Applied Economics Letters

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rael20

Digitalization and firm performance: channels and heterogeneities

Ruonan Cheng, Chun-Yu Ho & Shaoqing Huang

To cite this article: Ruonan Cheng, Chun-Yu Ho & Shaoqing Huang (2023) Digitalization and firm performance: channels and heterogeneities, Applied Economics Letters, 30:17, 2401-2406, DOI: 10.1080/13504851.2022.2097178

To link to this article: https://doi.org/10.1080/13504851.2022.2097178

Published online: 04 Jul 2022.

APPLIED ECONOMICS LETTERS
2023, VOL. 30, NO. 17, 2401–2406
https://doi.org/10.1080/13504851.2022.2097178

Digitalization and firm performance: channels and heterogeneities

Ruonan Cheng[a], Chun-Yu Ho[b] and Shaoqing Huang[a]

[a]Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, P. R. China; [b]Department of Economics, University at Albany, SUNY, NY, USA

ABSTRACT

In this paper we develop a text-based measure of digitalization using firms' annual reports with which we examine the effects of digitalization on firm performance using a dataset covering manufacturing and service firms. Our findings show that digitalization increases profitability in manufacturing firms by improving the efficiency of asset utilization to generate sales, yet it has no significant effects on overall profitability in service firms. It is because, on the one hand, digitalization improves asset utilization in these service firms while, on the other hand, cutting into their profits from sales. We also find evidence suggesting that digitalization enhances performance more for firms operating in highly competitive industries, smaller firms and firms with fewer skilled workers. Finally, our main results are robust to the use of instrumental variable estimation, the inclusion of R&D expenses or spatial autocorrelation term as a confounder, the use of mediating model, and the use of alternative fixed effects.

KEYWORDS

Digitalization; firm performance; information technology; textual analysis

JEL CLASSIFICATION

G32; O33

I. Introduction

Digitalization has been transforming businesses for decades.[1] In the 1960s and '70s, many firms began replacing manually performed labour, such as document processing and file storage, with computerized systems. After several waves of digitalization, the ongoing trend now involves big data, artificial intelligence, cloud computing and so on, enabling firms to use data to improve their operations and customer service. Such technology development poses an empirical challenge, however, to efforts to assess the impacts of digitalization on firm performance.

During the earlier waves of digitalization, studies quantified information technology (IT) inputs, such as installing computers, and assessed the impact of IT on firm performance (Brynjolfsson and Hitt 2003). Nonetheless, IT inputs cannot be used to fully assess the impacts of the current wave of digitalization on firm performance because digitalization not only represents an input but has also contributed to the transformation of business operations and product delivery. Unfortunately, the missing statistical link between digitalization and firm performance has already been identified (Brynjolfsson, Rock, and Syverson 2019). Since traditional measures of new technology are inadequate to meet the empirical challenge, it is imperative to develop new empirical methods with which to assess the impact of the current wave of digitalization.

In this paper we apply a text-mining method to annual reports for constructing a measure of digitalization and examine the effects of digitalization on firm performance using a dataset covering Chinese manufacturing and service firms over the 2010–2019 period. We were motivated to use China as our case in part because digitalization has had a strong impact on its economy. McKinsey Global Institute (2014) forecasts that new internet applications could fuel up to 1% increments in annual GDP growth in China from 2013 to 2025. By 2021 in China, there were roughly 1 billion internet users, supporting robust growth in e-commerce (McKinsey & Company 2021).

Our study relates to the literature that examines the effects of new technology on firm performance, see Brynjolfsson and Hitt (2003) as a seminal study.

CONTACT Shaoqing Huang sqhuang@sjtu.edu.cn Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, 200030, P. R. China.

This article has been republished with minor changes. These changes do not impact the academic content of the article.

[1]Digitalization is the process of employing digital technologies and information to transform business operations, according to Gartner, Inc. See https://www.gartner.com/en/information-technology/glossary

In several studies that are closely related to ours, researchers construct text-based measures of digitalization to examine its effects on firm performance [2] Chen and Srinivasan (2019) find a positive effect of digitalization on asset turnover ratio (ATR) but no effect on return on assets (ROA) for a sample of US firms. Industry-specific studies also report positive effects of digitalization on performance in US energy firms (Lyu and Liu 2021). Examining China, Guo and Xu (2021) study a sample of manufacturing firms and find positive and negative effects of digitalization on gross margins and ROA, respectively.

