Broadband Policy, Digitalization, and Firm Performance: Evidence from China*

Pavel Chakraborty† Wenbohao Zhu‡

April 2024

Abstract

Do broadband policies affect firm performance? Exploiting a quasi-experimental pilot program, which is a part of a bigger policy design to improve the broadband infrastructure and connectivity, implemented by the Chinese Central Govt. and a novel dataset for Chinese listed firms which provides detailed information of digital processes of firms, we investigate the effect of this policy on firms' digitalization or digital adoption of different processes, innovation, and other firm level performance for the years 2007–2019. We find that the program increased the digital characteristics of an average Chinese firm located in a city where the policy was implemented, compared to a firm in a city which was not part of the policy, by 31%. We also find significant innovation effects – R&D expenditure and patent filings went up by 27% and 22%, resepctively. All these led to increase in sales and value-added by 15% and stock-market valuation by 16%. Overall, our results underscore that broadband is a important factor for improved firm performance, especially in emerging economies.

Keywords: Broadband Policy, Digitalization, Innovation, Firm Performance

JEL Codes: O12, O3, D2, G3, L25, M51

---

*All errors are our own.

†Department of Economics, Management School, Lancaster University, UK; E-mail: p.chakraborty1@lancaster.ac.uk

‡School of Economics and Trade, Guangdong University of Foreign Studies, China; E-mail: 13032264827@163.com

1 Introduction

Digital technologies have advanced more rapidly than any innovation in our history – reaching around 50% of the developing world's population only in two decades (World Bank, 2021). By enhancing connectivity, financial inclusion, access to trade and public services, technology can be a great equaliser and a key driver for transforming societies. Digitalization of information and its dissemination via the internet can also bring substantial benefits for firms. It can facilitate innovation, improve matching of workers to firms, reduce the time and effort required to learn new skills, expand firms' market reach, etc (EBRD, 2022). Governments, multilateral aid agencies now promote digitalization as a powerful driver of economic growth and development. Among them, the development of digital infrastructure, such as high-speed internet networks, is of first order importance. And, this has become even more important after the COVID-19 pandemic.

However, there is still limited causal evidence in terms of how access to improved digital technologies can affect firms. This is mainly due to paucity of quasi-experimental policy changes which can reflect new and improved access to digital technologies, and in addition reliable micro level data (for firms) which can measure different aspects of digitalization, such as internet adoption, use of smart and intelligent technology, etc.

This paper overcomes these constraints by exploiting a quasi-experimental policy undertaken in China in 2013 and studying its effect on firm level digitalization, innovation, and performance by utilizing a unique dataset for Chinese firms which contains data on different aspects of digitalization as well as economic activities such as R&D expenses, patents filed, labour compensation, sales, value-added, etc. Our study is one of the first to utilize such a unique policy change for a developing country which is quasi-random in nature and can directly affect the digitalization or digital adoption of different processes by firms. This is the primary contribution of our study.

The setting for our paper is China, a country that decided to invest significantly in broadband infrastructure, due to its previous poor broadband infrastructure and high internet service charges, during 2013–2019. The main objectives of the strategy is to develop a nationwide coordinated broadband network in conjunction with improvement in network speed and reliability.[1] However, before the implementation of the bigger plan, known as "Broadband China Strategy" China decided to adopt a pilot policy over three-year period (between 2014–2016) to construct broadband infrastructure in 39 cities in each of the three years of the pilot policy. This pilot policy gives an ideal setting to test for the effect of broadband infrastructure on firm digitalization and performance for the following reasons.

First, the broadband pilot policy can be regarded as an exogeneous positive shock that can affect firms' digitalization and performance by upgrading broadband infrastructure in cities which are covered by the pilot policy compared to firms in cities which were not. According to Wang et al. (2023), the pilot policy was largely unexpected as it was implemented by the Chinese central government in a top-to-bottom approach. Second, the pilot policy was targeted to implement across different cities in three different phases thereby helping us to avoid a common spatial bias that can be possibly make the policy endogenous. For example, in 2014, it was implemented in Beijing in the North to Tibetan autonomous region in the far West to Chengdu in the Middle to Guangzhou in the South.

However, the policy could still be endogeneous in nature given that the pilot policy could only target those cities where firms had predominantly low index of digitalization or firms were performing worse than firm in non-pilot cities. Or firms in the pilot cities could be on completely different trends, in terms of their overall performance, digitalization, internet adoption, etc., than firms in the non-pilot cities. We check for all those endogeneity concerns by (a) regressing past values of firm performance, digitalization

[1] Details of the plan is discussed in Section 2.

measures on the broadband policy, and (b) interacting year fixed effects with city level dummies where the pilot policy was implemented. We do not find either any firm level characteristics to be influencing the policy implementation or any discernible differences in firm performance across pilot and non-pilot cities before the implementation of the policy.

Our analysis using this quasi-random policy change centers around the construction of an extensive micro level dataset on digitalization for about 4500+ Chinese firms listed in the major stock exchanges for the years 2007–2019 using the following steps. First, we download detailed annual reports of all the listed firms across all these years using the R language. The reports mention all the key features regarding the digital characteristics of a firm. For example, it would mention if the firm has adopted a new internet connection or the firm uses digital currency or the firm uses digital platform for its transactions, etc. We then created a thesaurus of digitalization based on the policy documents by the Ministry of Industry and Information Technology, People's Republic of China. Lastly, we used textual analysis by Python to extract the core keywords of the digitalization of a firm and match those keywords to our thesaurus to build the digitalization measure which is based on the frequency of the words related to digitalization that had appeared in the annual reports of a firm. The higher the frequency, the higher is the digital adoption of that firm or higher the digitalization index. Using this dataset which would represent the digital character of a firm at the most disaggregated level is the secondary contribution of the paper.

We matched this data on firms' digital frequency with detailed information on firm characteristics using the China Stock Market & Accounting Research Database based on an unique firm identifier and firm name. This dataset rolls out details regarding the name of the firm, its industry affiliation, R&D expenditure, number of patents and inventions, as well as data on sales, value-added, total employees (divided into R&D workers, senior managers, etc.), wages, etc.

Having the policy change in the background and equipped with these datasets, we use a simple differences-in-differences design to carry out our empirical exercise. The key point in exploiting this particular policy change regarding broadband infrastructure for causal inference is that the implementation of the policy provides a plausible exogenous change in city level dynamics, in our case the cities which were part of the pilot program (representing the 'treatment' group) relative to the rest of the cities (representing the 'control' group). We carry out our exercise at two different levels.

In the first part, we find a remarkably persistent and economically meaningful positive effect of the broadband policy on the digitalization of firms. In particular, our main finding is that the pilot policy led to a significant increase in the digital adoption of different processes by firms located in cities where the pilot policy was implemented by 31% relative to firms located in control or non-pilot cities. In terms of extensive margin, the policy increased the probability of adopting a new digital characteristic or feature of a firm by 0.28. This increase in overall digitalization of firms is significantly driven by adoption of smart and intelligent systems (41%), internet (22%) and digital (15%) technology.

In our reduced form regressions, the identifying assumption is that the location of a firm, in a city where the pilot program was implemented, does not affect its digitalization level except through the implementation of broadband infrastructure due to the broadband pilot policy program. We address the validity of this assumption in a variety of ways: (a) we test how our diff-in-diff estimate is affected by inclusion of variety of controls at firm and city level, both time-varying and time-invariant; and (b) to address the concerns that the implementation of broadband infrastructure could be correlated with firm level characteristics which are predisposed to overall digital frequency, we employ instrumental variable estimation.

Our instrument variable strategy is motivated by historical telephone infrastructure, in particular the dial-up connections, and topographical characteristics. We use an instrument that can plausibly exploit spatial variation in access to broadband connections due to historic differences in firms' access to the infrastructure which is key to the development of broadband infrastructure, the telephone network (DeStefano et al., 2018). In particular, we use the number of telephone users per 10,000 people in 1984 for each city. Our identifying assumption is that the historical telephone infrastructure may affect the choice of the cities where the pilot program was implemented and that eventually will have a positive effect on the digitalization of firms. We also check our result using geographical characteristics such as, elevation or gradient of cities. The higher the gradient, the difficult would be to establish a good broadband network, therefore the lower is the probability of that city to feature in the pilot program of the broadband strategy. Our estimates using the IV are consistent with that of OLS.

In the second part of the results, we investigate what happened to factors related to innovation – share of R&D workers, their average wages, R&D expenses, and number of patents filed. We find that share of R&D workers increased by about 8% in a firm located in a city where the pilot program was implemented. Similar was the effect for average wage for a R&D worker; it increased by 16%. We also find that the pilot program significantly increased R&D expenditure (27%) and patents filing (22%).

Lastly, we find that all these changes related to digitalization and innovation led to higher employment of both the factors of production (labour by 12% and capital by 21%), higher sales, value-added (15%), and stock-market valuation (16%). Our results also show that all these effects are driven by private and foreign firms across all sizes (with higher effects for marginally big firms (2nd tercile), followed by big, and the smallest firms) with no effects for State-owned enterprises (SoEs).

Related Literature: Our paper relates to several strands of the literature. First, it is closely related to the literature on how internet adoption affects various firm level outcomes. A sizeable amount of literature provides evidence on how internet connectivity can improve labour productivity (Akerman et al., 2015), labour market matching (Bhuller et al., 2023), management practices (Gokan et al., 2019), female employment (Chun & Tang, 2018), organization of production (Lin, 2019), firm entry (Hjort & Poulsen, 2019), firm level economic policy uncertainty (Wang et al., 2023), productivity (Commander et al., 2011), trade – both exports (Hjort & Poulsen, 2019) and imports (Malgouyres et al., 2021), etc.[2]

We deviate from the literature in two different ways: (a) we exploit a pilot program, which is part of a large future broadband infrastructure upgradation strategy, undertaken in a developing economy; and (b) focus on how broadband infrastructure affects firm level digitalization and innovation factors. In essence, our paper is closest to DeStefano et al. (2018). They utilize the arrival of a new communication technology, ASDL broadband, in the UK to study the effects on ICT use by firms. They find that ASDL broadband significantly improved ICT use by firms and increased firm size (measured through employment), but not firm productivity. Our paper is also close in spirit to Akerman et al. (2015), Haller & Lyons (2013), Fabling & Grimes (2021), Czernich (2014).

Second, our paper is also related to studies which focus on correction of information frictions. A large part of the literature investigates how information frictions play a crucial role and help explain trading patterns (Allen, 2014). We show that information frictions can also help firms to undertake higher innovation by employing more R&D workers. Houngbonon et al. (2021) exploits the variation in broadband infrastructure across cities

[2]There is also evidence that internet adoption can benefit human capital development, such as: poverty (Bahia et al., 2023), long-run academic achievement (Bianchi et al., 2022), students test scores in schools (Malamud et al., 2019), computer and cognitive skills (Malamud & Pop-Eleches, 2011), etc.

in Africa to show that firms are likely to undertake more process (20 percentage points) and product (12 percentage points) innovation when fast internet becomes available.

Finally, our paper is related to how broadband infrastructure and connection impacts overall firm performance. We show that it not only increases firm size (through use of higher production factors, such as higher labour and capital) and/or sales, but also leads to higher productivity (value-added), and valuation in the stock-market. This is similar to what Cariolle et al. (2019) finds evidence of positive association between internet use and firm performance (sales) for a sample of 30,000 firms in 38 developing and transition countries, DeStefano et al. (2018) finds for UK (increase in firm size), Chen et al. (2019) for Chinese firms (increase in firm productivity), Commander et al. (2011) for Brazilian and Indian firms (increase in firm productivity), etc.

