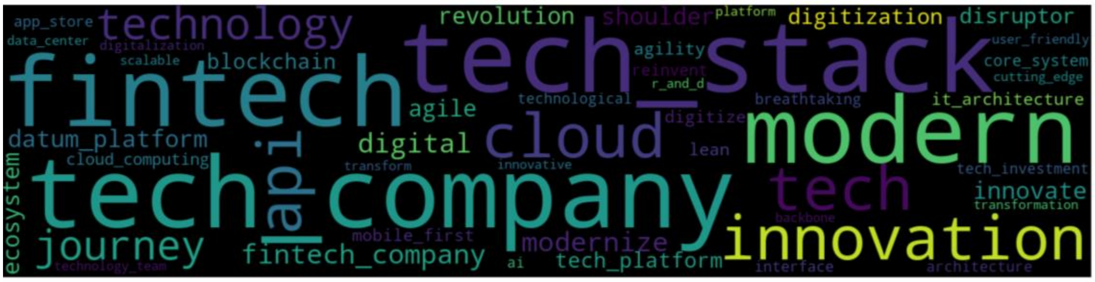
[14] Details of hyperparameters used in the model are presented in the Appendix.

Figure 3.1: Word cloud examples of the learned topic vector identified by Top2vec algorithm

(Panel A) Wordcloud for one representative topic vector for "digital"



(Panel B) Wordcloud for one representative topic vector for "fintech"



These identified topic vectors can be viewed as a "yardstick" to contextually match relevant paragraphs related to digital or fintech-related topics in banks' earning call transcripts. Specifically, if a paragraph j from transcript *i* for event in year t for bank b (denoted as $pagh_{b,i,j,t}$) is determined by the Top2Vec model to have the same topic vector from our 15 identified topic vectors, the paragraph is considered relevant to our targeted concept. An indicator function is then applied to this paragraph which takes a value of one if it is related to our target concept $1(pagh_{b,i,j,t} = Digital\ or\ Fintech)$, and zero otherwise.

3.2.2 Text-based indicator of banks' digitalisation exposures

After identifying fintech-related topics in banks' transcripts, we measure the extent of banks' digitalisation adoption by calculating the proportion of text in the earning call transcripts that is contextually relevant to digitalisation topics, denoted as $DFscore_{b,t}$. Specifically, $DFscore_{b,t}$ is computed by the following formulas (1) and (2) using outputs from Top2vec algorithm:

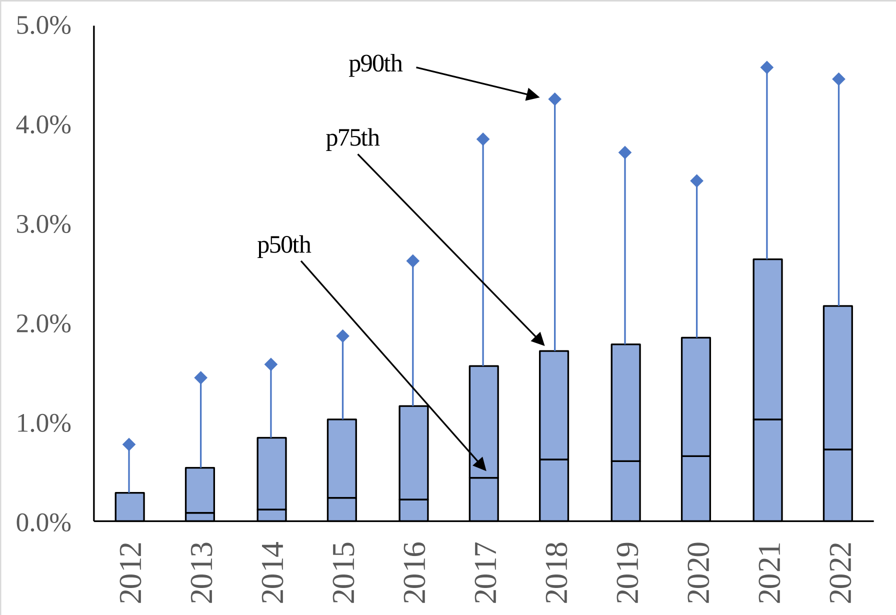
$$DFscore_{b,i,t} = \sum_{j=1}^{Npagh_{b,i,t}} 1(pagh_{b,i,j,t} = Digital\ or\ Fintech) \Big/ Npagh_{b,i,t} \quad (1)$$

$$DFscore_{b,t} = \sum_{i=1}^{Ntran_{b,t}} DFscore_{b,i,t} \Big/ Ntran_{b,t} \quad (2)$$

where $1(pagh_{b,i,j,t} = Digital\ or\ Fintech)$ takes a value of one if the $paragraph_{b,i,j,t}$ is determined by the Top2vec algorithm to be closest to the identified fintech/digital topic vectors, $Npagh_{b,i,t}$ is the number of paragraphs in an earnings call transcript i for bank b in year y. If a bank holds more than one earnings call during a year, we average the DFscore for that bank by the total number of call transcripts available in year t for bank b (i.e. $Ntran_{b,t}$). As a result, the average annual bank-specific digitalisation adoption measure based on (2) includes data points for around 3,300 bank-year observations.

Figure 3.2 displays the distribution of DFscore for our sampled banks, using the mean, 50^{th}, 75^{th} and 90^{th} percentiles levels. It is observed that while digitalisation-related topics do not yet contribute to mainstream discussion in banks' earning calls, they have gained increased traction in recent years. As suggested by Figure 3.2, there was prominent focus in 2021 alongside the tech boom in the stock market during the same year. The heterogeneity across banks is notable in that it may suggest some banks are lagging behind in technological adoption progress.

Figure 3.2: Distribution of DFscore across banks 2012-2022



Note: The p50th, p75th and p90th values represents the median, 75th, and 90th percentiles value of the distribution of DFscore in each calendar year.

3.3 Sample coverage and descriptive statistics

Regarding our sample, we include all listed banks that have a market capitalisation of at least US$ 1 billion and at least 5 analysts' coverage as of end-2022. [15] These screening criteria leave us with more than 500 banks[16]. Among them, 361 banks have their earnings transcripts available for download in the S&P Capital IQ database. The earning transcripts of these 361 banks are used to train the Top2Vec topic model. For estimation analysis later on, we restrict our sample to those banks with available transcripts for at least 5 consecutive years. Our final sample covers 219 publicly listed commercial banks globally.

Around 40% of our sampled banks are located in North America, and 30% and 20% are located in the European and Asia Pacific region respectively. As the S&P Capital IQ database collects earnings call transcripts for companies outside North America comprehensively since 2012, we restrict our bank-year level data sample for analysis to the period between 2013 and 2022. Our sample consists of around 2,000

[15] It is because having some analyst coverages on banks is associated with a higher likelihood for earnings call transcripts to be available in S&P capital IQ database.

[16] As consecutive earnings conference call transcripts are more readily available among traditional commercial banks, we retain in our sample only those commercial banks of which their primary standard industry codes are either "commercial banks" of "financial service".

unbalanced bank-year observations for the empirical analysis. Table A1 lists the sources and descriptions of our variables employed in the empirical analysis while Table 3.1 presents the summary statistics for the variables in our sample at the bank-year level.

Table 3.1: Summary Statistic Table

Panel A (bank-year-level)								
Variable names	N	mean	sd	p10	p25	p50	p75	p90
Dfscore (all)	2157	0.013	0.023	0.000	0.000	0.005	0.016	0.036
Dfscore (insample)	1891	0.011	0.014	0.000	0.000	0.005	0.016	0.034
No. of TechEvents (Ntech)	1898	0.27	0.74	0	0	0	0	1
Dum(Ntech>=1)	1898	0.16	0.37	0	0	0	0	1
Size (ln(asset in HK$))	1891	27.3	1.7	25.0	25.9	27.1	28.6	29.7
Capital Ratio (ppt)	1885	15.92	3.22	12.57	13.55	15.07	17.57	20.70
NPL (ppt)	1846	2.47	2.71	0.36	0.62	1.35	3.23	6.32
Cost efficiency	1871	0.98	0.69	0.54	0.64	0.74	0.97	1.73
ROA (ppt)	1887	0.95	0.64	0.23	0.58	0.94	1.27	1.67
LTD	1879	0.91	0.21	0.66	0.77	0.90	1.02	1.21
Loanast	1891	0.60	0.13	0.41	0.53	0.62	0.70	0.76
Secur_ast	1891	0.23	0.11	0.11	0.15	0.21	0.29	0.38
PBR	1862	1.31	0.70	0.52	0.84	1.22	1.61	2.13
ROE (ppt)	1886	9.73	5.82	3.61	7.02	9.64	12.85	16.54
sh_instinv (ppt)	1839	50.72	24.09	16.04	27.55	50.64	75.10	82.20
sh_ACT_instinv (ppt)	1862	36.28	17.19	9.75	20.63	37.70	52.02	58.91
12m return VOL	1860	0.29	0.14	0.15	0.20	0.25	0.35	0.50
12m return	1854	0.07	0.27	-0.26	-0.13	0.04	0.24	0.44
Panel B (bank-investor-year level)								
Variable names	N	mean	sd	p10	p25	p50	p75	p90
Inst_int_own_share b,i,t (1 unit = 1 ppt holding of common share)	665559	0.10	0.41	0.0000	0.0004	0.0041	0.0316	0.1677
Dummy_EM banks	667460	0.18	0.39	0	0	0	0	1
Dummy_Foreign investors	667460	0.53	0.50	0	0	1	1	1

Note: “ppt” stands for percentage point. Values for the winsorized variables are reported here.