Although we follow the previous studies to adopt the word frequency approach constructing our text-based measure of digitalization, our study contributes to this growing literature by providing novel insights in how digitalization affects firm performance. First, we employ the relationship $ROA = ROS \times ATR$ to explore how digitalization affects profitability. Specifically, there are two channels – a profit channel based on return on sales (ROS), i.e. profits retained from sales, and an operations channel based on ATR, i.e. the utilization of assets to generate sales. Second, digitalization is potentially implemented at different stages of production in manufacturing and service firms. We provide novel evidence pertaining to cross-industry and cross-firm heterogeneities in the effects of digitalization on firm performance.

The remainder of our paper is organized as follows: Section 2 presents the empirical methodology. We discuss the results and its implications in Section 3.

II. Data and empirical methodology

Our empirical analysis is based on an unbalanced dataset of Chinese firms covering 2281 firms from 29 manufacturing industries and 306 firms from 7 service industries over the 2010–2019 period. The financial data are collected from the CSMAR database and the textual data are obtained from annual reports issued by those listed firms. All the continuous variables used in the analysis are winsorized at the 1st and 99th percentiles.

To investigate the effects of digitalization on firm performance, our baseline estimating equation is:

$$y_{ijt} = DText_{ijt} \times \beta + X_{it}\gamma + \alpha_j + \alpha_t + u_{it} \qquad (1)$$

where firm, industry, and year are denoted by i, j, and t, respectively.

The dependent variable, y_{ijt}, represents several measures of firm performance. To understand how digitalization affects performance, we include three performance measures, namely ROA, ROS and ATR. We employ ROA as the overall performance measure, while the other two measures enable us to test the channels through which digitalization affects ROA. ROS is the net profit generated by each dollar in sales, which represents profits retained from sales after accounting for depreciation, interest, and operating costs. The higher the ROS, the more efficient a firm is at generating bottom-line profits from its top-line revenue. ATR is sales per dollar of assets, which represents the efficiency with which a firm uses its assets to generate sales. The higher the ATR, the more efficient a firm is at generating sales from its assets. In other words, ROS can be used to measure profitability efficiency and ATR can be used to measure operational efficiency. The results we report in Table 1 indicate that our sample firms earn ROA of 4%, ROS of 7%–9%, and ATR of 64%–98%.

The key variable of interest in our study is $DText$, which measures the level of digitalization. We construct $DText$ in several steps using a text-mining approach. First, we perform word segmentation on the contents of firms' annual reports using Jieba, a widely employed Python-based text-segmentation software for Chinese texts. Second, we customize and expand Jieba's lexicon to include proprietary words associated with digitalization. Third, we filter the segmented texts manually to generate a word list associated with digitalization, denoted by P. We then construct the $DText_{ijt}$ measure in reference to the frequencies with which those keywords appear in the annual report of firm i operating in industry j in year t relative to the industry total, as follows:

[2]See Loughran and McDonald (2016) for a recent survey of the use of textual analysis in finance research..

Table 1. Variable definitions and descriptive statistics.

	Manufacturing							Service						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Variables	Mean	SD	Min	P25	Median	P75	Max	Mean	SD	Min	P25	Median	P75	Max
Dep. Variables														
ROA	0.04	0.063	−0.26	0.02	0.04	0.07	0.20	0.04	0.051	−0.24	0.02	0.04	0.06	0.17
ROS	0.07	0.150	−0.80	0.03	0.07	0.13	0.43	0.09	0.183	−0.64	0.02	0.04	0.11	1.05
ATR	0.64	0.370	0.11	0.40	0.56	0.78	2.24	0.98	0.912	0.04	0.32	0.72	1.35	5.44
Indep. Variable														
DText	0.02	0.025	0.00	0.00	0.01	0.02	0.16	0.02	0.045	0.00	0.01	0.01	0.02	0.33
RD	14.88	6.723	0	16.28	17.46	18.40	21.48	5.91	8.053	0	0	0	16.08	19.76
Control Variables														
Size	21.90	1.165	19.71	21.05	21.74	22.56	25.36	22.52	1.375	19.24	21.58	22.49	23.40	26.02
Leverage	0.39	0.203	0.05	0.23	0.38	0.54	0.92	0.50	0.207	0.07	0.36	0.50	0.65	0.97
Liquidity	1.19	0.795	0.19	0.67	0.99	1.46	4.82	2.02	1.568	0.14	0.96	1.70	2.56	9.22
Subsidy	12.58	6.515	0	12.60	15.37	16.59	19.91	12.23	6.656	0	11.51	15.08	16.58	20.53
State	0.28	0.449	0	0	0	1	1	0.56	0.497	0	0	1	1	1