The rest of the paper is organized as follows: Section 2 outlines the details of the broadband policy undertaken in China in 2013. We describe the data and present a few stylized facts in Section 3. The empirical strategy is illustrated in Section 4. Section 5 describes all the results while Section 6 provides some concluding remarks.

2 Broadband Policy

Broadband connection has a positive and significant effect by streamlining economic production activities and by making daily life more efficient for individuals. Establishment of widespread and faster connection to broadband technology has been promoted and used by many countries such as Norway (Akerman et al., 2015), France (Malgouyres et al., 2021), Turkey (Demir et al., 2023), etc.

China had a relatively poor broadband infrastructure and high internet service charges during the first decade of 21st century (Zhou et al., 2022). In particular, (i) broadband penetration rate was less than half that of OECD countries, around 21%; (ii) average broadband speed was less than one-tenth of that of OECD countries; and (iii) average cost of broadband use was three times that of OECD countries. To promote broadband technology, China outlined a plan known as "Broadband China Strategy" on August, 2013.[3]

This plan was a national strategy primarily implemented by the Ministry of Industry and Information Technology. The plan set specific targets in terms of promoting the diffusion and application of broadband technology, primarily increasing broadband penetration and speeding up existing networks to spread information and digitalization in China. The plan set a period of seven years, 2013 to 2019, to build a reliable broadband network.

The detailed goals of the strategy can be outlined as follows: (a) in terms of connection to households, by the end of the plan more than 70% of the households should have access to internet with atleast a speed of 20 Mbps and 100 Mbps in some developed cities; (b) in terms of connection to firms, the broadband strategy aims to boost business network speed to 1000 Mpbs; (c) promotion of coordinated regional broadband network development along with optimization and upgrading of existing broadband network and improvement of speed and reliability; (d) firms were encouraged to upgrade their intelligence levels; (e) digital industries will be developed, such as cloud computing and big data; and (f) use of broadband networks would be expanded to education, medical care, employment, and social security.

However, "Broadband China Strategy" also outlined a pilot policy phase (2014–2016) as a part of the bigger plan to complete broadband infrastructure construction within a three-year period in 39 pilot cities in each of the three yearly phases of the pilot policy. In

[3]See: http://www.gov.cn/zhengce/content/2013-08/16/content_5060.htm

particular, broadband infrastructure will be constructed in 39 cities[4] in 2014, 39 cities[5] in 2015, and 39 cities[6] in 2016.[7]

The pilot cities were selected by the Chinese government for the pilot program of the "Broadband China" based on the applications submitted by each city across China. However, two criteria were kept in mind: (a) pilot cities to have a certain level of foundation for the development of broadband facilities; and (b) pilot cities were distributed in different regions to increase the representativeness of the sample.

Figure 1 shows the geographical distribution of pilot cities across China. In this figure, the orange areas represented the first batch of cities where the policy was implemented in 2014; the pink and purple areas represent the cities for 2015 and 2016 phase of implementation. Blue areas represent cities that were not the part of the pilot phase. Therefore, our treated group is a batch of 117 cities and control group consists of all other cities across China. The spatial distribution of the pilot cities indicate no serious selection bias in the sample. For example, cities in the far west to the far north to cities in Tibet were also part of the program.

[4]The cities are: Beijing, Tianjin, Shanghai, Chang-Zhu-Tan urban agglomeration, Shijiazhuang, Dalian, Benxi, Yanbian Chaoxian Nationality Autonomous Prefecture (Korean), Harbin, Daqing, Nanjing, Suzhou, Zhenjiang, Kunshan, Jinhua, Wuhu, Anqing, Fuzhou (including Pingtan), Xiamen, Quanzhou, Nanchang, Shangrao, Qingdao, Zibo, Weihai, Linyi, Zhengzhou, Luoyang, Wuhan, Guangzhou, Shenzhen, Zhongshan, Chengdu, Panzhihua, Tibetan Qiang Autonomous Prefecture of Ngawa, Guiyang, Yinchuan, Wuzhong, and Alaer.

[5]The cities are: Taiyuan, Hohhot, Ordos, Anshan, Panjin, Baishan, Yangzhou, Jiaxing, Hefei, Tongling, Putian, Xinyu, Ganzhou, Dongying, Jining, Dezhou, Xinxiang, Yongcheng, Huangshi, Xiangyang, Yichang, Shiyan, Suizhou, Yueyang, Shantou, Meizhou, Dongguan, Jiangjin District of Chongqing, Rongchang District of Chongqing, Mianyang, Neijiang, Yibin, Dazhou, Yuxi, Lanzhou, Zhangye, Guyuan, Zhongwei, Karamay.

[6]The cities are: Yangquan, Jinzhong, Wuhai, Baotou, Tongliao, Shenyang, Mudanjiang, Wuxi, Taizhou, Nantong, Hangzhou, Suzhou, Huangshan, Maanshan, Ji'an, Yantai, Zaozhuang, Shangqiu, Jiaozuo, Nanyang, Ezhou, Hengyang, Yiyang, Yulin, Haikou, Jiulongpo District of Chongqing, Beibei District of Chongqing, Ya'an, Luzhou, Nanchong, Zunyi, Wenshan Zhuang and Miao Autonomous Prefecture, Lhasa, Nyingchi, Weinan, Wuwei, Jiuquan, Tianshui, Xining.

[7]The policy led to about 70% of all firms using a fixed broadband in those pilot cities along with more than 98% villages having access to broadband technology. In addition, the broadband speed reached 50Mbps and 12Mbps for urban and rural households, respectively.

3 Data and Stylized facts

3.1 Firm level data

We examine the effect of the broadband policy initiated in 2013 by the State Council of China on digitalization and performance of China's listed firms for the period 2007-–2019 using three comprehensive datasets: (a) the China Stock Market & Accounting Research Database (CSMAR, hereafter); (b) the Ministry of Civil Affairs of the People's Republic of China (MCA, hereafter), and (c) CNINFO provided by the China Securities Regulatory Commission (CSRC).

CSMAR provides detailed information on all the economic activities of more than 4500 firms listed in the major stock exchanges of China, starting from 1990. In addition, the dataset also contains information on some small and medium-scale enterprises. The dataset is maintained by a Hong Kong based company called GTA (Global Technology Alliance). The dataset rolls out details regarding the name of a firm, its industry affiliation, its ownership criteria (private domestic firms or state-owned enterprises (SoEs) or joint ventures or foreign-owned multinational affiliates), important indicators regarding both the innovation input and output, such as R&D expenditure and number of patents filed, as well as data on sales, value-added, operating cost, amount of capital employed, assets, different indicators regarding financial performance such as amount of stocks held by the firm, stock market value, number of shareholders, etc.

A key information that we utilize in our analysis is the details on the composition of employment in a firm and its corresponding wages. In particular, apart from mentioning total employees of a firm, CSMAR also categories workers of a firm into managers, and R&D workers. The managers category comprises of Chairman, members of Senior Management, and Supervisors. We exploit this information to understand the effects on different types of labour as digitalization of firms went up. This dataset has previously been used by Du & Boateng (2015), Chan et al. (2012), and Chen et al. (2011) among others.

The second dataset, MCA, provides us with the administrative code of the province and city where the firms are located. We match this database with CSMAR based on an unique firm identifier and the firm name. MCA also provides with the industry affiliation code for each firm according to the "Industry Classification for National Economy" (GB/T 4754-2017).

Lastly, the most important dataset that we use for our purpose is CNINFO. It contains detailed annual reports of all the listed firms from 2007–2019.[8] The annual reports are downloaded from CNINFO using R language. These reports give details of the key features regarding the digital characteristics of firms. In particular, it mentions all the keywords regarding the new digital process(es) adopted or upgrade of digital process(es) by a firm. For example, the annual report would mention if a firm has used artificial intelligence or robots or biometric technology or neuromorphic computing or cloud computing, etc. in their production process or any of the services rendered. In addition, these reports would repeat those words the number of times that particular firm has used this service. We use these keywords regarding the digital characteristics a firm to create our index on digitalization or digital adoption using the following steps:

First, we identify the keywords related to digitalization based on the policy documents by the Ministry of Industry and Information Technology, People's Republic of China. Secondly, we then use these words to create a thesaurus for a firm's digitalization of processes. Third, we use textual analysis or text mining and term frequency by Python to extract the core keywords of the digitalization of a firm and match those keywords to our thesaurus.[9]

[8]For details, please see: www.cninfo.com.cn

[9]Simultaneously, we also create another dictionary to delete auxiliary words, interjections, modal par-

Table A.1 lists all the keywords that are used to form the digitalization index. There are 144 keywords (or variables) highlighting the digitalization of a firm which we further break into 9 different heads: (i) internet technology; (ii) digital technology; (iii) smart and intelligent technology; (iv) automatic technology; (v) information management; (vi) big-data and cloud; (vii) AI and learning; (viii) integrated technology; and (ix) others. We calculate the frequency of those keywords the number of times they appear in a firm's annual report. The higher the frequency or the use of the keywords mentioned in the annual report of a firm, the higher is the digitalization frequency or process of the firm. Lastly, we match our indicator on digital frequency of firms with the other two datasets using the unique firm identifier. We use the logarithm of the frequency of digitalization as our main outcome variable of interest.

3.2 Stylized Facts

Figure 2 plots the overall digitalization frequency across all firms in a year **(Panel A)** and normalized digitalization frequency for an average firm located in cities which had a broadband policy implemented either in 2014 or 2015 or 2016 and the ones that never received such treatment **(Panel B)**. **Panel A** clearly shows that the rate of increase of digitalization increased leaps and bounds from 2014 onward. For example, the overall digital frequency across all firms in the year 2014 was 36,731 and this increased to 58,855 in 2015 to 79,485 in 2016, and 111,598 in 2017. This means that the reporting of keywords by firms related to digital characteristics over these three years increased by more than 200%.

Panel B shows that a large part of this increase in digital frequency of a firm is driven by firms located in cities with the implementation of the broadband policy. While there was no discernible difference in the trend of digital frequency for firms before 2013 across

ticles, onomatopoeia, other functional, and symbolic words in the process of text mining and frequency statistics.

cities which have the broadband policy and which have not, the difference started to grow from 2014 onward.

Table 1 calculates the median of digital frequency index for an average firm before and after the broadband policy at the aggregate **(Panel A)** and divided into cities with broadband policy and no broadband policy **(Panel B)**. The median frequency of digitalization for an average firm in any city across China was 4 (i.e., 4 keywords was mentioned in that firm's annual report regarding the use of digital processes) between 2007–2013 and it jumped to 20 afterwards; a 5-fold or 400% increase. In case of cities where the broadband policy was implemented, the median frequency for a firm was 5 and it increased to 23 (a very similar increase like the aggregate), whereas for its counterpart (cities with no broadband policy) the median frequency was 3 before 2014 and 15 after. The median values also portrays similar picture as the **Figure 2** – that the difference in the digitalization of firms across these two types of cities increased significantly after the broadband policy was implemented.

