Data Marketization: Paving the Way for Digital Innovation

Yitian Zhang* Shunwei Zhu† Sipeng Zeng‡

October 18, 2024

Abstract

In the digital economy, data has emerged as a pivotal production factor, on par with labor, capital, and technology. Our study examines the impact of marketization of data on firms' digital innovation output and production efficiency. By constructing an endogenous growth model that incorporates data intermediaries and conducting empirical tests facilitated by the staggered establishment of data trading platforms in China, we find that data marketization encourages firms to file a greater number of digital patents and significantly enhances their total factor productivity. Our paper highlights the significance of data mobility among enterprises for the growth of the business economy.

Keywords: Digital innovation, Data exchange, Data trading platform, Factor marketization

JEL Codes: G14, G32, O32

*Email: ytzhang.19@saif.sjtu.edu.cn, Shanghai Advanced Institute of Finance, Shanghai Jiao Tong University.

†Email: zhushunwei@sjtu.edu.cn, Antai College of Economics and Management, Shanghai Jiao Tong University.

‡Email: spzeng-bp@ckgsb.edu.cn, Cheung Kong Graduate School of Business.

1 Introduction

In the digital economy, data has become an integral part of production, combining with factors like labor, capital, and technology to drive economic activity (Jones and Tonetti 2020a; Farboodi and Veldkamp 2021; Veldkamp and Chung 2024). The effective utilization of data can enhance decision-making efficiency, improve prediction accuracy, and ultimately boost productivity and profitability for enterprises (Wamba et al. 2017; Müller et al. 2018). From the perspective of knowledge production, Cong et al. (2021) and Cong et al. (2022) note that data plays a crucial role in knowledge accumulation, enhancing firm innovation efficiency, leading to increased knowledge generation and patent output, and improving production processes. However, as an intangible asset, data faces significant challenges in pricing and distribution (Veldkamp 2023; Farboodi and Veldkamp 2023).As a result, data is less likely to circulate effectively compared to human resources, capital, or even knowledge. Fortunately, open trading platforms may enhance the pricing and trading of data. Numerous studies have discussed how these platforms can facilitate the free flow of factors such as knowledge and improve resource allocation, highlighting several benefits for corporate innovation and technological advancement (Rosenkopf and Almeida 2003; Boudreau 2010; Han et al. 2024). This raises an important question: can trading platforms promote the flow of data and, in turn, benefit firms through digital innovation[1] and increased productivity? Our study aims to answer this question both theoretically and empirically.

Countries like the U.S. and Japan have developed data markets to reduce costs, combat monopolies, and enhance data utilization.[2] Our study focuses on China for two main reasons. First, China's data volume is growing rapidly, with the total data production reaching 32.85 zettabytes in 2023.[3] This substantial data volume underscores China's vibrant digital economy and technological advancement. Moreover, the scale of China's digital economy, exceeding 50 trillion yuan and accounting for 42.8 percent of GDP, further highlights the critical role of data in driving economic growth.[4]The wide range of industries and applications relying on data in China provides a rich context for exploring how data can be effectively used to drive innovation and improve productivity.

Second, identifying the causal effects of data commercialization on digital innovation and productivity poses challenges in other country contexts. In these countries, data trading platforms are typically either private or national, enabling firms to easily access

[1]Digital innovation involves the creation and transformation of market offerings, business processes, or models through digital technologies (Nambisan et al. 2017). This includes both the development of new technologies (Benassi et al. 2020;Glinkina et al. 2020) and their integration into products and services (Johansson et al. 2020).

[2]For example, Snowflake Data Marketplace in the U.S. and NTT Data's Data Trading Market in Japan.

[3]Source:https://www.gov.cn/yaowen/liebiao/202405/content_6953440.htm.

[4]Source:https://www.cnii.com.cn/rmydb/202408/t20240830_597714.html.

the services provided by these platforms. In contrast, China's strategy of establishing data trading platforms across different provinces offers a unique research context. This fragmentation of interregional trade, particularly concerning intangible assets like intellectual property, is influenced by China's province-based intellectual property system and the difficulties involved in resolving legal disputes across provincial boundaries (Bai et al. 2004; Poncet 2005). Consequently, leveraging the unique institutional setting of data trading platforms in China allows us conduct a difference-in-differences (DiD) analysis to assess the impact of data trading on digital innovation.