4. Empirical analysis

4.1 How useful is our textual-based measure (DFscore) derived from banks' earnings call transcripts in gauging digitalisation adoptions by banks?

Before using our constructed text-based DFscore for empirical analyses, we investigate whether our text-based indicator can reasonably gauge the digitalisation exposure of individual banks. We use two empirical models to examine whether the level of DFscore is positively correlated with technological adoptions by individual banks' in the same year (as proxied by the "*TechEvent*" identified from the Key development database of S&P Capital IQ, see section 3.1.1 for details).

Specifically, a probit and a poisson regression (i.e. equations (3) and (4)) are used:

$$Pr(No.of\ TechEvent_{b,t} \geq 1|X) = \Phi(X\beta) \tag{3}$$

$$Pr(No.of\ TechEvent_{b,t} = h_i|X)_{b,t} = {e^{-\exp\{X\beta\}}\exp\{X\beta\}^{h_i}}\Big/{h_i!}\ , \tag{4}$$

with $h_i = \{0,1,2,3\ldots\}$

where Φ is the standard cumulative Gaussian distribution function and $X\beta = \beta_0 + \beta_1 DFscore_{b,t} + \gamma_1 bankctrl_{b.t-1} + \vartheta_1 macroctrl_{c.t-1} + FEs$. In this specification, the subscript indices *b*, *c*, *t* represents bank *b* located in country *c* and year *t* dimensions respectively. For the dependent variables in (3) and (4), $Pr(No.of\ TechEvent_{b,t} \geq 1|X)$ is the probability of bank b having at least one identified *TechEvent* in year t, while $Pr(No.of\ TechEvent_{b,t} = h_i|X)_{b,t}$ is the probability of bank *b* having exactly h_i *TechEvent* in year *t*. $bankctrl_{b.t-1}$ denotes the vector of bank-specific control variables, which includes the log of bank assets (*SIZE*), return on assets (*ROA*), non-performing loan ratio (*NPL*), price-to-book ratio (*PBR*), loan-to-deposit ratio (*LTD*) and Capital ratio (*CapRatio*). $macroctrl_{c.t-1}$ is the vector of country macro control variables which includes real GDP growth, and the inflation rate and the short term interbank interest rate. These control variables are lagged by one year. FEs represents the vector of country-level and year-level time fixed effects.

Our parameter of interest is β_1, which estimates the partial effect of a higher DFscore on the likelihood of a *TechEvent* as described in specification (3), and the partial effect of a higher DFscore on the expected number of *TechEvents* in specification (4), in the same calendar year. If DFscore can adequately capture the extent of banks' digitalisation, we expect it to be associated with digital transformative action or events undertaken by banks. In which case, we expect a positive coefficient on β_1.

Table 4.1 reports the estimated results for equations (3) and (4) respectively in column (1) and (2). The coefficient of β_1 in both specifications (3) and (4) is indeed positive and statistically significant, which suggest that DFscore is a useful proxy of the extent of digitalisation adoption by banks. In terms of size, the results show that a 3.5-ppt increase in DFscore (that is, the mid-point of 90^{th} percentile of DFscore relative to the median level) is associated with an increase in the probability of a TechEvent of close to 7% in specification (3), while the results in specification (4) suggest that the expected number of TechEvents increases by (8.34)*(0.035)/100%=29% in response to a 3.5-ppt increase in the DFscore.[17]

To further strengthen the validality of our DFscore variable, we regress banks' DFscore on firms' information technology, equipment expenditure to total non-interest operating expenditure ratios for a subset of US sample banks that have reported these figures and are included in the S&P Capital IQ database. The intuition follows Modi et al. (2022) and Presbitero et al. (2024) in examining the determinants of banks' IT spending and its usages in research. We focus on the subsample of US banks because their data disclosures in IT spending is more consistently reported in the S&P Capital IQ database (i.e. mainly sourced from Call Report). Specifically, Modi et al. (2022) find that banks with more Fintech exposure tend to spend more on IT, based on Call Report data for a sample of US banks. The regression specification is similar to the ones used in the research above, and together with the associated results is reported in column (3) of Table A4 in Appendix III. As expected, we find a positive and statistically significant β_1 coefficient on DFscore among our subsample of US banks,

[17] To quantify, we also estimated a fixed effects linear regression $No.of\ TechEvent_{b,t} = X\bar{\beta}$ and reported the results in Column (3) of Table 4.1. Based on results in Column (3), it is estimated that a 3.5-ppt higher DFscore will increase the expected number of TechEvents by 0.12 unit. Given that the mean and SD of number of $\text{TechEvents}_{b,t}$ are 0.27, the positive impact compared with the mean level is also economically sizable (around 44%).

which is consistent with Modi et al. (2022). This also indicates that banks with a higher DFscore tend to spend more on IT, which serves as an important backbone in greater digital technology adoption by banks. [18]

Taken together, these results suggest that banks with a higher DFscore are more likely to experience a higher number of TechEvents in a given year compared with their peers. This suggests that DFscore is a useful proxy for banks' digitalisation. It supports using this text-based indicator to analyse the potential impact of banks' digitalisation on their financial performance in the following sections.

4.2 Is the degree of a bank's digitalisation associated with better performance?

In this subsection, we investigate whether a higher degree of digitalisation leads to better business fundamentals for individual banks. To do this, we run the following regression across five different dimensions: capital adequacy, asset quality, management efficiency, earnings, and liquidity (CAMEL):

$$CAMEL_{b,t} = \beta_0 + \beta_1 DFscore_{b,t} + bkctrl_{b,\,t-1} + macro\ ctrl_{c,\,t-1} + FEs + \varepsilon_t \quad (5)$$

where the subscript indices b, c, t denote the same dimensions as for equations (3) and (4). $CAMEL_{b,t}$ represent the fiscal year-end position of a bank's financial performance, and includes its total capital ratio (*CapRatio*), non-performing loan ratio (*NPL*), cost-to-income ratio (*CIR*), return on assets (*ROA*) and loan-to-deposit ratio (*LTD*).[19] A similar set of bank and macro control variables as described in the previous section is used. The definition and source of these variables are described in Table A1. β_1

[18] Other than the additional regression exercise, we also examine the correlation between banks' average DFscore between 2019 and 2022 with the November 2023 Evident AI index for a subset of 41 largest commercial banks in North America, Europe, and Asia. The Evident AI Index in essence assesses various approaches that banks are taking towards AI readiness. It is expected to be positively correlated with the DFscore if relevant. It is found that the average DFscore are positively correlated with the logarithm of Evident AI's AI readiness score, as well as the sub-category in talent score and innovation score. These scatter plot findings are available upon request.

[19] Since the fiscal year-end position of a bank's performance is generally disclosed few months after the fiscal year end, which is later than dates of all earnings call meeting that are held in that calendar year. Thus, the potential issue of contemporaneous bias in equation (5) should be limited.

in (5) now reflects the partial effect of higher DFscore on five different bank performance CAMEL indicators respectively.[20]

The results from equation (5) for the five selected $CAMEL_{b,t}$ indicators are reported in Table 3.2. Consistent with economic intuition and the literature, we find that banks with a higher DFscore tend to exhibit better performance in terms of a higher capital ratio and a lower *NPL*, *CIR* and *LTD*, relative to other banks (columns 1 – 3 and 4 respectively). The expected signs are consistent with those reported in the literature, and suggest that greater digitalisation adoption can improve banks' financial performance through better credit monitoring and a more efficient allocation of resources. While we find a positive coefficient for *ROA* in Column (4), it is not statistically significant, but is consistent with the mixed findings in the literature so far.[21]

To quantify these results, we compare the estimated effect of a hypothetical bank with a $DFscore_{b,t}$ of 3.5-ppt and the estimated increase in CAMELb,t relative to a bank with zero $DFscore_{b,t}$. It is estimated that a 3.5-ppt increase in DFscore is associated with an increase of 0.27ppts in a bank's total capital ratio and a reduction in their NPL, CIR and LTD of 0.24 ppts, 8.4 ppts and 1.4 ppts respectively. By comparing the effects with the mean level reported in the summary statistic Table 3.1, these estimates represent economically significant effects. Given that the longer term beneficial effects may not yet be fully reflected in these estimates with the ongoing technological adoption progress, the estimates may only reflect a fraction of the fully advantages from technological adoptions to be realised.