4 Empirical strategy

Having shown that firms in the cities with the "Broadband Policy" have on average higher digitalization than others, we now test for the direct effect of the policy on digitalization or digital characteristics and other outcomes of a firm using a simple OLS fixed effects diff-in-diff strategy. In our framework, we designate firms in the cities with the broadband policy as the "treated" group, whereas firms in all other cities as the "control" group. We use the following specification:

$$Ln(y_{fjct}) = \beta \left(Broadband\ Policy_{2013} \times City_{fc}\right) + \phi_f + \delta_c + \theta_{jt} + \epsilon_{fjt} \tag{1}$$

y_{fjct} represents our outcome variable of interest for firm f in sector j located in city c at time t. For our analysis, y takes a host of indicators – the overall digitalization index/frequency of a firm and its sub-components, such as internet adoption, digital technology adoption, smart and intelligent technology adoption, technology related to big-data and cloud, adoption of artificial intelligence and learning, etc.; different factors of innovation – R&D workers, R&D expenditure, number of patents filed, etc.; demand for different kinds of labour – total and managerial; and firm performance – capital employed, value-added, total sales, stock market value, etc.

$Broadband\ Policy_{2013}$ is a binary variable that takes the value of 1 in the post-policy, i.e., post-2013 years. The policy was announced by China State Council in 2013, indicating the pre-policy period to be 2007–13 and the post-policy 2014–2019. $City_{fc}$ is a binary variable that takes a value of 1 if the f-th firm is located in city c where the broadband policy was implemented either in 2014 or 2015 or 2016. For example, $City_{fc}$ equals 1 for firms located in 39 cities (where the broadband policy was implemented) in 2014. Similarly, for 2015 and 2016 in 39 cities each, respectively.

Certainly, the broadband policy is not the only external shock that could influence the performance of Chinese firms in terms of their digitalization or adoption of different upgraded digital and/or technological processes. There could be many other potentially exogenous events that would impact firm level decision-making process including time-varying domestic macroeconomic factors. It is quite likely that many of such events could impact our estimates, especially the events happening around 2013. In order to potentially avoid this problem with identification, we include firms located in other cities during the same time-frame as the control group. The argument goes as follows: any event that is not related to the broadband policy (e.g., a domestic macroeconomic policy change), would impact the firms located in cities with broadband policy and others in an identical fashion. Thus the net effects shown by the firms located in the treated cities over and above the firms in the control group (i.e., firms located in cities with no broadband policy in this context), would represent an effect attributable to factors specific only to the broadband policy.

Our coefficient of interest in Eqn. (1) is β. It measures relative the effect of the broadband policy on digitalization index, innovation, and performance of a firm located in a city treated with broadband policy relative to other firms. The underlying idea is that control group of firms potentially have the same characteristics as the treated firms, but they are not affected by the treatment. In summary, we expect that the firms' digitalization index and performance increase due to its location in cities where the broadband policy is implemented.

We also control for firm-specific time-invariant effects (ϕ_f) to control for other unobservable characteristics, city fixed effects (δ_c) to control for location-specific unobservable characteristics, and industry-year fixed effects (θ_{jt}) to control for time-varying changes at the industry level. We cluster our standard errors two-way at city and year level.

However, the basic estimates still may not provide conclusive evidence of the causal effect of the broadband policy because of the following two major reasons: (a) differential time trends; and (b) reverse causality. Below we will consider each of them separately and show that our results are robust. We address the problem of differential time trends for firms located in cities with the broadband policy and no broadband policy by showing that the digitalization index or other measures of technology adoption or firm performance between firms across these two types of cities are not significantly different before the policy was announced. In addition, we also show that no firm level feature, that could possibly be associated with the implementation of the broadband policy, did not influence the policy through a series of explicit exogeneity checks in the following section.

4.1 Dealing with the endogeneity of the Broadband Policy

We do a series of endogeneity checks in **Table 2**. In **Panel A** we show that firms located in cities where the broadband policy was implemented in 2014, 2015, and 2016 are not different in terms of their overall digitalization frequency, internet adoption, digital technology adoption, R&D expenses, and sales. In other words, we show that our treatment (firms in cities with broadband policy) and control (other firms) group were not on different time trends in the pre-policy period, i.e., for the years 2007–2013. We regress the overall digitalization, internet adoption, digital technology adoption index, R&D expenses, and sales on the interaction of year fixed effects for the years 2007–2013 with $City_{fc}$ using the following equation:

$$Ln(y_{fjct}) = \beta \left(\lambda_t \times City_{fc}\right) + \phi_f + \delta_c + \theta_{jt} + \epsilon_{fjt} \tag{2}$$

λ_t's denotes the year fixed effects from 2007 to 2013. The estimates from **Panel A** across columns (1) – (5) suggest that there is no differential time trend in any of the firm level indicators – neither in their technology adoption or performance. Interpreting it differently, the idea here is to introduce some counterfactual policies to see if they had any impact on firms' digitalization index or performance. As our estimates show, we find limited evidence of any consistent impact of the counterfactual policies. In addition, the estimates across the years in the pre-policy period switch their signs between positive and negative, thereby showing no consistent patterns.

Secondly, one of the most obvious points here to note is that the policy may target cities or firms which have low digitalization indices or low technology adoption or low innovation expenditure or performance in the intital period. Alternatively, big firms with high digitalization index may also influence the government to implement such policies in order to further increase their level of technology adoption to reap higher gains. Therefore, firms pre-located in cities with broadband policy could anticipate the announcement of the policy and can start adopting higher digital processes or technology and this can drive up their performance more than others.

To check whether such is the case or not, we regress $Broadband\ Policy_{2013} \times City_{fc}$ (our main variable of interest) on few of the important outcomes of interest such as overall digitalization index, internet adoption, digital technology, R&D expenses, and sales of a firm in **Panel B**. In effect, we run the following specification:[10]

$$\left(Broadband\ Policy_{2013} \times City_{fc}\right) = \pi X_{f,<2013} + \phi_f + \delta_c + \theta_{jt} + \epsilon_{fct} \tag{3}$$

$X_{f,<2013}$ is a vector of firm level characteristics that can possibly influence the implementation of the policy in city c. It includes overall digitalization index (a low or high digitalized firms may lobby for the policy in order to reap higher gains), internet or digital technology adoption (cities with low internet adoption may pressurize the Govt. for some policy to be implemented), R&D expenses (this captures whether more innovative firms are instrumental in driving the policy), and overall sales (firms with greater market power may also push for the policy in order to reap benefits from higher technology adoption). All the characteristics are used for the pre-2013 period. The idea here is to check whether any of these characteristic(s) at a previous period could influence the 2013 broadband policy undertaken by the China State Council. The coefficients indicate no statistical correlation between the complementary effect of the broadband policy and firm location with any of the firm characteristics. Combining all the above observations, we can conclude that the policy was exogenous to the prevailing conditions of the digitalization and

[10] We also use, not the interaction term, but only $Broadband\ Policy_{2013}$ as the outcome variable of interest; the results remain the same.

other key outcomes related to digitalization of Chinese firms located in cities with the broadband policy or not.

5 Results

5.1 First order effects – Digitalization

We start by investigating what happened to the overall digital index or frequency of firms in **Table 3**. Columns (1) and (2) regress the intensive and extensive margin of the digital frequency of firms on $Broadband\ Policy_{2013} \times City_{fc}$, controlling for firm, city, and interactions of industry-year fixed effects. In particular, column (1) uses the natural logarithm of the total frequency of the digital characteristics of a firm, and column (2) uses an indicator 0 or 1 if a firm adds a new digital process after 2013. Our estimates that show that the broadband policy implemented by the China State Council increased the digital frequency of a firm by 31% and the probability of using a new digital process after 2013 by 0.28 or 28%.

Columns (3) – (4) and (5) – (6) employ propensity score matching and nearest neighborhood matching methods. We compare firms based on two different types of characteristics – firm and city level. Columns (3) and (5) use firm level characteristics – age of a firm and value-added, whereas columns (4) and (6) use city level characteristics – GDP of a city and number of workers employed in R&D and IT industry. Our estimates across all these estimations are positive, significant, and higher than OLS estimates. Lastly, in column (7) we use an alternative method – kernel matching based on automatic bandwith selection. King & Nielsen (2019) argue that methods like propensity score matching approximates complete randomization. And, complete randomization analogy only works for observations with the same propensity score. Other methods such as Mahalanobis

Distance Matching (MDM) on which kernel matching procedure is based approximates fully blocked randomization. A fully blocked design is more efficient. It leads to less data imbalance and less "model dependence". We use this by including both firm and level characteristics to calculate the distance between a firm located in a city with broadband policy and another firm with no broadband policy. We find that the broadband policy continues to significantly influence digitalization index of firms.

Column (8) controls for two key firm level characteristics – age and size (value-added) of a firm in the regression. Both age and size is significantly correlated with the digitialization of a firm. In other words, mature and large firms have higher adoption of digital processes. We drop value-added and control for capital employed by a firm in column (9). We find that firms which have higher amount of capital employed also have higher digital adoption. Columns (10) and (11) use city level controls – GDP and number R&D and IT employees of a city, respectively. Both the variables are also significantly correlated with the digitalization frequency of a firm. All these results show that even after controlling for a series of factors, both at the firm and city level, our key variable of interest, $Broadband\ Policy_{2013} \times City_{fc}$, continues to be significant.

We check for some heterogeneity in columns (12) – (14). In column (12) we explore whether this increase in digitalization is significantly driven by firms which had higher digital frequency before the policy was implemented. We create a binary variable 0 or 1 to denote if a firm is highly digitized or not. If the overall digitalization frequency of a firm in the initial period, which is 2007, is greater than the median digital frequency of the corresponding sector, that particular firm is categorized as high-digitalized firm, $H - Digitalized_f$. We interact this indicator with $Broadband\ Policy_{2013} \times City_{fc}$ and check for the required effect. Our triple interaction term turns out to be negative and significant informing that the increase in digital frequencies after the broadband policy is driven by initial low digitalized firms. This is encouraging as the broadband policy led to a decrease and not increase in inequality in the digital processes of firms. The low digitalization firms had about 43% more increase in digital characteristics than high digitalized firms in the aftermath of the broadband policy in 2013.

Column (13) checks for spatial heterogeneity. We classify firms into three different regions – West, East, and Middle. We use "West" as the reference or excluded category as this is the interior most region of China. Our results show that the increase in digital frequency of firms is pervasive across all the regions of China; however, it is highest is case of firms located in the middle of China, followed by East, and the West. Firms in the western part of China had the lowest increase; in particular the effect is about one-fourth to one-fifth of the increase for middle and eastern part of China.

Lastly, in column (14) we check for industry level heterogeneity. We divide industries into – technology-intensive, capital-intensive, and labour-intensive. We use technology-intensive industries as the base category. Industries which has a higher average capital employed for the years 2007–2013 than the median of all the firms are classified as capital-intensive industries, $K - Intensive_j$. Similarly, industries which have a higher wages to capital employed ratio (average) than the median across all the firms before 2013 are termed as labour-intensive industries, $L - Intensive_j$. We interact these industry level binary variables with our key variable of interest, the double interaction term. Our estimates portray that firms belonging to the technology-intensive industry has the highest effect in terms of increase in the digital frequency followed by labour–intensive and capital-intensive. The effects are about 74–88% less than technology-intensive industries.

Although we use the interaction between industry-year fixed effects, our results could still possibly driven by the differences in the pre-trends between our treated (firms in cities with the broadband policy) and control (other firms) group. To control for such, we interact our treated dummy, $City_{fc}$, with year dummies that can possibly influence our results and plot the coefficients using the following regression equation:

$$Ln(y_{fjct}) = \left(\lambda_t \times City_{fc}\right) + \phi_f + \delta_c + \theta_{jt} + \epsilon_{fct} \tag{4}$$

where λ_t are the year dummies. **Figure 3** plots the year wise coefficients of the differences in digitalization frequencies between the firms located in the cities with broadband policy and other cities. Even when controlling for pre-trends, the differences are well-observed. The coefficient plot indicates that the difference between the firms located in the cities where broadband policy was implemented either in 2014 or 2015 or 2016 and the rest of the cities in terms of digital characteristics is statistically zero during the pre-policy period, i.e. on or before the China State Council announced the implementation of the policy in 2013. However, an examination of the post-policy period clearly indicates that the digital frequency rose differentially for firms located in cities with the broadband policy from 2014 onward. In particular, it took a sharp rise in the year following the policy and continues to increase further thereafter.