Initially, we evaluate the role of data intermediaries in economies from a theoretical perspective. Our benchmark model, following Cong et al. (2022), includes three sectors: consumers, the production sector, and the innovation sector. Here, the social planner's goal is to maximize consumer utility by optimally allocating human capital between the production sector and the innovation sector. In this scenario, the innovation growth rate is determined by exogenous factors such as innovation efficiency, consumer time preferences, and substitution elasticity. Recognizing the important role that data can play in the production and innovation sectors, we introduce a new sector to the benchmark model: data intermediaries. Operating in a competitive market, these intermediaries operate under zero-profit conditions, where prices for buying and selling data are endogenously determined by supply and demand. Similarly, the social planner maximizes consumer utility by determining the labor allocation and the quantity of data that consumers sell to achieve optimal social welfare. In this extended model, the innovation growth rate is further influenced by the importance of data in the production and innovation sectors, as well as the severity of data privacy concern. Our simulation analysis compares the digital innovation growth rate and total factor productivity (TFP) with and without data intermediaries, demonstrating that economies benefit significantly from the presence of data intermediaries, resulting in higher digital innovation growth rate and TFP.

Next, our study empirically tests the implications of the model and obtains consistent results. One of the primary challenges in conducting empirical work on digital innovation is that the extent of digital innovation is typically not reflected in traditional information disclosure channels, such as annual reports and financial statements. As a consequence, most existing studies focus on the application of specific digital technologies (Yang 2022; Beltagui et al. 2020; Scott et al. 2017). However, patent information is widely regarded as a benchmark for assessing a firm's innovation capabilities. Identifying a firm's digital innovation performance through patent information can help capture the level of digital technological innovation using digital patents (Ahn 2020; Wang and Hsu 2021; Corvello et al. 2023; Corvello et al. 2023), thereby overcoming measurement difficulties. This study follows the current academic approach to define "green innovation" by identifying and classifying patents according to International Patent Classification (IPC) codes that correspond to relevant technology fields (Haščič and Migotto 2015; Cohen et al. 2020;

Pan et al. 2021; Bolton et al. 2023; Hege et al. 2023; Favot et al. 2023; Hege et al. 2024). This method allows us to quantify the number of digital patents at the firm level.

We utilize the document "Classification of Core Industries in the Digital Economy and Corresponding International Patent Classifications (2023)" published by the China National Intellectual Property Administration (CNIPA) in 2023. This document, based on the "Statistical Classification of the Digital Economy and Its Core Industries (2021)" and the "International Patent Classification (IPC)", provides a reference table that aligns IPC codes with the core industries of the digital economy. It comprehensively covers patents related to these core industries. This identification approach leverages the strengths of IPC classification to accurately capture digital patents held by firms while ensuring scientific rigor of the identification process and consistency with the standards from similar studies through the use of CNIPA's official classification criteria.

Empirically, we exploit the staggered establishment of data trading platforms in China and conduct a DiD analysis to assess their impact on firm innovation. The establishment of these platforms serve as an exogenous shock for data marketization, justified for several reasons. First, data trading platforms and exchanges employ privacy encryption technologies that make data "usable but not visible." This approach addresses privacy concerns, protects the rights of data owners, and promote large-scale, normalized data trading activities, thereby facilitating data flows and transactions. Second, these platforms implement a multi-level data use mechanism, formulate scientific market access standards for participants, and ensure that data circulation and transactions are compliant, orderly, transparent, and efficient. This effectively prevents data monopolization and supports the multi-level flow of data elements. Third, data platforms serve as a critical component of the basic infrastructure for data markets, significantly expanding the scale and scope of data transactions. Currently, data platforms have become important market instruments that integrate data trading, analysis, and management.

Our baseline DiD estimation results suggest that the establishment of data trading platforms is associated with a 6.29% [5] increase in firm digital patenting output. In addition, we examine the relationship between a firm's distance from the city where the data trading platform is located and its innovation output. Our results show that firms located farther from the data city—usually a provincial capital or economically developed area—exhibit lower innovation levels prior to the establishment of the platform. However, following the platform's establishment, these firms experience a more pronounced growth in innovation output. This finding implies that data platforms significantly enhance digital innovation, especially for firms in remote areas within the same province, highlighting the role of data intermediaries in promoting innovation and production efficiency.