One limitation of the above analysis is that these empirical findings are associative, and thus may be subject to potential endogeneity issues. For instance, one might argue that the positive relationship between DFscore and bank performance indicators may be because better performing firms tend to spend more on IT. To alleviate such endogeneity concern, we follow Chen and Srinivasan (2023) and conduct a two-year period lead-lag regression between DFscore and banking performance

[20] REGHDFE command in STATA by Correia (2016) is employed to estimate the multiple dimension fixed effects linear regressions in (5).
[21] For instance, Zhao et al. (2022) found banks' patent applications have no effect on bank profitability based on a sample of Chinese banks. By contrast, Beltrame et al. (2022) and Bian et al. (2023) found evidences for a positive effect from higher patent filings on banks' ROAs using another Chinese banks sample.

indicators to mitigate potential contemporaneous effects between the two variables. Lead-lag regression analysis can help address endogeneity concerns as it is in general difficult to forecast future fundamental performances. If banks merely intend to promote more digital adoption after recording better performance, any positive effects on their fundamentals should dissipate in future periods (i.e. the effect of lagged DFscores on CAMEL indicators will likely vanish). At the same time, better performing banks will likely spend more time discussing digital adoption issues in future earnings call events (i.e. a positive effect of lagged CAMEL on future DFscore). The related specifications and associated estimation results are provided in Table A5 and A6 in Appendix IV. Overall, we still find statistically significant and same sign coefficients if we employ DFscore(t-2) as our targeted independent variables for capital, asset quality, cost efficiency in year t as shown in Table A5. However, we do not find any statistically significant results from regressing lagged period values of each of the CAMEL(t-2) on DFscore(t) in Table A6. Therefore there is no strong evidence supporting the idea that better performing banks (in previous two-year period) spend more time discussing digital technology related issues in earnings call events. In summary, this study provides fresh empirical findings that lend support to a positive impact of digitalisation on banks' performance. As digitalisation progresses, we can expect to see even more beneficial impacts in the future.

4.3 Is higher exposure to technological adoption associated with better perceived profitability prospects?

As digitalisation and artificial intelligence continue to advance, they have the potential to significantly enhance banks' productivity and profitability. This could lead to higher bank valuations, as banks with higher exposure to digital technology adoption may have a more sustainable business outlook than their peers. In this section, we extend our analysis to examine whether a bank's exposure to digitalisation adoption is positively associated with stronger profitability prospects as perceived by market participants. To measure this, we use banks' price-to-book ratio (PBR), which is the ratio of the market value of a bank's equity to its accounting value. The PBR measures investors' expectations of how much shareholder value the bank will create from a given stock of assets and liabilities. As such, the PBR is commonly used in the literature as an indicator of a bank's resilience and business prospects (Bogdanova et al., 2018; Kerry, 2019; Gambacorta et al., 2020; Simoens and Vennet, 2021; Caparusso et al., 2023, etc.).

To assess the impact of DFscore on banks' PBR, we regress banks' PBR on DFscores and other relevant bank control variables. Specification (6) considers the average effect over the whole sample period while specification (7) distinguishes the effect from the end of 2020 and onwards. Specification (7) takes into account the increasing interest in tech and AI-related concepts in financial markets since 2020, due to a wider public adoption of AI. This period captures the enthusiasm in the financial market for digital technological advancements, which has driven the share prices of tech leaders to historical high levels.[22]

$$PBR_{b,t} = \beta_0 + \beta_1 DFscores_{b,t} + \delta_1 bk\ controls_{t\,/\,t-1} + bank\ FE + EXCHctry\#year\ FE + \varepsilon_{b,t} \tag{6}$$

$$PBR_{b,t} = \beta_0 + \beta_1 DFscores_{b,t} + \beta_2 DFscores_{b,t} * dum(after2020) + \delta_1 bk\ controls_{t\,/\,t-1} + bank\ FE + EXCHctry\#year\ FE + \varepsilon_{b,t} \tag{7}$$

where $PBR_{b,t}$ is the average price-to-book ratio between the last month (i.e. December) of calendar year t and the first month (i.e. January) of calendar year t+1, so that all information contents entailed in $DFscores_{b,t}$ are already available ahead of the period we observe banks' PBR. We include key relevant bank characteristics control variables commonly employed in the literature (such as Bogdanova et al. 2018; Gambacorta et al., 2020; Simoens and Vennet, 2021; Caparusso et al. 2023; Mücke, 2023) that are important determinants of banks' PBR, including ROE, NPL, bank size, capital ratio, 12-month share prices return and 12-month return volatility. As noted in Simoens and Vennet (2021), it is also important to capture the effect of general developments in the stock markets in the regions where the banks are listed. However, it is difficult to

[22] The periods from 2020 onwards also cover the COVID-19 pandemic in which firms have increasingly adopted digital technologies to cope with the challenging operating during lockdowns. Abidi et al. (2022) and Xia et al. (2022) showed that more digitalised and technologically adopted firms can better cope with the difficult operating environment during the pandemic.

identify all of the stock market indices that are relevant for our global sample, and so we replace the year fixed effect with the country of stock exchange*year fixed effect to approximate stock market developments.

Table 4.3 presents the results of the specifications (6) and (7). While there are no statistically significant effects of a higher DFscore on PBR for the whole sample period as reflected in the coefficient of β_1 in column (1), there is evidence to suggest that a higher DFscore increases banks' PBR after 2020 as shown by the statistically significant β_2 in Column (2). This may reflect the idea that digitalisation and AI-related developments have gained greater investors' attention in more recent years. In terms of the magnitude of the effect, it is estimated that a 3.5-ppt increase in the DFscore is associated with a 0.08 unit increase in PBR, which is equivalent to 6.1% higher PBR than the mean level (1.31 unit) for the whole sample. This suggests that increased technological adoption by banks is associated with more positive business prospects as perceived by market participants.[23]

4.3.1 Identification of effects through cross-sectional differences among banks' institutional investor ownership

To strengthen our identification on the effect of DFscore, we further exploit cross-sectional differences in institutional investor shareholdings for our sampled banks. While the information content in the DFscore is valuable in capturing the relative extent of banks' digitalisation exposure, such non-financial information is more challenging to be extracted and obtained by less sophisticated retail investors. In contrast, institutional investors, especially those active institutional investors, tend to explore alternative information alongside publicly available financial statements to identify investment opportunities. It is typically recognised that professional investors use a wider variety of information, including non-financial information, in their investment decisions than non-professional investors (Sutton et al., 2010; Cohen, Holder-Webb and Zamora, 2015; Bird and Karolyi, 2016; Ilhan, et al. 2023). As such, institutional

[23] The statistically significant and negative effect of DFscore on PBR prior 2020 disappeared if we drop bank fixed effect and keep only time varying bank-specific control variables, yet the stronger effect after 2020 remained robust in this alternative specification. For conservativeness, we keep the specifications including bank fixed effects but limit our discussions on the estimated effects prior 2020.

investors can be expected to pay more attention to information embedded in the earnings call transcripts, and hence to be more responsive to changes in banks' DFscore.

Given this differences in the sensitivity to banks' DFscore between retail and institutional investors, the information content captured in DFscores should be more revealing for banks with higher institutional investor ownership. If greater digitalisation leads to higher future profitability as perceived by investors, positive effects on PBR should be more pronounced for banks with higher institutional investor ownership relative to other banks, conditional on the same level of change in DFscore.

To empirically examine this, we modify (6) by adding an interaction term between DFscore and the share of institutional investors (denoted as $sh\_instINV_{b,t}$) to test whether the size of the effect on PBRs is dependent on the share of institutional investor ownership. Column 3 in Table 4.3 reports the results of the modified regressions using $sh\_instINV_{b,t}$ as the interaction term. Overall, our estimation results are consistent with our conjecture. Conditional on the level of DFscore, banks with higher institutional investor ownership are found to experience a higher level of PBR compared to their counterparts.[24]

Taken together, our findings suggest that greater digitalisation benefits banks' current bank performance and fundamentals, and also leads to an improvement in their future business prospects. This finding is relevant for bank business strategies in view of the compression of price-to-book ratios in the banking industry globally.