5.1.1 IV Results

Although we run a few endogeneity checks to investigate whether firm level attributes (such as location, size, innovation, etc.) during the period 2007–2012 influence the implementation of the "Broadband Policy" undertaken by the China State Council in 2013, a few potential concerns still remains. For example, the broadband policy may have predominantly reached economically attractive locations or locations that are predisposed to digitalization a priori or the policy was implemented in those places where the local leaders were more connected to the central governing system in China. In addition, as stated in the policy document, the construction of broadband in China should "make full use of the existing network infrastructure", which may mean that the choice of the cities may not have been completely exogenous. To address this, we employ instrumental variables estimation. We present our results in **Table 4**.

We use two types of instruments – (a) existing infrastructure which may possibly influence the choice of cities where the broadband policy was designed to be implemented. In particular, we use city level telephone users for the year 1984; (b) geographical characteristics, such as elevation and gradient of a city.

Our choice of the former instrument, which is the city level differences in the number of telephone users per 10,000 people in 1984 is driven by the following thought process: before the implementation of broadband internet, all firms with a telephone connection would have dial-up access to internet, albeit with a limited rate less than the broadband. Now, these firms with pre-existing dial-up connection could be the primary beneficiaries of the new broadband policy as certain level of network infrastructure which is needed for the implementation of broadband is already constructed. We follow DeStefano et al. (2018) and posit that our use of historical records of telephone infrastructure is driven by the fact that pre-existing telephone lines can significantly influence the implementation of broadband policy. We run a simple unconditional correlation between the cities where the broadband policy was implemented in 2014 or 2015 or 2016 and telephone users per 10,000 in 1984 (at the city level) and plot it in **Panel A** of **Figure 4**. Our plot shows a strictly positive and significant correlation between these two thereby satisfying the relevance of our instrument.

In addition to a strong first stage, we also need to satisfy the exclusion restriction. Namely, the historical telephone infrastructure should affect the digital frequency of firms located in those cities where the broadband policy was implemented, but after the implementation and not before. Therefore, we should not find any correlation between the digital frequency of those cities before the policy was implemented and our historical records of telephone users of those cities, which is our IV. We use the sum of digital frequencies across firms of those cities where the broadband policy was implemented before 2014 and plot a simple unconditional correlation in **Figure B.1**. The plot shows no pre-existing correlation. This alleviates any concern pertaining to common unobserved shocks across these cities based on pre-existing infrastructure levels.

Column (1) presents our 2SLS or IV results with the number of telephone users per 10,000 people in a city in 1984. The first stage shows a strong and significant positive correlation justifying that pre-existing or historical telephone infrastructure can act as a significant driver to the implementation of the current broadband infrastructure. Our second stage estimate continues to be positive and significant like our OLS coefficients.

We use another set of instruments, namely geographical characteristics following Kolko (2012). Geographical characteristics, such as topography of cities does not directly affect the digitalization of firms, but can possibly influence the implementation of the broadband in terms of say, laying out of the fibre cables or broadband infrastructure. For example, topographical undulation, such as elevation or gradient of cities can impact the development of digital infrastructure. The larger the degree of undulation or the elevation of the cities, the higher would be the construction costs which can impact the network infrastructure's operational efficiency and therefore possibly worse would be the signal quality of the broadband network. And, this could affect the choice of the cities in terms of where the broadband policy could be implemented. We run a simple unconditional correlation test between the cities which were treated with broadband policy and average elevation and gradient of cities in **Panels B** and **C** of **Figure 4**. The correlation plots confirm the selection and relevance of our instrument.

Columns (2) – (3) use average elevation and standard deviation of elevation of a city, respectively as the instruments. Likewise our conditional correlation graphs, our first stage appears to be negative – higher elevation of a city reduces the probability of the implementation of the broadband program in that city or the choice of the city to be included in the program. On the other hand, our 2SLS estimate is positive and significant. Columns (4) and (5) repeat the same exercise as columns (2) and (3), but using the gradient of a city instead of elevation. The results turn out to be the same – a strong and negative first stage with a positive second stage estimate.

5.1.2 Components of the Digitalization Index

Table 5 breaks down the overall digitalization index of a firm into 9 different sub-indices. We start by looking at internet technology in column (1). This index would contain if a firm mentions some of the following keywords in its annual report, for example internet mode, internet business, internet strategy, internet solutions, internet marketing, internet applications, internet platform, internet action, etc. Our estimate shows that the broadband policy has increased internet-related activities of a firm by about 22%. Column (2) substitutes internet with digital technology. If a firm mention words, such as digital terminal, digital communications, digital network, digital currency, digital intelligence, digital finance, digital marketing, etc. it would be counted as part of the digital technology index. A firm's digital participation increased by 15% in a city where the broadband policy was implemented.

Columns (3) – (9) use other sub-components of the overall digitalization index, such as smart and intelligence (for example, smart logistics, smart technology, smart marketing, smart warehousing, smart manufacturing, intelligent transportation, etc.), automatic technology (automatic control, automatic production, automatic monitoring, etc.), information (information integration, information management, information network, information sharing, etc.), big-data and cloud (big data, data mining, data science, data network, cloud computing, cloud platform, cloud services, etc.) AI and learning (intelligent robot, machine learning, deep learning, artificial intelligence, etc.), integrated technology (integrated solutions, integrated systems, integrated control, etc.) , and others (electronic commerce, fin-tech, block-chain, natural language processing, unmanned retail, etc.) as the outcomes of interest. Across all the sub-components of digitalization index, the effect is most pronounced for smart and intelligence (41%), followed by internet technology (22%), and big-data and cloud (20%).

5.2 Innovation Effects

Internet access is an essential determinant of innovation (Rampersad & Troshani, 2020); however, but the effect may be limited (Ford, 2018). Two mechanisms have been highlighted regarding the effect of broadband connection on innovation: (a) broadband favors innovation by lowering information costs. Xu et al. (2019) show that access to internet is a mechanism for lowering discovery or information costs, which can increase patent filings. In particular, they argue that if agents can learn at a low cost whether the innovation has been patented, how many others are competing in the patent race, who is and will be pursuing such protection, etc. they are more likely to pursue the project; and (b) broadband enhances innovation by easing collaborations. Agrawal & Goldfarb (2008) utilizing data for US universities find that decrease in collaboration costs resulting from adopting Bitnet (an early version of internet technology) seem to have facilitated a general increase in multi-institutional collaboration. Similarly, Ding et al. (2010) find that the availability of Bitnet on a university's campus not only provided greater access to materials and equipment to the scientists but also allowed researchers to share ideas at a lower rate and across greater distances, which had a positive effect on their productivity and collaborative network.

We follow this literature and check for the effects of the broadband policy on different factors related to research and development. **Table 6** presents the results. We start by looking at research and development expenses in column (1). We find that firms located in cities with the broadband policy increased their R&D expenditure by 27%. Columns (2) and (3) substitutes the R&D expenses by another productive factor required to undertake innovation – employment of R&D workers. Column (2) use share of R&D workers and column (3) the average wage of R&D workers as the outcome variables of interest. We find positive effects on both counts: share of R&D workers increased by 8% while wage went up by 47%. Lastly, we focus on the output side of innovation in column (4), which resembles patents filed by a firm. We find that patent filings also shot up by 22%. Our results are similar to findings in case of US (Xu et al., 2019), Germany (Bertschek et al., 2013), Australia (Rampersad & Troshani, 2020), etc.

5.3 Firm Performance

5.3.1 Labour Demand

Hjort & Poulsen (2019) show that introduction of fast interest resulted in higher net job creation. This is driven by increase in employment in higher-skill occupations, but less-educated workers also gained employment, but less. Similar outcomes have been noted in case of the US, especially in the rural areas (Atasoy, 2013). On the other hand, Kolko (2012) show that county level broadband expansion does not increase overall wages or improve employment opportunities in the US.

We follow this literature and explore the effects on labour demand of a firm. One unique feature of the firm level dataset that we use is that it not only gives detailed information on the total number of employees and total wages, it also provides detailed information on senior managers. We use these different categories to check the compositional effect on different types of workers. **Panel A** of **Table 7** presents the required result.

Our findings show that the 2013 broadband policy result in (a) significant increase in employment of a firm – both the total number of workers (extensive margin) and their average wage (intensive margin) increased. In particular, a firm employed 12% more workers and paid 16% higher average wages (columns (1) and (2)); (b) price of senior managers went by 12% with no effect on their employment or on the extensive margin (columns (3) and (4)). Our results are similar to what Akerman et al. (2015) find for Norway.

5.3.2 Firm Performance

All these changes in terms of employment can have a significant effect on the overall performance of a firm. DeStefano et al. (2018) find that for UK firms higher use of ICT (internet and communications technology) led to increase in their sales, but not productivity. Reports on digitalization by EBRD (2022) and World Bank (2021) highlight that increase in digitalization across Europe and South Asia led to increase in performance for firms, regions, etc.

Panel B of **Table 7** starts by looking at what happened to what factor of production, capital employed in column (5). Like labour, our estimates show that the policy led to a significant increase in capital by 21%. We then substitute capital employed by value-added (proxy for productivity) in column (6), sales in column (7), and stock market valuation in column (8). Our results show positive effects on all counts – value-added and sales by around 15% and stock market valuation by 16%. Our results echo for Hvide et al. (2022) for Norway. Their results show that internet use causes a substantial increase in stock market participation (driven primarily by increased fund ownership) and firm performance.

5.4 Firm Characteristics

Lastly, we check how firm characteristics play a role in driving our main findings. We focus on two main characteristics – size and ownership. **Table 8** presents our results using three key variables: total digitalization index, R&D expenses, and total sales as the outcome variables.

Panel A divides firms into three different terciles based on their initial total assets and estimates the effect for each tercile on digitalization, R&D expenditure, and sales. A firm belongs to the 1st tercile if the assets (initial) of a firm is less than 33% of the total assets of the corresponding industry, and so on. In particular, a firm in 2nd tercile belongs to 34-67th percentile. Our results show that firms across terciles have a positive effect of the broadband policy, with certain amount of heterogeneity. The biggest effect of the broadband policy for each of the outcome variables is for the second tercile of firms. These firms, which are marginally big, got the biggest push in terms of upgrading their digital processes (it went up by 36% compared to 27% for 1st tercile and 33% for 3rd tercile), increase in R&D expenses (increased by 31% compared to 23% for 1st tercile and 26% for 3rd tercile), and increase in sales (went up by 19% compared to 14% of 3rd tercile and no increase for 1st tercile).

We divide firms into three categories of ownership in **Panel B** – private, SoEs (state-owned enterprises), and foreign firms. We find that the overall effect for digitalization and R&D expenditure is entirely driven by domestic private (36% and 31%) and foreign (31% and 23%) firms with former having the higher effect. However, in case of total sales, foreign firms register the biggest effect.