Observation = 16,438 for manufacturing and 2,556 for service industries. *ROA* is the net profit-to-assets ratio. *ROS* is the net profit-to-operating income ratio. *ATR* is the operating income-to-assets ratio. *RD* is the logarithm of R&D expenditures. *Size* is the logarithm of total assets. *Leverage* is the debt-to-assets ratio. *Liquidity* is the liquid assets turnover ratio. *Subsidy* is the logarithm of government subsides. *State* is the dummy variable for state-owned enterprises.

$$DText_{ijt} = \sum_k 1\{\omega_{ijkt} \in P\} \Big/ \sum_m \sum_k 1\{\omega_{mjkt} \in P\}$$

where ω_{ijkt} is the k^{th} word appearing in the annual report of firm i in industry j in year t. In Table 1 we report the mean of *DText* as 2%, but its distribution is right-skewed, suggesting that a group of firms mentions digitalization-related keywords more often than the other firms in the sample. The idea underlying *DText* is that the more frequently a firm uses digitalization keywords, the higher is its level of digitalization. Figure 1 depicts the 30 most frequently appearing digitalization-related keywords we find in the annual reports of our sample firms. ‘Information’ and ‘Intelligence’ are two commonly observed keywords for both manufacturing and service firms. ‘Automation’ and ‘E-Commerce’ are used frequently by manufacturing and service firms, respectively. These intuitive results illustrate the relevance of our text-based measure of digitalization. In Equation (1), the coefficient β captures the effects of digitalization on firm performance.

We include a vector of firm characteristics X_{it} that include firm size (*Size*), the debt-to-asset ratio (*Leverage*), the ratio of liquid assets to total assets (*Liquidity*), government subsidies (*Subsidy*), and whether a firm is state-owned (*State*) as control variables. We also include a full set of industry-specific fixed effects (α_j) and year-specific fixed effects (α_t).[3] Industry-specific fixed effects control for time-invariant unobservable factors across industries that affect performance. Year fixed effects control for macroeconomic shocks that affect the performance of all firms. The error term u_{it} denotes all omitted idiosyncratic shocks to firm performance.

III. Results

In columns 1–3 of Panel A in Table 2 we report the main results derived by estimating Equation (1) for manufacturing industries. For column 1 we employ y_{ijt} = *ROA* and report a positive and significant coefficient of *DText*, indicating that digitalization enhances the overall profitability of manufacturing firms. A one-standard-deviation increase in *DText* (0.025) raises *ROA* by 0.0014 (= 0.054 × 0.025), which is about 2.2% of average *ROA*. In column 2 we report a nonsignificant coefficient of *DText* when we employ y_{ijt} = *ROS*, whereas in column 3 we report a positive and significant coefficient of *DText* when we employ y_{ijt} = *ATR*. These results suggest that digitalization improves the efficiency of asset utilization to generate sales, which in turn enhances overall profitability.

In columns 4–6 of Panel A in Table 2 we report the main results derived by estimating Equation (1) for service industries. For column 4 we employ y_{ijt} = *ROA* and report a nonsignificant coefficient of *DText*, indicating that

[3] We employ Hausman test to select whether the fixed effects model or the random effects model should be used. The empirical results consistently suggest using the fixed effects model, see Panel A of Table 2 for the details.

Figure 1. Word clouds of the most frequent keywords.

digitalization does not affect overall profitability in service firms. In column 5 we report a negative and significant coefficient of $DText$ when we employ $y_{ijt} = ROS$, whereas in column 6 we report a positive and significant coefficient of $DText$ when we employ $y_{ijt} = ATR$. Consistent with the results our model generates for manufacturing firms, here we find that digitalization increases the efficiency of asset utilization to generate sales. In contrast to the effects we find for manufacturing firms, here we find that digitalization reduces profits retained from sales. This result may reflect the increased costs, such as higher sales and marketing expenses, associated with digitalization.

We then conduct a series of robustness checks. First, reverse causality may lead our measure of digitalization to suffer from endogeneity. For example, a less profitable firm may implement digitalization to boost its profits. We instrument our measure of digitalization with its one-year lag and report the results of the IV estimation in Panel B. Second, R&D may be an omitted variable that drives performance and digitalization. We include RD in Equation (1) and report the results in Panel C. Third, performance of nearby firms may be an omitted variable that drives performance and digitalization. We include a lagged spatial autocorrelation term of firm performance,

Table 2. Main results.