4.3.2 Robustness analysis

The empirical analyses so far has primarily focused on banks by examining their fundamental performances and market valuation impact from greater digitalisation. Our textual measures can also be used to empirically examine whether institutional investors ownership is influenced by the digital technological exposure of banks, given that our sample covers global listed banks that have detailed shareholding ownership

---

[24] As the investment behaviour may differ between active and passive institutional investors, with the latter being driven by mainly by market indexes given their investment mandate to track targeted indexes. Thus, passive institutional investors may not necessarily pay attentions to banks' non-financial information, as suggested by Sakaki and Jory (2019). To address this, we have conducted a robustness check by replacing the share of institutional investor ownership with that of active institutional investor ownership in specification (6). Result is quantitatively similar to table 4.4 in panel B of Table A7.

information available over time. The existing literature suggests that there is a positive relationship between the degree of digitalisation and and institutional shareholder ownerships of listed firms (for instance, Aghion, Van Reenen and Zingales, 2013; Sakaki and Jory, 2019, etc). More digitalisation at a firm level can benefit shareholders through a lower cost of equity (Zhang and Wang, 2024). It is unclear whether a similar pattern holds in the banking industry because of scant research in this area.

In Appendix V, we provide empirical evidence of the relationship between our constructed measure DFscore and institutional investor ownership for our sample of large banks. Based on bank-year-level regressions in Table A7 and bank-investor-year level regressions in Table A8, we find a positive relationship between DFscores and the share of active institutional ownership of banks. Furthermore, we find evidence to suggest that a higher DFscore could alleviate the home bias of foreign institutional investors that deters investment in banks in emerging economies (Table A8). Such bias can arise from information asymmetries between foreign institutional investors and banks in emerging economies. This provides further support for the view that a higher degree of digital technological adoption among banks, as reflected in a higher DFscore, can be a source of valuable extra information for active institutional investors to enrich their investment holding decisions. These empirical findings support our adoption of banks' institutional ownership structure to identify the effects of a higher DFscore on banks' PBRs. The specifications, estimation results and related discussion for this additional content are given in Appendix V in a more detailed manner.

5. Conclusions

In conclusion, this study utilises a state-of-the-art NLP technique to gauge the degree of banks' digitalisation adoption based on their earnings call transcripts. It uses this indicator to examine the potential impact of digitalisation on banks' fundamentals. Based on a large sample of listed banks globally, this analysis provides fresh insights that help to broaden our understanding of the trend in greater digitalisation of banks' business models and its implications for the financial performance of the banking sector.

First, our textual measure supports the view that there has been a rapid pace of digitalisation by banks in recent years. However, there is significant heterogeneity in the degree of digitalisation by among banks globally. This suggests that some financial institutions could be at risk of lagging behind their peers in terms of their digitalisation adoption, which could undermine their long-term competitiveness. Second, our analysis provides fresh empirical evidence to support the view that banks with greater digitalisation benefit from improved cost efficiency, better asset quality, and a stronger capital and liquidity positions relative to their peers. This is consistent with the view that greater digitalisation can help to improve banks' performance and fundamentals. In addition, it can enhance their future profitability as perceived by market investors. Specifically, banks with greater digitalisation adoption have brighter business prospects as perceived by investors. An important policy implication is that it is crucial for banks to continue embracing and enhancing their digital transformation, so that they do not lag behind in the digitalisation trend which could hinder their long-term competitiveness.

References

Abidi, N., El Herradi, M., & Sakha, S. (2022). Digitalization and resilience: firm-level evidence during the COVID-19 pandemic. *IMF Working Paper*, No. 2022/034, International Monetary Fund.

Aghion, P., Van Reenen, J., & Zingales, L. (2013). Innovation and institutional ownership. *American economic review*, *103*(1), 277-304.

Alexopoulos, M., Han, X., Kryvtsov, O., & Zhang, X. (2023). More than words: Fed Chairs' communication during congressional testimonies. *Journal of Monetary Economics*, (forthcoming). https://doi.org/10.1016/j.jmoneco.2023.09.002

Angelov, D. (2020). Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*.

Beltrame, F., Zorzi, G., & Luca, G. (2022). The effect of FinTech investments on listed banks: Evidence from an Italian sample. *Risk Governance & Control: Financial Markets & Institutions*, 12(2), 47-55.

Bian, W., Wang, S., & Xie, X. (2023). How valuable is FinTech adoption for traditional banks?. *European Financial Management*. (forthcoming), https://doi.org/10.1111/eufm.12424.

Bird, A., & Karolyi, S. A. (2016). Do institutional investors demand public disclosure?. *The Review of Financial Studies*, 29(12), 3245-3277.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993-1022.

Bogdanova, B., Fender, I., & Takáts, E. (2018). The ABCs of bank PBRs. *BIS Quarterly Review*, March, 2018, Bank for International Settlements.

Caparusso, J., Lewrick, U., & Tarashev, N. (2023). *Profitability, valuation and resilience of global banks-a tight link BIS Working Paper*, No. 1144, Bank for International Settlements.

Chen, W., & Srinivasan, S. (2023). Going digital: Implications for firm value and performance. *Review of Accounting Studies*, 1-47.

Chen, M. A., Wu, Q., & Yang, B. (2019). How valuable is FinTech innovation?. *The Review of Financial Studies*, *32*(5), 2062-2106.

Cheng, M., & Qu, Y. (2020). Does bank FinTech reduce credit risk? Evidence from China. *Pacific-Basin Finance Journal*, 63, 101398.

Cohen, J. R., Holder-Webb, L., & Zamora, V. L. (2015). Nonfinancial information preferences of professional investors. *Behavioral Research in Accounting*, 27(2), 127-153.

Correia, S. (2016). A feasible estimator for linear models with multi-way fixed effects. *Preprint*, http://scorreia.com/research/hdfe.pdf.

Dangl, T., Halling, M., & Salbrechter, S. (2023). Firm-specific Climate Risk Estimated from Public News. Available at SSRN: https://ssrn.com/abstract=4575999

Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, *8*, 439-453.

Fang, L., Li, X., Subrahmanyam, A., & Zhang, K. (2023). Does FinTech Innovation Improve Traditional Banks' Efficiency and Risk Measures? A New Methodology and New Machine-Learning-Based Evidence from Patent Filings. Available at SSRN: https://ssrn.com/abstract=4350734.

Fang, Y., Wang, Q., Wang, F., & Zhao, Y. (2022). Bank fintech, liquidity creation, and risk-taking: Evidence from China. *Economic Modelling*, 127, 106445.

Fritzsch, S., Scharner, P., & Weiß, G. (2021). Estimating the relation between digitalization and the market value of insurers. *Journal of Risk and Insurance*, *88*(3), 529-567.

Gambacorta, L., Oliviero, T., & Shin, H. S. (2020). Low Price-To-Book Ratios and Bank Dividend Payout Policies. *BIS Working Paper*, No. 907, Bank for International Settlements

Gosselin, M. A., & Taskin, T. (2023). What Can Earnings Calls Tell Us About the Output Gap and Inflation in Canada?, *Discussion Papers*, 2023-13, Bank of Canada.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Guo, P., & Zhang, C. (2023). The impact of bank FinTech on liquidity creation: Evidence from China. *Research in International Business and Finance*, 64, 101858.

Hasan, I., Li, X., & Takalo, T. (2023). Technological innovation and the bank lending channel of monetary policy transmission. *BOFIT Discussion Papers*, No. 9/2023.

Hassan, T. A., Hollander, S., Van Lent, L., & Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, 134(4), 2135-2202.

Heinrichs, A., Park, J., & Soltes, E. F. (2019). Who consumes firm disclosures? Evidence from earnings conference calls. *The Accounting Review*, *94*(3), 205-231.

Hornuf, L., Klus, M. F., Lohwasser, T. S., & Schwienbacher, A. (2021). How do banks interact with fintech startups?. *Small Business Economics*, 57, 1505-1526.

Ilhan, E., Krueger, P., Sautner, Z., & Starks, L. T. (2023). Climate risk disclosure and institutional investors. *The Review of Financial Studies*, *36*(7), 2617-2650.

Kerry, W. (2019). Finding the bad apples in the barrel: using the market value of equity to signal banking sector vulnerabilities. *IMF Working Paper*, No. 2022/034, International Monetary Fund.

Krishnan, A., & Kennedyraj (2023). Exploring the Power of Topic Modeling Techniques in Analyzing Customer Reviews: A Comparative Analysis. *arXiv preprint arXiv:2308.11520*.

Kueschnig, M., & Schertler, A. (2023). Fusing futures: Financial institutions' stock price response to fintech acquisitions. *Finance Research Letters*, 104779.

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196). PMLR.