6 Concluding Remarks

Do broadband policy affects digital adoption, innovation, and overall performance of firms? We investigate this question by exploiting a quasi-experimental policy adopted by China in 2013. This particular policy was a pilot program of a large future broadband infrastructure development program. Utilizing this pilot program, we find that the policy led to significant increase in digitalization of firms' processes or digital adoption. This is significantly driven by adoption of internet and digital technology, smart and intelligence processes, and big-data and cloud. The policy also led to an increase in both employment in innovation inputs and innovation output for firms. All these changes led to increase in firm size (employment and sales), productivity, and higher performance in the stock market.

With rapid development of the digital economy, digital transformation of firms is of the first order importance. On the other hand, digital adoption is one of the crucial constraints that firms in developing countries still face. And, governments should play a key role in this aspect by providing new infrastructure for firms as broadband infrastructure and widespread use of digital technologies have the potential to enhance all aspects of an economy and society. Overall, digital adoption is an important driver of global economic integration and can provide significant opportunities for developing and emerging countries. Our results also suggest that policy makers may want to consider the potential heterogeneous effects on various economic outcomes.

References

Agrawal, A. & Goldfarb, A. (2008). Restructuring research: Communication costs and the democratization of university innovation. *American Economic Review*, 98(4), 1578–1590.

Akerman, A., Gaarder, I., & Mogstad, M. (2015). The skill complementarity of broadband internet. *The Quarterly Journal of Economics*, 130(4), 1781–1824.

Allen, T. (2014). Information frictions in trade. *Econometrica*, 82(6), 2041–2083.

Atasoy, H. (2013). The effects of broadband internet expansion on labor market outcomes. *ILR Review*, 66(2), 315–345.

Bahia, K., Castells, P., Cruz, G., Masaki, T., Rodríguez-Castelán, C., & Sanfelice, V. (2023). Mobile broadband, poverty, and labor outcomes in Tanzania. *The World Bank Economic Review*, 37(2), 235–256.

Bertschek, I., Cerquera, D., & Klein, G. J. (2013). More bits – more bucks? Measuring the impact of broadband internet on firm performance. *Information Economics and Policy*, 25(3), 190–203.

Bhuller, M., Ferraro, D., Kostøl, A., & Vigtel, T. (2023). The internet, search frictions and aggregate unemployment. *NBER Working Paper No. 30911*, NBER, Massachusetts, USA.

Bianchi, N., Lu, Y., & Song, H. (2022). The effect of computer-assisted learning on students' long-term development. *Journal of Development Economics*, 158, 102919.

Cariolle, J., Le Goff, M., & Santoni, O. (2019). Digital vulnerability and performance of firms in developing countries. *Banque de France Working Paper No. 709*, Banque de France, Paris, France.

Chan, K. S., Dang, V. Q., & Yan, I. K. (2012). Chinese firms' political connection, ownership, and financing constraints. *Economic Letters*, 115(2), 164–167.

Chen, S., Liu, W., & Song, H. (2019). Broadband internet, firm performance, and worker welfare: Evidence and mechanism. *Economic Inquiry*, 58(3), 1146–1166.

Chen, S., Sun, Z., Tang, S., & Wu, D. (2011). Government intervention and investment efficiency: Evidence from China. *Journal of Corporate Finance*, 17(2), 259–271.

Chun, N. & Tang, H. (2018). Do information and communication technologies empower female workers? Firm-level evidence from Vietnam. *ADB Economics Working Paper Series No. 545*, Asian Development Bank, Manila, Philippines.

Commander, S., Harrison, R., & Menezes-Filho, N. (2011). ICT and productivity in developing countries: New firm-level evidence from Brazil and India. *The Review of Economics and Statistics*, 93(2), 528–541.

Czernich, N. (2014). Does broadband internet reduce the unemployment rate? Evidence for Germany. *Information Economics and Policy*, 29, 32–45.

Demir, B, T., Javorcik, B., & Panigrahi, P. (2023). Breaking invisible barriers: Does fast internet improve access to input markets? *Mimeo*, (pp. University of Oxford).

DeStefano, T., Kneller, R., & Timmis, J. (2018). Broadband infrastructure, ICT use and firm performance: Evidence for UK firms. *Journal of Economic Behavior and Organization*, 155, 110–139.

Ding, W. W., Levin, S. G., Stephan, P. E., & Winkler, A. E. (2010). The impact of information technology on academic scientists' productivity and collaboration patterns. *Management Science*, 56(9), 1439–1461.

Du, M. & Boateng, A. (2015). State ownership, institutional effects and value creation in cross-border mergers & acquisitions by Chinese firms. *International Business Review*, 24(3), 430–442.

EBRD (2022). EBRD transition report 2021-22: System upgrade – Delivering the digital dividend. *EBRD Transition Report 2021-22*.

Fabling, R. & Grimes, A. (2021). Picking up speed: Does ultrafast broadband increase firm productivity? *Information Economics and Policy*, 57, 100937.

Ford, G. S. (2018). Is faster better? Quantifying the relationship between broadband speed and economic growth. *Telecommunications Policy*, 42(9), 766–777.

Gokan, T., Kichko, S., & Thisse, J.-F. (2019). How do trade and communication costs shape the spatial organization of firms? *Journal of Urban Economics*, 113, 103191.

Haller, S. & Lyons, S. (2013). Broadband adoption and firm productivity: Evidence from Irish manufacturing firms. *Telecommunications Policy*, 39(1), 1–13.

Hjort, I. & Poulsen, J. (2019). The arrival of fast internet and employment in Africa. *American Economic Review*, 109(3), 1032–1079.

Houngbonon, G. V., Mensah, J. T., & Traore, N. (2021). The impact of internet access on innovation and entrepreneurship in Africa. *World Bank Policy Research Working Paper No. 9945*, World Bank, Washington DC, USA.

Hvide, H. K., Meling, T. G., Mogstad, M., & Vestad, O. L. (2022). Broadband internet and the stock market investments of individual investors. *NBER Working Paper No 30383*, NBER, Massachusetts, USA.

King, G. & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27(4), 435–454.

Kolko, J. (2012). Broadband and local growth. *Journal of Urban Economics*, 71(1), 100–113.

Lin, T. (2019). Division of labor and productivity advantage of cities: Theory and evidence from Brazil. *CEPR Discussion Paper No. 16590*, CEPR, London, UK.

Malamud, O., Cueto, S., Cristia, J., & Beuerman, D. W. (2019). Do children benefit from internet access? Experimental evidence from Peru. *Journal of Development Economics*, 138, 41–56.

Malamud, O. & Pop-Eleches, C. (2011). Home computer use and the development of human capital. *The Quarterly Journal of Economics*, 126(2), 987–1027.

Malgouyres, C., Mayer, T., & Mazet-Sonilhac, C. (2021). Technology-induced trade shocks? Evidence from broadband expansion in France. *Journal of International Economics*, 133(103520).

Rampersad, G. & Troshani, I. (2020). Impact of high-speed broadband on innovation in rural firms. *Information Technology for Development*, 26(1), 89–107.

Wang, B., Wang, L., Gong, B., Yan, Z., & Hu, P. (2023). Does broadband internet infrastructure mitigate firm-level economic uncertainty? Evidence from Broadband China Pilot Policy. *Economics Letters*, 232(111325).

World Bank (2021). Shifting gears: Digitization and services-led development. *World Bank South Asia Focus*.

Xu, X., Watts, A., & Reed, M. (2019). Does access to internet promote innovation? A look at the U.S. broadband industry. *Growth and Change*, 50(4), 1423–1440.

Zhou, F., Wen, H., & Lee, C.-C. (2022). Broadband infrastructure and export grwoth. *Telecommunications Policy*, 46(102347).

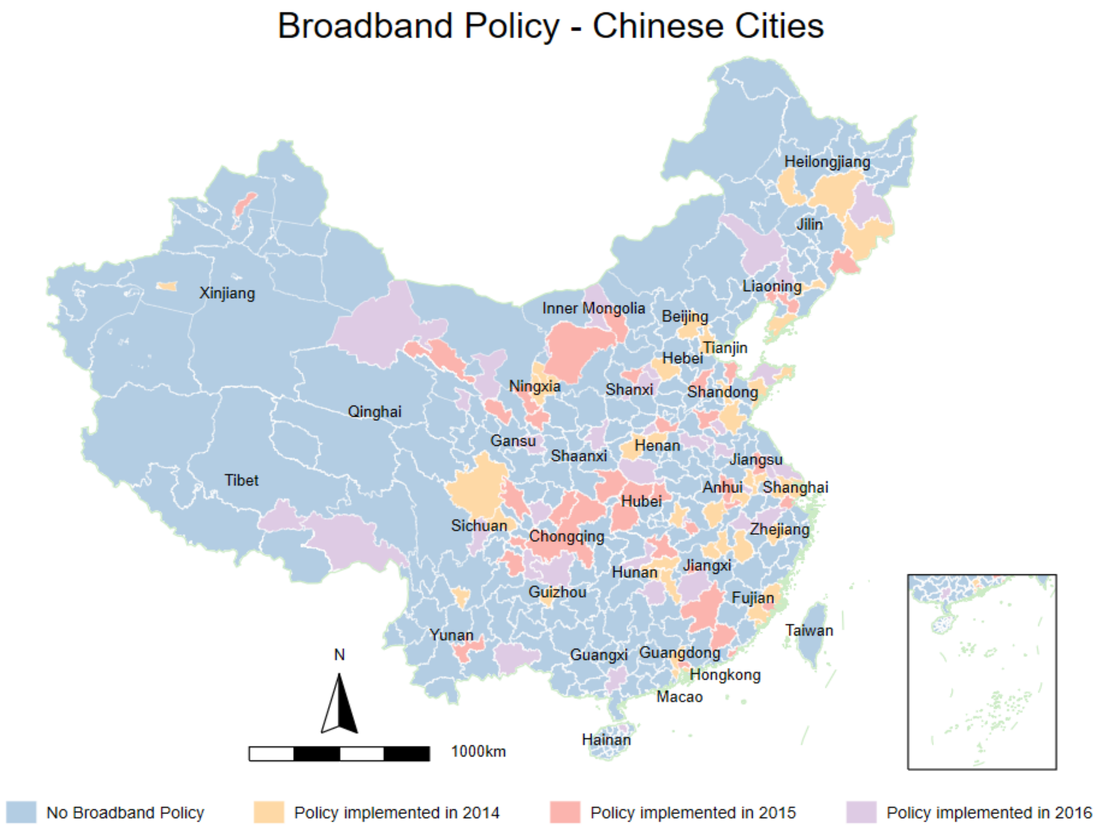

Figure 1: Broadband Policy – Chinese Cities

Notes: Figure presents the Chinese cities where the broadband policy was implemented in 2014, 2015, and 2016, respectively. Due to size constraints, we only marked the provinces on the map.