	Manufacturing			Service		
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var y_{ijt} =	*ROA*	*ROS*	*ATR*	*ROA*	*ROS*	*ATR*
Panel A: Baseline						
DText	0.054**	0.082	0.273**	0.018	−0.507***	3.012***
	(0.027)	(0.065)	(0.117)	(0.038)	(0.131)	(0.456)
R-squared	0.259	0.236	0.694	0.221	0.283	0.647
Hausman Test (P value)	0.000	0.000	0.000	0.000	0.000	0.000
Panel B: IV						
DText	0.018	0.039	0.314**	−0.015	−0.726***	3.605***
	(0.035)	(0.085)	(0.155)	(0.034)	(0.128)	(0.445)
F-statistic	342.23	342.23	342.23	95.71	95.71	95.71
R-squared	0.262	0.217	0.694	0.218	0.281	0.654
R-squared						
Panel C: Confounder						
DText	0.050*	0.080	0.203*	0.033	−0.361***	2.630***
	(0.027)	(0.065)	(0.117)	(0.039)	(0.134)	(0.467)
RD	0.015**	0.008	0.239***	−0.021*	−0.216***	0.565***
	(0.008)	(0.019)	(0.028)	(0.013)	(0.045)	(0.157)
R-squared	0.259	0.236	0.696	0.222	0.289	0.649
Panel D: Spatial Spillover						
DText	0.081***	0.082	0.310**	0.010	−0.584***	3.313***
	(0.030)	(0.065)	(0.129)	(0.043)	(0.150)	(0.511)
$Wy_{-ijt\text{-}1}$	0.018	−0.024*	0.013	−0.012	−0.008	−0.129***
	(0.023)	(0.014)	(0.016)	(0.034)	(0.014)	(0.035)
R-squared	0.261	0.236	0.693	0.222	0.283	0.661
Panel E: Mediating Effect						
DText	0.044*		0.273**	−0.007		3.012***
	(0.026)		(0.117)	(0.038)		(0.456)
ATR	0.075***			0.008***		
	(0.002)			(0.002)		
R-squared	0.313		0.694	0.229		0.647
Panel F: Interactive FEs						
DText	0.063**	0.100	0.226*	0.018	−0.543***	3.076***
	(0.029)	(0.069)	(0.123)	(0.040)	(0.138)	(0.493)
R-squared	0.289	0.262	0.701	0.287	0.336	0.659
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Province FEs	A-E	A-E	A-E	A-E	A-E	A-E
Industry FEs	A-E	A-E	A-E	A-E	A-E	A-E
Year FEs	A-E	A-E	A-E	A-E	A-E	A-E
Province x Year FEs	F	F	F	F	F	F
Industry x Year FEs	F	F	F	F	F	F

Observations = 16,438 for manufacturing and 2,556 for service industries in Panel A, C, E and F. Observations = 14,949 for manufacturing and 2,433 for service industries in Panel B and D. Panel A: Hausman test is a specification test between random effects and fixed effects models. Panel C: For exposition, we divide *RD* by 100. Panel D: $Wy_{-ijt\text{-}1}$ refers to the weighted average of lagged dependent variables of other firms in the same province, using inverse distance weighting *W* computed by the latitude and longitude of the firms' headquarters. Firm controls include *Size, Leverage, Liquidity, Subsidy*, and *State*. Standard errors in parentheses, ***p < .01, **p < .05, *p < .1

$Wy_{-ijt\text{-}1}$, in Equation (1) and report the results in Panel D. Fourth, to verify the efficiency of asset utilization as a mechanism for digitalization to improve firm performance, we employ the efficiency of asset utilization as a mediating variable in the model of *ROA* and report the results in Panel E. And finally, we expand the additive fixed effects to include interaction fixed effects at the province-year and industry-year levels in Panel F, which should better capture policy and industry effects on firm performance over time. Encouragingly, our results are robust to those checks.

Finally, we explore heterogeneities in our findings and report the results in Table 3. For Panel A we include the Herfindahl–Hirschman Index (HHI) and its interaction term with *DText* in Equation (1). The coefficients of the interaction terms have signs that are the opposites of those attaching to *DText*. The same pattern appears, as reported in Panel B, when we include *Status* (the firm-to-industry operating-income ratio) and its interaction with *DText* in Equation (1). The same pattern once again appears, as reported in Panel C, when we include *Technician* (the proportion of technicians to all workers) and its interaction with *DText* in Equation (1). Overall, our

Table 3. Heterogeneities.