Leippold, M., & Yu, T. (2024). The Green Innovation Premium: Evidence from US Patents and the Stock Market. *Swiss Finance Institute Research Paper*, (23-21). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4391444.

Li, K., Liu, X., Mai, F., & Zhang, T. (2021a). The role of corporate culture in bad times: Evidence from the COVID-19 pandemic. *Journal of Financial and Quantitative Analysis*, *56*(7), 2545-2583.

Li, K., Mai, F., Shen, R., & Yan, X. (2021b). Measuring corporate culture using machine learning. *The Review of Financial Studies*, *34*(7), 3265-3315.

Li, Q., Shan, H., Tang, Y., & Yao, V. (2024). Corporate climate risk: Measurements and responses. *The Review of Financial Studies*, (forthcoming).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Modi, K., Pierri, N., Timmer, Y., and Peria, M. S. M. (2022). "The anatomy of banks' IT investments: Drivers and implications". *IMF working paper*, No. 2022/244.

Mücke, C. (2023). Bank Dividend Restrictions and Banks' Institutional Investors, *SAFE Working Paper Series*, No. 392, Leibniz Institute for Financial Research SAFE.

Presbitero, A., Rebucci, A., & Zhang, G. (2024). "Bank Technology Adoption and Productivity", SSRN working paper, available at SSRN 4688583.

Qi, Y. D., & Cai, C. W. (2020). Research on the multiple effects of digitalization on the performance of manufacturing enterprises and its mechanism. *Learning and Exploration*, *7*, 108-119.

Sakaki, H., & Jory, S. R. (2019). Institutional investors' ownership stability and firms' innovation. *Journal of Business Research*, *103*, 10-22.

Sautner, Z., Van Lent, L., Vilkov, G., & Zhang, R. (2023). Firm-level climate change exposure. *The Journal of Finance*, *78*(3), 1449-1498.

Simoens, M., & Vander V., R. (2021). Bank performance in Europe and the US: A divergence in market-to-book ratios. *Finance Research Letters*, *40*, 101672.

Soto, P. E. (2021). Breaking the word bank: Measurement and effects of bank level uncertainty. *Journal of Financial Services Research*, *59*, 1-45.

Sutton, S., V. Arnold, J. Bedard, and J. Phillips. (2010). Where do investors prefer to find nonfinancial information? *Journal of Accountancy* (September). Available at: http://www.journalofaccountancy.com/Web/20102682

Tavakkolnia, A., & Smeulders, D. (2023). How the Spread of Risk Information Affects the Informativeness of Firms Textual Risk Disclosures. Available at SSRN: https://ssrn.com/abstract=4474130.

Vandevoort, B., Bex, G. J., Crevecoeur, J., & Neven, F. (2023). Topic modelling and text classification models for applications within EFSA. *EFSA Supporting Publications*, *20*(8).

Wu, F., Hu, H., Lin, H., & Ren, X. (2021). Enterprise digital transformation and capital market performance: Empirical evidence from stock liquidity. *Management World*, *37*(7), 130-144.

Xia, Y., Qiao, Z., & Xie, G. (2022). Corporate resilience to the COVID-19 pandemic: The role of digital finance. *Pacific-Basin Finance Journal*, *74*, 101791.

Xie, X., & Wang, S. (2023). Digital transformation of commercial banks in China: Measurement, progress and impact. *China Economic Quarterly International*, *3*(1), 35-45.

Xie, P., & Xing, E. P. (2013). Integrating document clustering and topic modeling. *arXiv preprint arXiv:1309.6874*.

Zhang, C., & Wang, Y. (2024). Is enterprise digital transformation beneficial to shareholders? Insights from the cost of equity capital. *International Review of Financial Analysis*, *92*, 103104.

Zhang, Y., Ye, S., Liu, J., & Du, L. (2023). Impact of the development of FinTech by commercial banks on bank credit risk. *Finance Research Letters*, 103857.

Zhao, J., Li, X., Yu, C. H., Chen, S., & Lee, C. C. (2022). Riding the FinTech innovation wave: FinTech, patents and bank performance. *Journal of International Money and Finance*, *122*, 102552.

Table 4.1: The impact of DFscores on the likelihood and number of occurrences of TechEvents

Column	(1)	(2)	(3)
Model	Probit reg	Poisson reg	Linear reg
Dependent Variables	Prob(TechEvents>=1)	Prob(TechEvents=h_i)	No. of TechEvents
DFscore $_{b,t}$ (β_1)	**10.092*** **	**8.336***	**3.399** **
	(3.564)	**(4.617)**	**(1.401)**
Size b,t-1	0.460***	0.584***	0.229***
	(0.041)	(0.041)	(0.035)
RoA b,t-1	0.240*	0.174	0.060
	(0.136)	(0.224)	(0.039)
PBR b,t-1	-0.094	-0.141	-0.061
	(0.092)	(0.137)	(0.038)
LTD b,t-1	-0.641*	-1.254***	-0.369**
	(0.338)	(0.432)	(0.159)
CapRatio b,t-1	-0.013	-0.021	0.005
	(0.020)	(0.028)	(0.008)
NPL b,t-1	-0.017	-0.030	-0.011
	(0.028)	(0.051)	(0.010)
rGDPg c,t-1	-0.025	-0.051	-0.005
	(0.028)	(0.044)	(0.009)
Infrate c,t-1	0.018	-0.045	-0.004
	(0.047)	(0.072)	(0.011)
ST_int c,t-1	0.048	0.156**	0.023**
	(0.048)	(0.072)	(0.011)
Constant	-13.318***	-17.327***	-5.723***
	(1.428)	(1.769)	(0.992)
No. of obs.	1,703	1,898	1,898
Adj. (Pseduo) R^2	0.26	0.32	0.26
Bank HQ country FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Standard Error type	Clustered-by banks	Robust	Clustered-by banks

Corresponding type of standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4.2: The effect of DFscore on banks' financial performance indicators

Column	(1)	(2)	(3)	(4)	(5)
Performance indicator	Capital	Asset Quality	Cost efficiency	Earnings Capacity	Liquidity
Dependent variables	CapRatio b,t	NPL b,t	Cost_eff b,t	RoA b,t	LTD b,t
DFscore b,t **(β_1)**	**7.588****	**-6.844****	**-2.389****	**0.795**	**-0.403****
	(3.536)	**(3.159)**	**(1.202)**	**(1.001)**	**(0.173)**
Size b,t-1	-0.820**	0.211	0.299***	-0.141**	0.062***
	(0.394)	(0.182)	(0.069)	(0.061)	(0.014)
Loan/Asset b,t-1	-3.046	0.608	0.275	0.139	0.845***
	(2.710)	(1.606)	(0.477)	(0.545)	(0.071)
Secur/Asset b,t-1	1.580	4.271***	-0.207	-0.414	0.200***
	(2.058)	(1.617)	(0.444)	(0.418)	(0.071)
RoA b,t-1	-0.067	-0.490***	-0.087***	n.a.	0.003
	(0.164)	(0.150)	(0.031)		(0.007)
CapRatio b,t-1	n.a.	-0.038	-0.003	-0.017*	-0.001
		(0.029)	(0.010)	(0.010)	(0.002)
NPL b,t-1	-0.052	n.a.	-0.013	-0.019	-0.000
	(0.064)		(0.013)	(0.013)	(0.002)
LTD b,t-1	-0.033	3.225***	-0.124	-0.542*	n.a.
	(1.218)	(0.928)	(0.244)	(0.285)	
rGDPg c,t-1	-0.035	0.007	-0.001	0.008	-0.005***
	(0.023)	(0.017)	(0.007)	(0.007)	(0.001)
Infrate c,t-1	-0.095	-0.019	-0.066***	0.020	-0.001
	(0.068)	(0.069)	(0.025)	(0.021)	(0.002)
ST_int c,t-1	0.134**	-0.027	0.063***	0.012	-0.004
	(0.065)	(0.063)	(0.022)	(0.016)	(0.003)
Constant	39.933***	-7.790	-7.009***	5.536***	-1.281***
	(11.057)	(5.274)	(1.889)	(1.795)	(0.410)
No. of Obs.	1,724	1,660	1,625	1,715	1,706
Within R^2	0.0306	0.156	0.0622	0.0276	0.236
Bank FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Clustered-by-bank standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4.3: The effects of DFscore on banks' price-to-book ratio (PBR)