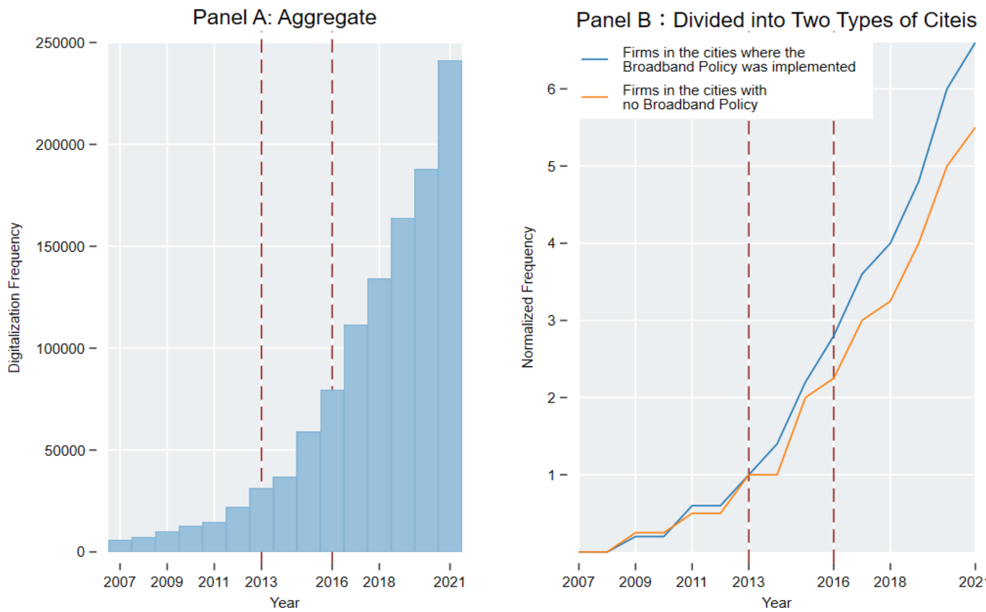

Figure 2: Digitalization Index: Aggregate and Divided into Cities

Notes: Panel A represents the overall trend of the digitalization frequency with 144 variables. The vertical red lines for the year 2013 and 2016 represent starting and closing year of the Broadband Policy, respectively. China State Council announced the "Broadband China Strategy and Implementation Plan" in 2013. Three batches (39 cities in 2014, 39 cities in 2015, and 39 cities in 2016) of total 117 cities were chosen to be a part of the agenda. Panel B presents the normalized median trend of the overall digitalization frequency divided into two types of cities – blue line represents the firms in the cities where the broadband policy was implemented and orange line represents the firms in the cities which were outside the purview of the broadband policy.

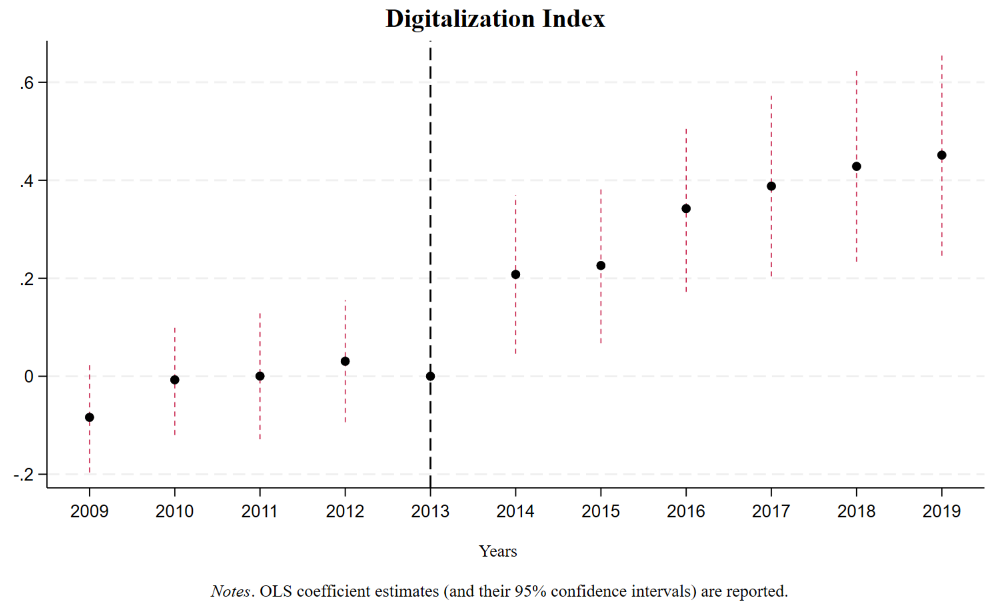

Figure 3: Event Study Framework – Overall Digitalization Index

Notes: Figure presents the differences in the overall digitalization index between firms in the cities where the broadband policy was implemented and all other cities. All the coefficients are plotted with respect to the year 2013 when the Broadband Policy was announced. Our coefficient estimates are controlled for industry-year and city fixed effects and standard errors clustered at the year-city level.

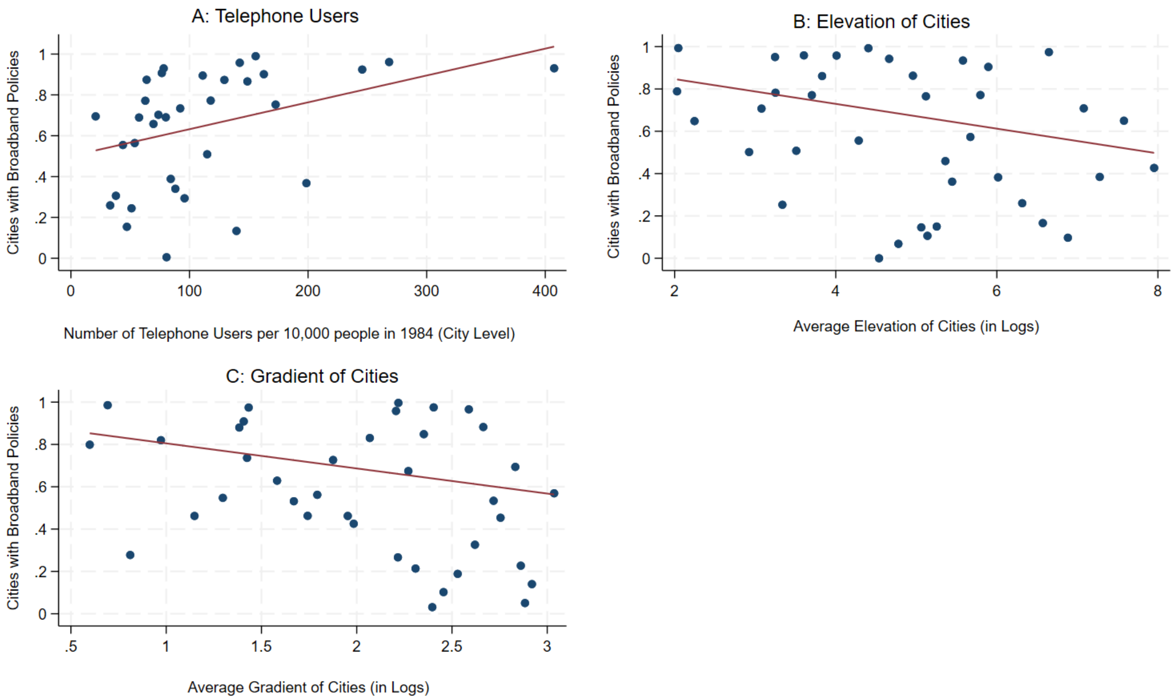

Figure 4: Correlation: Broadband Cities and IVs

Notes: Panel A of the Figure presents unconditional correlation between cities which were selected for the implementation of the broadband policy in 2014 or 2015 or 2016 and the number of telephone users per 10,000 in 1984. Panel B of the Figure presents unconditional correlation between cities which were selected for the implementation of the broadband policy in 2014 or 2015 or 2016 and average elevation of the cities. Panel C of the Figure presents unconditional correlation between cities which were selected for the implementation of the broadband policy in 2014 or 2015 or 2016 and average gradient of the cities.

Table 1: Digitalization Index

	(1)	(2)
	2007–2013	2014–2021
Panel A: Aggregate		
$\text{Digitalization}_{Total}$	4	20
Panel B: Divided into 2 Types of Cities		
$\text{Digitalization}_{Broadband\ Cities}$	5	23
$\text{Digitalization}_{No\ Broadband\ Cities}$	3	15

Notes: The digitization index is based on the lexical frequencies of 144 internet related variables mentioned by a firm in its annual reports. Numbers represent median values of the lexical frequencies. It is a combination of: (1) internet related activities, such as internet mobile, internet mode, internet platform, internet strategy, internet marketing, etc; (2) digital technology related variables such as digital marketing, digital technology, digital network, digital finance, etc; (3) smart and intelligent technology related variables such as intelligent robots, smart logistics, smart factory, smart grid, digital intelligence, etc; (4) automatic technology related variables such as automatic production, automatic control, automatic face recognition, etc; (5) information technology related variables such as information management, information network, information software, etc; (6) big-data and cloud related variables, such as big data, data science, could platform, cloud services, data visualization, etc; (7) AI (Artificial Intelligence) and learning related variables such as artificial intelligence, intelligent robot, machine learning, etc; (8) integrated technology related variables such as integrated solutions, integrated system, etc; (9) all other types of technology related variables such as electronic commerce, financial technology, quantitative finance, block chain analysis, e-commerce, etc by a firm.

Table 2: Differences in Pre-Broadband Policy Time Trends and Endogeneity Checks

	(1)	(2)	(3)	(4)	(5)
Panel A: Differences in Pre-Trends					
	Total Digitalization Index	Internet Adoption	Digital Technology	R&D Expenses	Total Sales
$D_{2007} \times City_{fc}$	0.174	–0.122	–0.131	–0.494	0.003
	(0.166)	(0.119)	(0.123)	(0.277)	(0.206)
$D_{2008} \times City_{fc}$	0.088	–0.184	0.091	0.258	0.212
	(0.145)	(0.099)	(0.074)	(0.312)	(0.158)
$D_{2009} \times City_{fc}$	–0.012	0.047	–0.115	–0.217	0.364
	(0.180)	(0.113)	(0.094)	(0.303)	(0.246)
$D_{2010} \times City_{fc}$	–0.013	0.096	–0.014	0.042	0.068
	(0.150)	(0.091)	(0.045)	(0.213)	(0.138)
$D_{2011} \times City_{fc}$	–0.019	–0.036	–0.059	–0.029	–0.145
	(0.117)	(0.108)	(0.070)	(0.209)	(0.108)
$D_{2012} \times City_{fc}$	–0.016	0.017	0.005	–0.257	–0.062
	(0.106)	(0.094)	(0.056)	(0.180)	(0.087)
$D_{2013} \times City_{fc}$	0.159	0.006	–0.079	0.163	0.073
	(0.098)	(0.086)	(0.056)	(0.213)	(0.062)
R-Square	0.96	0.84	0.77	0.89	0.89
N	29,341	29,341	29,341	18,308	28,715
Panel B: Endogeneity Checks					
	$Broadband_{2013} \times City_{fs}$				
Total Digitalization Index$_{ft-1}$	–0.002				
	(0.002)				
Internet Adoption$_{ft-1}$		0.001			
		(0.001)			
Digital Technology$_{ft-1}$			–0.0004		
			(0.001)		
R&D Expenses$_{ft-1}$				0.00004	
				(0.002)	
Total Sales$_{ft-1}$					–0.001
					(0.002)
R-Square	0.89	0.89	0.89	0.89	0.89
N	19,460	19,460	19,460	14,355	19,047
Firm FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Industry (2-digit) × Year FE	Yes	Yes	Yes	Yes	Yes

Notes: In Panel A, columns (1) – (5) use logarithm of lexical frequency of total digitalization index, frequency of internet adoption index, frequency of adoption of digital technology, R&D expenses, and total sales of a firm as the dependent variables, respectively. D_{2007}, D_{2008}, D_{2009}, D_{2010}, D_{2011}, D_{2012}, D_{2013} are year dummies. These dummies equal to 1 for the respective years. In Panel B, columns (1) – (5) use $Broadband_{2013} \times City_s$ as the dependent variable. $Broadband_{2013}$ is a time dummy which takes a value 1 if year is greater than 2013. $City_{fc}$ is a spatial dummy – it takes a value 1 if a firm f is located in city c where the broadband policy was implemented between 2014 and 2016. Numbers in the parentheses are two-way clustered standard errors at the year and city level. Intercepts are not reported. *,**,*** denotes 10%, 5%, and 1% level of significance, respectively.