Despite the staggered establishment of data trading platforms providing multiple shocks for our DiD analysis, two potential concerns arise: reverse causality and the

[5] $e^{0.061} - 1$.

omission of important variables. To address the issue of reverse causality, we perform a dynamic treatment effect analysis. This analysis finds no significant differences in digital innovation between the treated and control groups before the establishment of data trading platforms, thereby rejecting the demand-driven interpretation of our results and ruling out reverse causality. Additionally, our regional-level dynamic treatment analysis indicates that the treated and control provinces do not significantly differ in terms of the size and development level of their regional economy, R&D, innovation, technological development, or the fiscal capacity of local governments before the platforms were established. The second concern is that data trading platforms may correlate with other government policies that encourage digital innovation. To mitigate this issue, we incorporating variables representing provincial policies on digitization. This approach allows us to isolate the impact of the establishment of data trading platforms from other provincial digitization policies or economic conditions, thus addressing the potential omission of important variables. Our results remain robust when using pseudo-treatment individuals and pseudo-treatment timing placebo tests simultaneously.

Optimal resource allocation, such as the transfer of production factors from inefficient to efficient sectors or firms, can significantly enhance the productivity of the economy. The establishment of data trading platforms facilitates the flow of data, which serves as a vital production factor, thereby maximizing its benefits. Our study further examines the impact of data commercialization on firms' TFP, emphasizing digital innovation as a key intermediary that effectively bridges the relationship between data commercialization and productivity growth. Our empirical findings support Acemoglu and Restrepo (2018)'s theory that innovative digital technology can reduce a firm's reliance on traditional resource factors, thereby promoting a transformation in business models that improves resource utilization efficiency.

Our contributions are as follows: First, this paper contributes to the literature on the data economy. Data plays a crucial role in the broader digital economy by enhancing decision-making, reducing costs, and driving sectoral transformations (Goldfarb and Tucker 2019; Veldkamp and Chung 2024). It serves as a key input in production and innovation processes, as evidenced by research exploring that data integration with other production factors fosters continuous growth and industrial innovation (Cong et al. 2022). The literature on the data economy underscores the transformative impact of data as a nonrival good that enhances economic growth, innovation, and productivity. A central theme across studies is the nonrival nature of data, which allows it to be used simultaneously by multiple firms without depletion, leading to increasing returns and efficiency gains (e.g., Farboodi and Veldkamp 2021; Jones and Tonetti 2020b). However, this characteristic also raises important questions about data ownership, privacy, and optimal usage. Balancing the benefits of data utilization with privacy protection is critical, as noted in discussions on the trade-offs between data sharing and individual privacy (Cong

et al. 2021). Building on existing research, our paper is the first to theoretically and empirically validate the role of data marketization in the broader economy by fostering corporate digital innovation and enhancing firms' total factor productivity.

Secondly, our study provides a new perspective on the digital innovation growth from the standpoint of data marketization. With the growing global impact of digital transformation (Verhoef et al. 2021), more scholars are focusing on the innovation processes tied to digital technologies (Hess et al. 2016; Scott et al. 2017; Beltagui et al. 2020; Bertello et al. 2021; Annarelli et al. 2021; Yang 2022; Li et al. 2023). Understanding the key factors that drive the effectiveness of these processes is crucial (Annarelli et al. 2021). Research identifies several influencing factors at various levels, including institutional (Wei et al. 2022), organizational (Cheng and Wang 2022), team (Ardito et al. 2021), and individual level (Opland et al. 2022; Corvello et al. 2023). Our paper contributes significantly to this literature in two ways. First, we introduce an innovative method for measuring digital innovation. Drawing on China's "Classification of Core Industries in the Digital Economy and Corresponding International Patent Classifications (2023)", we define digital innovation based on the IPC codes of patents filed by companies. This novel approach transcends the limitations of normative studies and case analyses, enabling extensive empirical research on a large scale without being confined to a single technology. It paves the way for a more comprehensive understanding of digital innovation. Second, while previous studies have overlooked the influence of data on digital innovation, we boldly expand the boundary of factors influencing it. By delving into this unexplored territory, we offer new insights and perspectives that can reshape the research landscape in this field. Our work not only enriches the theoretical framework but also has practical implications for businesses and policymakers seeking to foster digital innovation.