Column	(1)	(2)	(3)
Dependent variables	PBR b,t	PBR b,t	PBR b,t
DFscore b,t (β_1)	-0.576	-2.251**	-5.674***
	(0.821)	(0.919)	(1.689)
DFscore b,t*dum_aft20 (β_2)		4.458***	
		(1.509)	
(β1 + β2)		2.207*	
		(1.339)	
sh_instinv b,t			0.004*
			(0.002)
DFscore b,t *sh_instinv b,t			0.111***
			(0.032)
12m returnVOL b,t	-0.469***	-0.485***	-0.493***
	(0.116)	(0.118)	(0.120)
12m return b,t	0.582***	0.586***	0.589***
	(0.055)	(0.054)	(0.054)
RoE b,t-1	0.017***	0.017***	0.016***
	(0.003)	(0.003)	(0.003)
Size b,t-1	-0.453***	-0.435***	-0.447***
	(0.088)	(0.089)	(0.090)
NPL b,t-1	-0.069***	-0.068***	-0.070***
	(0.019)	(0.019)	(0.018)
CapRatio b,t-1	-0.019***	-0.018***	-0.018***
	(0.007)	(0.007)	(0.007)
Constant	14.058***	13.569***	13.708***
	(2.430)	(2.440)	(2.451)
No. of obs.	1,665	1,665	1,640
Within R2	0.324	0.332	0.339
Bank FE	Yes	Yes	Yes
ExchangeCtry#Year FE	Yes	Yes	Yes

Clustered-by-bank standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendices

Appendix I – Additional tables

Table A1: Variables description

Variable	Description	Source
Bank-level variables		
DFscore	Banks' yearly digital technological adoption exposure variable constructed from earnings transcripts	Applying Top2vec on transcripts from S&P Capital IQ
Number of TechEvents (Ntech)	No. of digital technology key development events from type "Product-related announcement" and "Business expansion" in a calendar year	S&P Capital IQ - Key development database
Bank size (Size)	Logarithm of total assets in HK dollar millions	S&P Capital IQ
Capital Ratio	Bank's Total Capital Ratio	S&P Capital IQ
Non-performing loan ratio (NPL)	Bank's non-performing loan ratio	S&P Capital IQ
Cost efficiency	Banks' non-interest expense over revenue minus interest expenses	S&P Capital IQ
Return-on-asset (ROA)	Bank's return on assets	S&P Capital IQ
Loan-to-deposit ratio (LTD)	Bank's loan-to-deposit ratio	S&P Capital IQ
loan-to-asset ratio (loanast)	Bank's loan-to-asset ratio	S&P Capital IQ
Securities asset-to-asset ratio	Bank's securities asset-to-asset ratio	
Price-to-book ratio (PBR)	Bank's daily averaged price-to-book ratio between December (Year) and January (Year+1).	S&P Capital IQ
Return-on-equity (ROE)	Bank's return on equity	S&P Capital IQ
12m share price return	Returns on share price of the bank in the past 12 months	S&P Capital IQ
12m return volatility (12m return VOL)	Volatility (standard deviation) of the monthly return on share price of the bank over the past 12 months.	S&P Capital IQ
% sh_instinv	Total percentage of common share holding of a bank by institutional investors with investment orientation (excluding holdings by controlling parent companies). The type of institutional investor follows definition by S&P Capital IQ.	S&P Capital IQ
% sh_ACT_instinv	Total percentage of common share holding of a bank owned by all institutional investors with investment orientation as "Active" (excluding controlling the holding by controlling parent companies). Type of institutional investor follows definition by S&P Capital IQ.	S&P Capital IQ
Country-level variables		
real GDP growth (rGDPg)	Real GDP growth rate of the HQ country of a bank (Supplemented by sources from S&P Capital IQ, CEIC and national statistical bureaus)	World bank- World economic outlook database
Inflation rate (infrate)	Inflation rate of the HQ country of a bank (Supplemented by sources from CEIC and national statistical bureaus)	World bank- A Global Database of Inflation
Short-term interest rates (ST_int)	Short term interbank/money market interest rate of the HQ country of the bank (Supplemented by sources from CEIC and national statistical bureaus)	IMF international financial statistic database
Bank-investor-level variables		
InstInv_own_sha b,h,t	Percentage of common share holding of a bank owned by individual institutional investors with investment orientation as "active", excluding controlling the holding by controlling parent companies).	S&P Capital IQ

DumEM bank	Dummy variable takes value 1 if the headquarter country location of the bank is located in an emerging economy following the definition used by Bank for international settlement in their quarterly reviews.	S&P Capital IQ
DumforINV	Dummy variable takes the value of 1 if the headquarter country location of an institutional investor is different from the headquarter country location of the bank in which it is holding shares.	S&P Capital IQ

Table A2: Illustrative examples of TechEvents identified from S&P Capital IQ Key development database

Date	Company	Event Type	Headline
07/09/2017	ANZ Group Holdings Limited (ASX:ANZ)	Business Expansion	Australia and New Zealand Banking Group Opens New Digital Branch in Wodonga
14/06/2019	Bank of China Limited (SEHK:3988)	Business Expansion	Bank of China Limited Announces Commencement of Operation of BOC Financial Technology Co., Ltd
13/07/2019	China Construction Bank Corporation (SEHK:939)	Business Expansion	China Construction Bank Launches Smart Banks in Beijing
09/11/2018	DBS Bank Ltd.	Product-Related Announcement	DBS to Introduce Mobile-Based QR Payment Collection Solution
08/10/2022	HSBC Holdings plc (LSE:HSBA)	Product-Related Announcement	Hsbc Launches Digital Platform That Revolutionises Trade Finance
30/11/2022	ICICI Bank Limited (NSEI:ICICIBANK)	Product-Related Announcement	Tata Consultancy Services Limited and ICICI Bank Announces the Launch of `iLens', Digital Lending Solution
01/09/2018	Standard Chartered PLC (LSE:STAN)	Business Expansion	Standard Chartered to Launch Digital Bank in Hong Kong; Appoints Deniz Guven as the CEO of the Virtual Bank
13/09/2019	Deutsche Bank Aktiengesellschaft (XTRA:DBK)	Business Expansion	Deutsche Bank Launches New Fintech Innovation Hub in Shanghai
14/04/2018	Banco Santander, S.A. (BME:SAN)	Product-Related Announcement	Santander Introduces New Blockchain-Based International Payments Service

Appendix II – Application of Top2vec algorithm to construct DFscore

In this Appendix II, we provide details of the pre-processing of textual data and the construction of our DFscore employed in this study. We largely follow the working approach in Li et al. (2021a)[25] to extract raw textual data from raw call transcripts in pdf format, removing irrelevant content, and applying pre-processing steps on the textual data before applying the Top2vec algorithm to create relevant topic vector clusters. Figure A1 displays a high level overview of the procedure used to generate our textual variable construction.

In the data cleaning and pre-processing steps, as widely recognized in previous research (Li et al., 2021a; Li et al. 2021b), learning and identifying multiword phrases in the textual corpus is essential, as phrases can express meanings not available from standalone words. In view of the potential presence of universal and sector-specific phrases in our data corpus, we follow Li et al. (2021a) and apply a two-step approaches in tagging both types of phrases in our document corpus. As suggested by Li et al. (2021a), standard natural language processing (nlp) packages, such as Stanford CoreNLP (also known as Stanza now), SpaCy, NLTK, etc, can help to identify commonly used multiword English expressions as they are trained by a huge amount of online textual sources. Whereas sector-specific phrases may only appear in specific textual sources and so are unavailable to the abovementioned general purpose nlp packages, Li et al. (2021a) propose using phraser modules of gensim package to identify these. Therefore, we follow their approach to apply sequentially the stanza package and then the gensim package to identify and concatenate universal and sector-specific multi-word expressions separately.[26] Finally, standard textual processing steps, such as transforming text to lowercase, removing numbers and currency symbols, and very frequent stop-words, are applied to each paragraph before applying the Top2vec algorithm on the text. Although thenTop2vec algorithm can embed text pre-processing and phrase identification steps within its framework, recent papers (Vandevoort et al., 2023; Krishnan and Kennedyraj, 2023; Tavakkolnia and Smeulders, 2023) argue that including common data pre-processing steps and phrase identification steps before

[25] We express gratitude again to the authors of Li et al. (2021a) for uploading their Python codes in Github and making them open source and available to other researchers. (link to Github: https://github.com/ssrn3632395/The-Role-of-Corporate-Culture-in-Bad-Times)

[26] Following the two-step approach, we also remove stop words, punctuation marks and single-letter words before applying the phraser module of the gensim package.

applying Top2vec can improve it performance especially in the presence of short texts. For a comprehensive discussion on potential issues and solutions in constructing a firm-specific variable, such as the corporate culture variables in their application, from textual earnings call transcripts data, we refer readers to Li et al. (2021a, 2021b).