	Manufacturing			Service		
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var y_{ijt} =	*ROA*	*ROS*	*ATR*	*ROA*	*ROS*	*ATR*
Panel A: HHI						
DText	0.090***	0.088	0.578***	−0.048	−0.738***	4.462***
	(0.031)	(0.074)	(0.128)	(0.049)	(0.171)	(0.596)
DText x *HHI*	−0.352**	−0.062	−4.224***	0.190**	0.651**	−4.063***
	(0.148)	(0.358)	(0.755)	(0.089)	(0.309)	(1.078)
R-squared	0.259	0.236	0.695	0.225	0.284	0.649
Panel B: Status						
DText	0.078***	0.103	0.343***	0.021	−0.562***	3.469***
	(0.028)	(0.069)	(0.119)	(0.038)	(0.133)	(0.458)
DText x *Status*	−0.732***	−0.371	−10.508***	−0.065	0.975**	−9.202***
	(0.163)	(0.395)	(0.852)	(0.122)	(0.425)	(1.465)
R-squared	0.260	0.236	0.702	0.222	0.284	0.657
R-squared						
Panel C: Technician						
DText	0.035	0.039	0.245**	0.025	−0.491***	3.152***
	(0.027)	(0.065)	(0.117)	(0.038)	(0.131)	(0.458)
DText x *Technician*	−0.521***	−1.176***	−0.745**	−0.144	−0.542	4.400
	(0.098)	(0.236)	(0.374)	(0.250)	(0.870)	(3.039)
R-squared	0.260	0.237	0.694	0.231	0.288	0.650
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes
Industry FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Observations = 16,438 for manufacturing and 2,556 for service industries. Firm controls include *Size, Leverage, Liquidity, Subsidy*, and *State*. Also, *HHI, Status* and *Technician* are included in the model used in Panel A, B and C, respectively. *HHI* is the Herfindahl-Hirschman index for each two-digit industry based on the market share of firm. *Status* is the firm-to-industry operating income ratio. *Technician* is the ratio of technician to total employees in a firm. Standard errors in parentheses, ***p < .01, **p < .05, *p < .1

results suggest that digitalization enhances performance more firms operating in highly competitive industries, smaller firms and firms with few skilled workers.

IV. Conclusion

Using a sample of Chinese firms over the 2010–19 period, this paper finds that digitalization increases profitability in manufacturing firms but not in service firms. Our results also highlight policy and managerial implications that industry and firm heterogeneities play a role in determining the benefit of digitalization.

Acknowledgement

Shaoqing Huang would like to acknowledge financial support from the National Natural Science Foundation of China (project number: 71973096) and the National Social Science Fund of China (project number: 22AZD036).

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Brynjolfsson, E., and L. M. Hitt. 2003. "Computing Productivity: Firm-level Evidence." *Review of Economics and Statistics* 85 (4): 793–808. doi:10.1162/003465303772815736.

Brynjolfsson, E., D. Rock, and C. Syverson. 2019. "Artificial Intelligence and the Modern Productivity Paradox." In *The Economics of Artificial Intelligence: An Agenda*, edited by A. Agrawal, J. Gans, and A. Goldfarb, 23–57. Chicago: University of Chicago Press.

Chen, W., and S. Srinivasan. 2019. *Going Digital: Implications for Firm Value and Performance*. Boston: *Harvard Business School*.

Guo, L., and L. Xu. 2021. "The Effects of Digital Transformation on Firm Performance: Evidence from China's Manufacturing Sector." *Sustainability* 13 (22): 12844. doi:10.3390/su132212844.

Loughran, T., and B. McDonald. 2016. "Textual Analysis in Accounting and Finance: A Survey." *Journal of Accounting Research* 54 (4): 1187–1230. doi:10.1111/1475-679X.12123.

Lyu, W., and J. Liu. 2021. "Artificial Intelligence and Emerging Digital Technologies in the Energy Sector." *Applied Energy* 303: 117615. doi:10.1016/j.apenergy.2021.117615.

McKinsey & Company. 2021. *The Future of Digital Innovation in China*. McKinsey & Company.

McKinsey Global Institute. 2014. *China's Digital Transformation*. McKinsey & Company.