Table 3: Broadband Policy and Digital Index of Firms: Benchmark Results

	Digitalization Index						
	Intensive Margin	Extensive Margin	Propensity Score Matching		Nearest Neighborhood Matching		Kernel Matching
			Firm	City	Firm	City	Firm + City
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Broadband_{2013} \times City_{fc}$	0.308***	0.279***	0.935***	0.597***	0.859***	0.458***	0.962***
	(0.077)	(0.042)	(0.018)	(0.044)	(0.019)	(0.064)	(0.022)
R-Squared	0.78	0.68	n/a	n/a	n/a	n/a	n/a
N	26,552	26,551	25,758	20,585	25,758	20,585	
	Firm Controls		City Controls		Heterogeneity		
					Digitized Firms	Spatial	Industry
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
$Broadband_{2013} \times City_{fc}$	0.193***	0.195***	0.221***	0.254***	0.510***	0.119*	0.546***
	(0.059)	(0.063)	(0.068)	(0.068)	(0.104)	(0.066)	(0.134)
Ln(Age)$_f$	0.322***	0.362***					
	(0.042)	(0.046)					
Ln(Value-added)$_f$	0.191***						
	(0.011)						
Ln(Capital Employed)$_f$		0.006***					
		(0.0004)					
Ln(City GDP)$_c$			0.805***				
			(0.129)				
Ln(R&D + IT Employees)$_c$				0.071**			
				(0.036)			
$Broadband_{2013} \times City_{fc} \times H - Digitalized_f$					–0.289**		
					(0.125)		
$Broadband_{2013} \times East_{fc}$						0.477***	
						(0.113)	
$Broadband_{2013} \times Middle_{fc}$						0.515***	
						(0.126)	
$Broadband_{2013} \times City_{fc} \times K - Intensive_j$							–0.479***
							(0.130)
$Broadband_{2013} \times City_{fc} \times L - Intensive_j$							–0.402***
							(0.095)
R-Squared	0.80	0.79	0.79	0.78	0.79	0.79	0.80
N	25,372	26,323	23,456	23,056	26,552	26,551	26,551
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry (2-digit) × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Columns (1) – (7) use overall digitalization index of a firm. This index is based on lexical frequencies of different indices mentioned by firms in their annual reports. It consists of 144 internet related variables. Column (1) uses log value of the sum of lexical frequencies of the overall digitalization index; column (2) uses a indicator value 0 or 1. It takes a value 1 if a firm's digitalization index after 2013 is greater than the mean digitalization index across all firms on or before 2013. In particular, a firm f located in city c (where the broadband policy was implemented) will assume the value 1 if it's digitalization index is greater than 9.573611. $Broadband_{2013}$ is a time dummy which takes a value 1 if year is greater than 2013. $City_c$ is a spatial dummy – it takes a value 1 if a firm f is located in city c where the broadband policy was implemented between 2014 and 2016. Age is the age of a firm. Value-added is the gross value-added of a firm defined as sales minus total raw material expenditure. Capital employed is the total amount of capital employed by a firm. City GDP is the gross domestic product of a city. R&D + IT Employees is the total number of employees at city level which are engaged in scientific research and comprehensive technology service industry. $H - Digitalized_f$ is a firm level indicator which takes a value 1 if the digitalization index of a firm is greater than the median digitalization index of the corresponding industry. $East_s$, $Middle_s$ takes a value 1 if a firm is located in a city which is geographically in the east and middle of China. $K - Intensive_j$ and $L - Intensive_j$ are industry level indicators which takes a value 1 if an industry's average capital intensity (amount of capital employed) and labour intensity (labour cost/capital employed) before 2013 is greater than the median capital and labour intensity across all industries. Numbers in the parentheses are two-way clustered standard errors at the year and city level. Intercepts are not reported. ***, **, * denotes statistical significance at 1%, 5%, and 10%.

Table 4: Broadband Policy and Digital Index of Firms: IV Results

	Digitalization Index				
	(1)	(2)	(3)	(4)	(5)
$Broadband_{2013} \times City_{fc}$	1.344***	2.531**	2.320**	2.667**	2.341**
	(0.295)	(1.018)	(0.966)	(1.294)	(1.015)
R-Square	0.71	0.87	0.82	0.85	0.82
N	25,131	26,552	26,551	26,552	26,551
1st-Stage Estimates					
	$Broadband_{2013} \times City_{fc}$				
$Broadband_{2013} \times No.\ of\ Telephones_{1984,c}$	0.135***				
	(0.005)				
$Broadband_{2013} \times Elevation_c$		–0.058***			
		(0.002)			
$Broadband_{2013} \times SD\ Elevation_c$			–0.055***		
			(0.002)		
$Broadband_{2013} \times Urban\ Gradient_c$				–0.126***	
				(0.004)	
$Broadband_{2013} \times SD\ Urban\ Gradient_c$					–0.131***
					(0.005)
F-Stat (Kleibergen-Paap)	184.02	140.80	170.55	122.04	122.34
Firm FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Industry (2-digit) × Year FE	Yes	Yes	Yes	Yes	Yes

Notes: Columns (1) – (5) use logarithm of lexical frequency of total digitalization index of a firm as the dependent variable. $Broadband_{2013}$ is a time dummy which takes a value 1 if year is greater than 2013. $City_c$ is a spatial dummy – it takes a value 1 if a firm f is located in city c where the broadband policy was implemented between 2014 and 2016. $No.\ of\ Telephones_{1984,c}$ is the number of telephones per 10,000 people in city c in 1984. $Elevation_c$ and $SD\ Elevation_c$ are average elevation and standard deviation of elevation of a city c. $Gradient_c$ and $SD\ Gradient_c$ are average slope/gradient and standard deviation of slope/gradient of a city c. Numbers in the parentheses are two-way clustered standard errors at the year and city level. Intercepts are not reported. *,**,*** denotes 10%, 5%, and 1% level of significance, respectively.

Table 5: Broadband Policy and Digital Index of Firms: Effects on Other Types of Digitalization Indices

	Internet Technology	Digital Technology	Smart and Intelligence	Automatic Technology	Information
	(1)	(2)	(3)	(4)	(5)
$Broadband_{2013} \times City_{fc}$	0.221***	0.153**	0.408***	0.074***	0.037
	(0.053)	(0.056)	(0.091)	(0.024)	(0.045)
R-Squared	0.73	0.66	0.76	0.60	0.66
N	30,538	30,538	30,538	30,538	30,538
	Big-Data and Cloud	AI and Learning	Integrated Technology	Others	
	(6)	(7)	(8)	(9)	
$Broadband_{2013} \times City_{fc}$	0.201***	0.100***	0.159***	0.219***	
	(0.052)	(0.037)	(0.045)	(0.044)	
R-Squared	0.76	0.65	0.65	0.71	
N	30,538	30,538	30,538	28,700	
Firm FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Industry (2-digit) $\times$ Year FE	Yes	Yes	Yes	Yes	Yes

Notes: Columns (1) – (9) use different indices of digitalization of a firm. These indices are based on lexical frequencies of these different indices mentioned by firms in their annual reports. It consists of 144 internet related variables. Column (1) uses adoption of internet related activities, such as internet mobile, internet mode, internet platform, internet strategy, internet marketing, etc. by a firm. It consists of 21 internet related variables. Column (2) uses adoption of digital technology related variables such as digital marketing, digital technology, digital network, digital finance, etc. by a firm. It consists of 10 related variables. Column (3) uses adoption of smart and intelligent technology related variables such as intelligent robots, smart logistics, smart factory, smart grid, digital intelligence, etc. by a firm. It consists of 33 related variables. Column (4) uses adoption of automatic technology related variables such as automatic production, automatic control, automatic face recognition, etc. by a firm. It consists of 8 related variables. Column (5) uses adoption of information technology related variables such as information management, information network, information software, etc. by a firm. It consists of 12 related variables. Column (6) uses adoption of big-data and cloud related variables, such as big data, data science, could platform, cloud services, data visualization, etc. by a firm. It consists of 16 related variables. Column (7) uses adoption of AI and learning related variables such as artificial intelligence, intelligent robot, machine learning, etc. by a firm. It consists of 5 related variables. Column (8) uses adoption of integrated technology related variables such as integrated solutions, integrated system, etc. by a firm. It consists of 6 related variables. Column (9) uses all other types of technology related variables such as electronic commerce, financial technology, quantitative finance, block chain analysis, e-commerce, etc by a firm. It consists of 38 related variables. All these indices are used as log values of the sum of lexical frequencies across these indicators. $Broadband_{2013}$ is a time dummy which takes a value 1 if year is greater than 2013. $City_c$ is a spatial dummy – it takes a value 1 if a firm f is located in city c where the broadband policy was implemented between 2014 and 2016. Numbers in the parentheses are two-way clustered standard errors at the year and city level. Intercepts are not reported. ***, **, * denotes statistical significance at 1%, 5%, and 10%.

Table 6: Broadband Policy and Innovation Effects

	Factors of Innovation			
	Inputs			Output
	R&D Expenditure	Share of R&D Workers	Avg. Wage of R&D Workers	Patents Count
	(1)	(2)	(3)	(4)
$Broadband_{2013} \times City_{fc}$	0.274**	0.079***	0.473***	0.217*
	(0.103)	(0.015)	(0.124)	(0.103)
R-Square	0.830	0.118	0.671	0.778
N	19,245	30,538	30,538	9,071
Firm FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Industry (2-digit) $\times$ Year FE	Yes	Yes	Yes	Yes

Notes: Column (1) uses amount of R&D expenditure of a firm; columns (2) and (3) use share of R&D workers, and average wage of an R&D worker; column (4) uses the number of patents filed by a firm as the dependent variable, respectively. $Broadband_{2013}$ is a time dummy which takes a value 1 if year is greater than 2013. $City_c$ is a spatial dummy – it takes a value 1 if a firm f is located in city c where the broadband policy was implemented between 2014 and 2016. Numbers in the parentheses are two-way clustered standard errors at the year and city level. Intercepts are not reported. ***, **, * denotes statistical significance at 1%, 5%, and 10%.

Table 7: Broadband Policy and Firm Performance

	Total Employees		Managers	
Panel A: Labour				
	Number	Avg. Wage	Share	Avg. Wage
	(1)	(2)	(3)	(4)
$Broadband_{2013} \times City_{fc}$	0.119**	0.160***	0.002	0.121**
	(0.045)	(0.044)	(0.004)	(0.049)
R-Square	0.865	0.642	0.613	0.771
N	30,475	30,322	30,538	30,516
Panel B: Other Measures				
	Capital Employed	Value Added	Total Sales	Stock Market Value
	(5)	(6)	(7)	(8)
$Broadband_{2013} \times City_{fc}$	0.208***	0.146***	0.154**	0.160**
	(0.052)	(0.063)	(0.059))	(0.059)
R-Square	0.872	0.809	0.853	0.821
N	21,285	29,419	29,917	30,069
Firm FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Industry (2-digit) × Year FE	Yes	Yes	Yes	Yes

Notes: Columns (1) – (2) use total number of employees, and average wage of an employee in a firm; columns (3) – (4) use share of managers and average wage of a managerial worker; columns (5) – (8) use capital employed (= total asset – current liability), value-added (= sales – operating cost), total sales, and stock market value of a firm, respectively as the dependent variable. $Broadband_{2013}$ is a time dummy which takes a value 1 if year is greater than 2013. $City_c$ is a spatial dummy – it takes a value 1 if a firm f is located in city c where the broadband policy was implemented between 2014 and 2016. Numbers in the parentheses are two-way clustered standard errors at the year and city level. Intercepts are not reported. ***, **, * denotes statistical significance at 1%, 5%, and 10%.