To measure bank-specific exposure to digital technological adoption, we employ the Top2vec algorithm (Angelov, 2020) to learn the topic distributions for paragraphs in the embedding space. Clustering-based topic models, such as Top2vec, BERTopic (Grootendorst, 2022), assume that documents relevant to the same topic have many common semantic and contextual features, and therefore can be represented by vectors close to each other and form a cluster in a high dimensional space. Because our objective is to construct a bank-level variable for measuring banks' digital technological exposure, we aim to pin down those paragraphs relevant to such a concept in a contextual manner. After preparing the pre-processed individual paragraphs from earnings call transcripts, we train the Top2Vec model (Angelov, 2020) using the pre-processed textual data to derive a topic distribution outcome. Table A3 lists the values of key hyper parameters in the Top2vec algorithm for it to derive the results. The resulting model outputs provide detailed high dimensional vector space information for all the input words, documents and topics, and allow us to identify the closest topic vector cluster for each input paragraph. The model outcome can then enable us to follow our variable construction details as described in Section 3.1.

Figure A1: Flow diagram of textual variable construction scheme for our DFscore variable

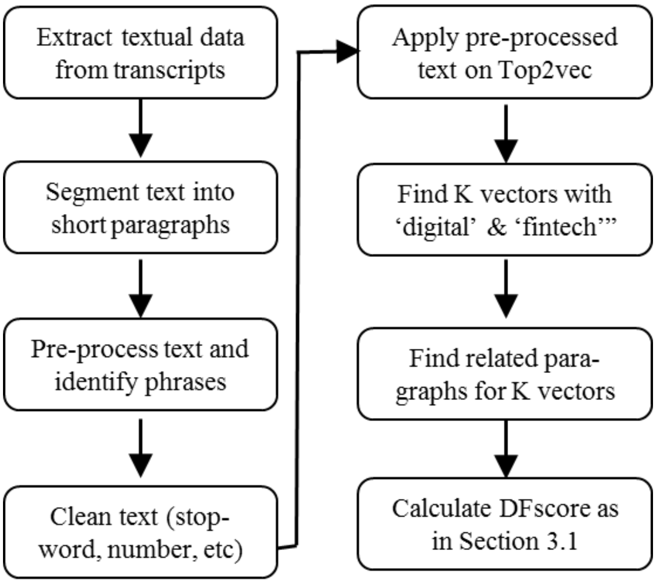


Table A3: Key hyperparameters applied during the Top2vec algorithm

Category	Parameter name	Value
Clustering hyper parameters (UMAP/HDBSCAN)	Minimum cluster size	50
	Minimum sample	5
	Cluster Metric	Euclidean
	Cluster selection method	Excess of mass
	No of neighbours	15
	Token metric	5
	Initial condition	Random
Top2vec parameters	Speed (no. of epoch)	Deep-learn (400)
	minimum occurrence of Token	30
	Embedding model	Doc2vec
	Minimum length of documents	At least 25 characters

Appendix III – Additional regression tables for robustness checks on the validity of DFscore

Table A4: Robustness regressions of the impact of DFscore on the likelihood and number of occurrences of TechEvents, and IT spending ratio based for a subset of US banks

Specification: $ITspending\ ratio_{b,t} = \beta_0 + \beta_1 DFscore_{b,t} + bkctrl_{b,\,t-1} + macro\ ctrl_{c,\,t-1} + FEs + \varepsilon_t$

VARIABLES	(1) TechEvent (dummy)	(2) TechEvent (count)	(3) IT spending ratio
DFscore (t)	**12.860***	**12.140***	**29.072****
	(7.346)	**(6.486)**	**(14.251)**
Size(t-1)	0.471***	0.516***	-1.494***
	(0.047)	(0.045)	(0.566)
ROA(t-1)	0.069	0.162	-0.166
	(0.204)	(0.231)	(0.334)
PBV(t-1)	-0.158	-0.218	0.268
	(0.154)	(0.169)	(0.500)
LTD(t-1)	-0.294	-1.757***	-5.333**
	(0.567)	(0.639)	(2.421)
CapRatio(t-1)	0.021	0.009	-0.051
	(0.040)	(0.047)	(0.092)
NPL(t-1)	-0.038	-0.083	0.484*
	(0.056)	(0.079)	(0.288)
Constant	-12.126***	-11.532***	48.989***
	(1.960)	(1.971)	(15.035)
Observations	712	677	646
regtype	Probit	Poisson	Reghdfe (clustered-S.E.)
Macro_control (t-1)	Yes	Yes	Yes
time_fixed	Yes	Yes	Yes

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix IV – Robustness analyses to address endogeneity issues related to the effect of DFscore on bank performance

Table A5: Regression results for lagged term of DFscore on CAMEL

Specification: $CAMEL_{b,t} = \beta_0 + \beta_1 DFscore_{b,t-2} + bkctrl_{b,\,t-1} + macro\ ctrl_{c,\,t-1} + FEs + \ \varepsilon_t$

Column	(1)	(2)	(3)	(4)	(5)
Performance indicator	Capital	Asset Quality	Cost efficiency	Earnings Capacity	Liquidity
Dependent variables	CapRatio b,t	NPL b,t	Cost_eff b,t	RoA b,t	LTD b,t
DFscore b,t-2 **(β_1)**	7.101*	-5.580*	-2.193*	0.032	0.138
	(3.859)	(3.067)	(1.256)	(1.055)	(0.134)
No. of Obs.	1,282	1,252	1,267	1,280	1,270
Within R^2	0.0314	0.0639	0.0201	0.0163	0.536
Bank_contrl(t-1)	Yes	Yes	Yes	Yes	Yes
Macro_control(t-1)	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Regression results for testing CAMELs as a determinant of DFscore

Specification: $DFscore_{b,t} = \beta_0 + \beta_1 CAMEL_{b,t-2} + bkctrl_{b,\,t-1} + macro\ ctrl_{c,\,t-1} + FEs + \ \varepsilon_t$

VARIABLES	(1) DFscore (t)	(2) DFscore (t)	(3) DFscore (t)	(4) DFscore (t)	(5) DFscore (t)
ROA (t-2)	0.043				
	(0.090)				
CapRatio (t-2)		-0.009			
		(0.022)			
CostEff(t-2)			0.030		
			(0.073)		
NPL(t-2)				-0.029	
				(0.031)	
LTD (t-2)					-0.736
					(0.509)
Constant	-2.275	-1.856	-2.330	-3.507	-1.756
	(5.034)	(5.185)	(5.085)	(5.050)	(4.975)
Observations	1,411	1,411	1,391	1,399	1,404
R-squared	0.634	0.634	0.633	0.639	0.636
Bank Control (t-1)	yes	yes	yes	yes	yes
Macro Control(t-1)	yes	yes	yes	yes	yes
bk_fixed	Yes	Yes	Yes	Yes	Yes
time_fixed	Yes	Yes	Yes	Yes	Yes
Within-adjusted R2	0.0174	0.0174	0.0183	0.0183	0.0181

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. DFscore is scaled by 100 times as dependent variable for easier interpretation.

Appendix V – Further analysis based on bank-investor-year level data

In this supplementary section, we include additional empirical analysis to support the findings on a positive relationship between banks' DFscore and their common shares owned by institutional investors based on both bank-year level and bank-investor-year level regressions. Aghion, Van Reenen and Zingales (2013) examine the relationship between institutional investor ownership and firm innovations. More recently, Sakaki and Jory (2019) find a positive association between stability in equity ownership by institutional investors and their investee firms' level of innovations, and also between active stock-picking institutional investors with the level of firm innovations. Zhang and Wang (2024) find that firms' digital transformation can lower cost of equity capital, confirming its benefits and importance for shareholder decisions. However, whether similar patterns exist in the banking industry remains unclear given limited research in this area. The same intuition could suggest that the information value in greater bank digital technological exposure (as reflected in our DFscore textual measure) could influence the institutional shareholder ownership of listed banks.

We aim to assess whether higher digital technological exposure by banks in terms of DFscore is positively associated with a higher active institutional investor holding share. Based on findings in Aghion, Van Reenen and Zingales (2013), Sakaki and Jory (2019) and Zhang and Wang (2024), we hypothesize that there exists a positive relationship between the investment holdings of banks by institutional investors and their level of digital technological adoption. The impact is expected to be more pronounced after 2020 given a notable increase in enthusiasm in financial markets for digital technological advancements. To empirically test this, we modify (6) and (7) by replacing the dependent variable of banks' PBR with $sh\_Act\_instINV_{b,t}$. Other specifications in (6) and (7) remain the same.

Consistent with the earlier literature, we find that a higher DFscore is associated with a higher share of active institutional investors in Column (1) of Table A7. The effect is more pronounced after 2020 as suggested by the results in Column (2). A 3.5-ppt increase in DFscore implies an increase in the share of active institutional investors of 1.1 ppt (equivalent to a 3.2% increase relative to the mean of 34 ppt for the variable $sh\_Act\_instINV_{b,t}$ in our sample). These results are consistent with our view that higher technological adoption among banks, as reflected in a higher DFscore, can contribute to better bank business prospects. This can lead to a higher PBR value and a higher share of active institutional investor holding relative to peer competitor banks.

We also examine whether the information of higher exposure to digital technological exposure becomes more relevant to foreign institutional investors for their investment in the shareholding of banks in emerging economies, as foreign investors will typically encounter more severe information asymmetries in these cases. To empirically test this, we disaggregate the bank-year level institutional investor holding share to bank-investor-year-level data to capture geographical information at an individual investor level. Empirically, in line with the specifications in Mücke (2023) to control for both bank fundamental and market pricing variables, we estimate specifications (8) to (9) using bank-investor-year data.

$$\begin{aligned}\%InstInv\_own\_sh_{b,h,t} &\\ &= \beta_0 + \beta_1 DFscores_{b,t} + \delta_1 bk\&market\ control_{b,t\ or\ b,t-1} \\ &+ FEs + u_{b,h,t}\end{aligned} \tag{8}$$

$$\begin{aligned}\%InstInv\_own\_sh_{b,h,t} &\\ &= \beta_0 + \beta_1 DFscores_{b,t} + \beta_2 DFscores_{b,t} * dumEMbank_b \\ &+ \beta_3 DFscores_{b,t} * dumforINV_h + \beta_4 DFscores_{b,t} \\ &* dumEMbank_b * dumforINV_h + \gamma_1 dumEMbank_b \\ &+ \gamma_2 dumforINV_h + \gamma_3 dumEMbank_b * dumforINV_h \\ &+ \delta_1 bk\&market\ control_{b,t\ or\ b,t-1} + FEs + u_{b,h,t}\end{aligned} \tag{9}$$

where $\%InstInv\_own\_sh_{b,h,t}$ is the percentage point of common share holdings of a bank b owned by active individual institutional investors h in the end of year t in (8) to (9). BK&market control includes the same variables as in specification (6). Bank, year and active institutional investor fixed effects are included in these two specifications. In specification (9), DumEMbank takes the value of 1 if the headquarters country location of the bank is in an emerging economy, and 0 otherwise. DumforINV takes the value of 1 if the headquarter country of the institutional investor is different from the headquarter country of the bank, and 0 otherwise. In specification (8), the coefficient of interest is β_1, which measures the effect of a higher DFscore on the share of ownership by active institutional investors of bank b. In specification (9), the coefficient β_4 captures whether there is positive difference in the slope effect of a higher DFscore on the shareholdings of foreign institutional investors in emerging economy banks, relative to local investors. The underlying intuition is that digital technological exposure is a valuable piece of non-financial information for institutional investors, and the value of such information is more prominent under more severe asymmetric information situations.

Table A8 presents the results for our specifications (8) and (9). We again find a positive and statistically significant coefficient β_1 in Column (1), which supports the findings in Table 4.3. The magnitude indicates that a 3.5-ppt increase in DFscore will be associated with a 32% increase in active institutional investor's shareholding of banks, compared with a mean level (0.1 ppt) from the summary statistics table. In Column (4), we report the evidence supporting the well-known home bias investment pattern among investors. The negative effect is stronger for banks in emerging economies, which suggests that information asymmetries effects are present in our bank-investor-year sample. In Column (5) of Table A5, consistent with the intuition above, we find that β_4 is positive and statistically significant. The combined slope estimate on DFscore of active institutional investor holding in this case is $\beta_1 + \beta_2 + \beta_3 + \beta_4$=0.194, which is slightly higher than the standalone β_1=0.172 in column (5) for the case of domestic investors on advanced economy banks. These results provide evidence that the information value of a higher DFscore is meaningful for attracting active institutional investors holdings, and that this is more important for foreign investors in emerging economy banks.

Table A7: The effect of DFscores on banks' share of active institutional investor holdings

	Panel A			Panel B
Column	(1)	(2)	Column	(1)
Dependent variables	sh_ACT_instinv b,t	sh_ACT_instinv b,t	Dependent variables	sh_ACT_instinv b,t
DFscore b,t (β_1)	17.727*	9.505	DFscore b,t (β_1)	-4.694***
	(9.533)	(11.639)		(1.520)
DFscore b,t*dum_aft20 (β_2)		21.611	Sh_act_instinv b,t	0.004
		(18.454)		(0.003)
(β1 + β2)		31.116**	DFscore b,t *sh_act_instinv b,t	0.143***
		(14.877)		(0.048)
12m returnVOL b,t	-1.325	-1.408	12m returnVOL b,t	-1.325
	(1.928)	(1.930)		(1.928)
12m return b,t	0.712	0.707	12m return b,t	0.712
	(0.654)	(0.653)		(0.654)
RoE b,t-1	-0.012	-0.011	RoE b,t-1	-0.012
	(0.043)	(0.043)		(0.043)
Size b,t-1	2.141*	2.196*	Size b,t-1	2.141*
	(1.288)	(1.288)		(1.288)
NPL b,t-1	0.170	0.170	NPL b,t-1	0.170
	(0.228)	(0.227)		(0.228)
CapRatio b,t-1	0.050	0.048	CapRatio b,t-1	0.050
	(0.106)	(0.106)		(0.106)
PBR b,t-1	0.604	0.569	PBR b,t-1	0.604
	(0.560)	(0.557)		(0.560)
Constant	-26.004	-27.436	Constant	-26.004
	(35.275)	(35.284)		(35.275)
No. of obs.	1,608	1,608	No. of obs.	1,608
Within R2	0.00656	0.00681	Within R2	0.00656
Bank FE	Yes	Yes	Bank FE	Yes
ExchangeCtry#Year FE	Yes	Yes	ExchangeCtry#Year FE	Yes

Clustered-by-bank standard errors reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Empirical results for specifications (8) and (9) based on bank-investor-year level data.

Columns	(1)	(2)	(3)	(4)	(5)
Dependent variables	%InstInv_own_sh$_{b,h,t}$	%InstInv_own_sh$_{b,h,t}$	%InstInv_own_sh$_{b,h,t}$	%InstInv_own_shb,h,t	%InstInv_own_shb,h,t
DFscore $_{b,t}$ (β_1)	0.0923**	0.1971	0.0851		0.1726
	(0.0401)	(0.2234)	(0.0643)		(0.2583)
DFscore $_{b,t}$*dumEMEbk$_b$ (β_2)			0.0146		-1.1914**
			(0.0796)		(0.4975)
DFscore $_{b,t}$*dumforINV$_h$ (β_3)		-0.1609			-0.1747
		(0.2998)			(0.4549)
DFscore $_{b,t}$*dumEMEbk$_b$*dumforINV$_h$ (β_4)					1.3877**
					(0.6561)
dumforINVh (γ2)		-0.1661***		-0.1578***	-0.1556***
		(0.0134)		(0.0137)	(0.0140)
dumEMEbkb*dumforINVh (γ3)				-0.1511***	-0.2012***
				(0.0372)	(0.0438)
12m return b,t	0.0002	0.0008	0.0002	0.0023	0.0005
	(0.0028)	(0.0027)	(0.0028)	(0.0029)	(0.0027)
12m returnVOL b,t	0.0041	0.0007	0.0041	-0.0049	0.0007
	(0.0086)	(0.0079)	(0.0086)	(0.0079)	(0.0079)
PBR b,t-1	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Size b,t-1	-0.0434***	-0.0394***	-0.0434***	-0.0382***	-0.0395***
	(0.0053)	(0.0052)	(0.0053)	(0.0055)	(0.0052)
CapRatio b,t-1	0.0027***	0.0028***	0.0027***	0.0032***	0.0028***
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0007)
RoE b,t-1	-0.0006**	-0.0004	-0.0006**	-0.0004	-0.0004
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
NPL b,t-1	-0.0023**	-0.0021**	-0.0023**	-0.0021*	-0.0021**
	(0.0010)	(0.0010)	(0.0010)	(0.0012)	(0.0010)
Outstanding amount of common share b,t	-0.0424***	-0.0395***	-0.0425***	-0.0411***	-0.0402***
	(0.0098)	(0.0098)	(0.0099)	(0.0097)	(0.0098)
Constant	1.2910***	1.2601***	1.2908***	1.2476***	1.2868***
	(0.1532)	(0.1487)	(0.1530)	(0.1575)	(0.1504)
No. of obs.	617,804	617,804	617,804	651,346	617,804
Within R2	0.000767	0.0177	0.000765	0.0178	0.0186
Bank FE	Yes	Yes	Yes	Yes	Yes
InstINV FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Note: Clustered-by-banks standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1