Table 8: Broadband Policy and Role of Firm Characteristics

	Total Digitalization			R&D Expenses			Total Sales		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Size									
	1st Tercile	2nd Tercile	3rd Tercile	1st Tercile	2nd Tercile	3rd Tercile	1st Tercile	2nd Tercile	3rd Tercile
$Broadband_{2013} \times City_{fc}$	0.274***	0.362***	0.328***	0.233**	0.307**	0.264**	0.028	0.193**	0.135*
	(0.096)	(0.106)	(0.103)	(0.113)	(0.123)	(0.131)	(0.083)	(0.089)	(0.078)
R-Square	0.953	0.962	0.953	0.998	0.998	0.998	0.998	0.998	0.998
N	9,904	10,022	10,078	6,070	6,733	6,396	9,828	10,013	10,072
Panel B: Ownership									
	Private	SoEs	Foreign	Private	SoEs	Foreign	Private	SoEs	Foreign
$Broadband_{2013} \times City_{fc}$	0.355***	0.352	0.311***	0.306**	0.110	0.230**	0.074*	0.202	0.248***
	(0.098)	(0.229)	(0.088)	(0.149)	(0.149)	(0.094)	(0.040)	(0.181)	(0.083)
R-Square	0.936	0.974	0.967	0.998	0.998	0.998	0.998	0.998	0.998
N	12,380	2,430	15,241	6,339	1,638	10,987	11,909	2,375	15,158
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry (2-digit) × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Columns (1) – (3) use logarithm of lexical frequency of total digitalization index of a firm; columns (4) – (6) use logarithm of research and development expenses of a firm; columns (7) – (9) use logarithm of total sales of a firm, respectively as the dependent variable. $Broadband_{2013}$ is a time dummy which takes a value 1 if year is greater than 2013. $City_c$ is a spatial dummy – it takes a value 1 if a firm f is located in city c where the broadband policy was implemented between 2014 and 2016. Numbers in the parentheses are two-way clustered standard errors at the year and city level. Intercepts are not reported. ***, **, * denotes statistical significance at 1%, 5%, and 10%.

Appendix for Online Publication

A Tables

Table A.1: Key Words Related to Lexical Frequency

Internet Technology	Digital Technology	Smart and Intelligence Technology	Automatic Technology	Information
Industrial Internet	Digital Control	Smart Environmental Protection	Automatic Control	Industrial Information
Internet Mode	Digital Terminal	Smart Healthcare	Automatic Monitored	Information Integration
Internet Business	Digital Communications	Smart Technology	Automatic Production	Information Software
Internet Platform	Digital Currency	Smart Mobility	Automatic Monitoring	Information Management
Internet Medical	Digital Network	Smart Marketing		Information Network
Mobile Internet	Digital Marketing	Smart Grid		Information Terminal
Internet Business Mode	Digital Finance	Smart Manufacturing		Information Sharing
Mobile Internet	Digital Intelligence	Smart Factory		
Internet Ecology		Industrial Intelligence		
Internet Solutions		High End Intelligence		
Internet Strategy		Intelligent Robot		
Internet Action		Intelligent Production		
Internet Marketing		Mobile Intelligence		
Internet Application		Intelligent Customer Service		
		Smart Warehousing		
		Business Intelligence		
		Intelligent Transportation		
		Intelligence Control		
		Smart Logistics		

Big-Data and Cloud	AI and Learning	Integrated Technology	Others	
Big Data	Intelligent Robot	Integrated Control	Electronic Commerce	
Data Platform	Robot Advisor	Information Integration	Financial Technology	
Data Management	Machine Learning	Integrated Solutions	Block-Chain	
Data Mining	Deep Learning	Integrated System	In Memory Computing	
Intelligent Data Analysis	Artificial Intelligence		Natural Language Processing	
Data Science			Unmanned Retail	
Data Network			NFC Payment	
Industrial Cloud				
Cloud Computing				
Cloud Ecology				
Cloud Platform				
Cloud Service				

Notes: The table presents examples of key words related to different sub-components of the digitalization index. We use these key words to count the lexical frequency of the digitalization index.

Table A.2: Sub-Components of Digitalization Index

	Cities with Broadband Policy		Cities with No Broadband Policy	
	2007–2013	2014–2021	2007–2013	2014–2021
	(1)	(2)	(3)	(4)
Internet	0	5	0	2
Digital Technology	0	3	0	1
Smart and Intelligence	0	11	0	2
Automatic Technology	0	4	0	0
Information Technology	2	9	2	4
Big-Data and Cloud	0	5	0	2
Artificial Intelligence and Learning	0	4	0	0
Integrated Solutions	1	4	1	2
Others	1	5	1	2

Notes: Numbers represent the 75the percentile of the lexical frequency of these sub-components of the total digitalization index. Internet related activities, such as internet mobile, internet mode, internet platform, internet strategy, internet marketing, etc. by a firm. It consists of 21 internet related variables. Digital technology related variables such as digital marketing, digital technology, digital network, digital finance, etc. by a firm. It consists of 10 related variables. Smart and Intelligent technology related variables such as intelligent robots, smart logistics, smart factory, smart grid, digital intelligence, etc. by a firm. It consists of 33 related variables. Automatic technology related variables such as automatic production, automatic control, automatic face recognition, etc. by a firm. It consists of 8 related variables. Information Technology related variables such as information management, information network, information software, etc. by a firm. It consists of 12 related variables. Big-data and Cloud related variables, such as big data, data science, could platform, cloud services, data visualization, etc. by a firm. It consists of 16 related variables. Artificial Intelligence and Learning related variables such as artificial intelligence, intelligent robot, machine learning, etc. by a firm. It consists of 5 related variables. Integrated technology related variables such as integrated solutions, integrated system, etc. by a firm. It consists of 6 related variables. All Other types of technology related variables such as electronic commerce, financial technology, quantitative finance, block chain analysis, e-commerce, etc by a firm. It consists of 38 related variables.

Table A.3: Broadband Policy and Digital Index of Firms: Effects on Other Digitalization Indices – Extensive Margin

	Internet Technology	Digital Technology	Smart and Intelligence	Automatic Technology	Information Technology
	(1)	(2)	(3)	(4)	(5)
$Broadband_{2013} \times City_{fc}$	0.188***	0.154***	0.267***	0.116***	0.311***
	(0.032)	(0.033)	(0.041)	(0.025)	(0.044)
R-Squared	0.59	0.56	0.64	0.51	0.73
	Big-Data and Cloud	AI and Learning	Integrated Technology	Others	
	(6)	(7)	(8)	(9)	
$Broadband_{2013} \times City_{fc}$	0.173***	0.080***	0.258***	0.219***	
	(0.028)	(0.023)	(0.040)	(0.029)	
R-Squared	0.59	0.56	0.59	0.57	
N	30,538	30,538	30,538	30,538	30,538
Firm FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Industry (2-digit) × Year FE	Yes	Yes	Yes	Yes	Yes

Notes: Columns (1) – (9) use different indices of digitalization of a firm. These indices are based on lexical frequencies of these different indices mentioned by firms in their annual reports. It consists of 144 internet related variables. Column (1) uses adoption of internet related activities, such as internet mobile, internet mode, internet platform, internet strategy, internet marketing, etc. by a firm. It consists of 21 internet related variables. Column (2) uses adoption of digital technology related variables such as digital marketing, digital technology, digital network, digital finance, etc. by a firm. It consists of 10 related variables. Column (3) uses adoption of smart and intelligent technology related variables such as intelligent robots, smart logistics, smart factory, smart grid, digital intelligence, etc. by a firm. It consists of 33 related variables. Column (4) uses adoption of automatic technology related variables such as automatic production, automatic control, automatic face recognition, etc. by a firm. It consists of 8 related variables. Column (5) uses adoption of information technology related variables such as information management, information network, information software, etc. by a firm. It consists of 12 related variables. Column (6) uses adoption of big-data and cloud related variables, such as big data, data science, could platform, cloud services, data visualization, etc. by a firm. It consists of 16 related variables. Column (7) uses adoption of AI and learning related variables such as artificial intelligence, intelligent robot, machine learning, etc. by a firm. It consists of 5 related variables. Column (8) uses adoption of integrated technology related variables such as integrated solutions, integrated system, etc. by a firm. It consists of 6 related variables. Column (9) uses all other types of technology related variables such as electronic commerce, financial technology, quantitative finance, block chain analysis, e-commerce, etc by a firm. It consists of 38 related variables. Our dependent variable uses a indicator value 0 or 1. It takes a value 1 if a firm i located in s (where the broadband policy was implemented) after 2013 mentions any of these related variables in their annual reports. $Broadband_{2013}$ is a time dummy which takes a value 1 if year is greater than 2013. $City_{fc}$ is a spatial dummy – it takes a value 1 if a firm f is located in city c where the broadband policy was implemented between 2014 and 2016. Numbers in the parentheses are two-way clustered standard errors at the year and city level. Intercepts are not reported. ***, **, * denotes statistical significance at 1%, 5%, and 10%.

B Graphs

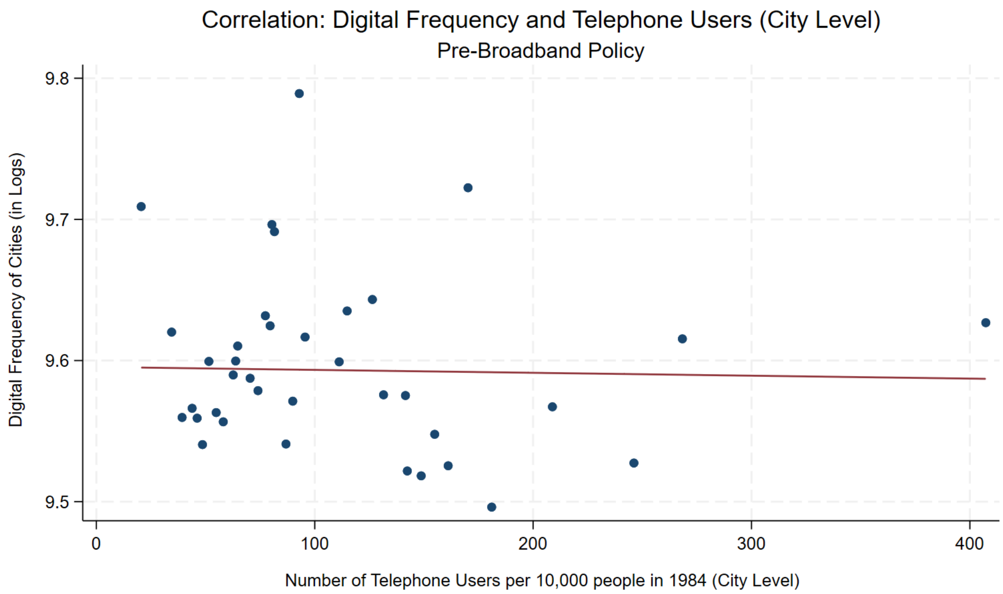

Figure B.1: Correlation: Digital Frequency of Cities with Broadband Policy and Telephone Users, 2007–2013

Note: Figure presents unconditional correlation between the digital frequency of cities which were selected for the implementation of the broadband policy in 2014 or 2015 or 2016 and the number of telephone users per 10,000 in 1984 before the policy was implemented, i.e., for the years 2007–2013.