Finally, it also enriches the body of theoretical research on the real effects of the establishment of data trading platforms. Akcigit and Liu (2016) find that data trading platforms, as third-party data institutions, can help compensate for innovation failures on the part of data buyers, thereby mitigating welfare losses caused by insufficient firm incentives. Bergemann and Bonatti (2019) argue that firms that purchase personal consumer information from data trading platforms may earn excess profits through third-degree price discrimination, but this behavior may harm consumers, resulting in a net loss of social welfare. In contrast to their micro-economic approaches, our endogenous growth model discusses the role of data trading platforms in driving the level of innovation and productivity of the economy as a whole from a macro perspective.

The rest of the paper is organized as follows. Section 2 provides the institutional background of data market implementation in China. In Section 3, we compare two models to examine how data intermediaries influence firms' digital innovation and total factor productivity. Section 4 presents the variable definitions and sample description. Section 5 performs a Difference-in-Differences analysis to test the conclusions derived from the

theoretical model and presents various robustness checks to strengthen our identification strategy. Section 6 empirically evaluates the impact of data market implementation on firms' TFP. Finally, Section 7 concludes the paper.

2 Institutional background of data market implementation in China

Since the 2017 Chinese government work report first proposed vigorously developing the digital economy, its growth has ascended to an unprecedented strategic height in China. Unlike traditional factors of production, data is characterized by non-rivalry and constitutes a key element of the digital economy (Jones and Tonetti 2020a). Consequently, numerous data trading platforms have mushroomed across the country to promote the commercialization of data. These data exchanges offer platform services, brokered transactions, data sales, and value-added services, facilitating the utilization of data and delivering economic and operational benefits to enterprises. The marketization of data establishes a comprehensive market mechanism for data resources, ensuring the circulation and trading of data. This integration enables data to be incorporated into the production process along with labor, capital, and technology. Through this incentive system, it encourages enterprises to produce and trade data within the market, thereby leveraging data dividends and accelerating the pace of innovation.

In 2014, Beijing established two big data trading platforms, marking China's entry into the exploratory stage of data trading platform construction. In 2015, the Guiyang Data Exchange launched in April with state-owned assets commission support. In August, Wuhan established the Yangtze River Big Data Trading Center. Subsequently, more data exchanges emerged. From 2019, central policies accelerated data trading platform and circulation system construction. Local governments included such construction in their plans. By 2023, 44 data trading platforms led or approved by local governments had been established. See Table 1 for their establishment times.

The types of data eligible for trading have expanded from the initial fields of finance and the internet to encompass healthcare, meteorology, transportation, logistics, communications, education, humanities, geography, and lifestyle services. For instance, in January 2023, the Guangzhou Data Exchange launched the country's first industry data index release platform, covering more than 80 industry data indices across 10 sectors. Moreover, with the development of sci-tech innovation enterprises, the demand for innovative intelligence data has been on the rise. This data comprises patents, trademarks, software copyrights, literature journals, corporate information, public opinion, and capital market data. Figure A1 depicts the website of a data trading platform, providing details about the types of traded products, applicable usage scenarios, product names,

the companies offering these products, and pricing information.

3 Theoretical model

In this section, we construct endogenous growth models (Romer 1990) to discuss changes in innovation and efficiency in the data economy. Specifically, consider the financial industry (e.g., JPMorgan Chase, Bank of America, etc.). By collecting and analyzing a large amount of customer data, including transaction records, behavioral preferences, risk tolerance and other information, financial companies can deeply understand customer needs and develop innovative digital financial products and services. For example, they can launch convenient mobile payment products, use big data and artificial intelligence analysis to provide digital credit services to customers, and offer intelligent investment advisory services based on big data. In such a scenario, consumer data is an important element for enterprises to carry out digital innovation of their products.

We first introduce the agents in the economy, i.e., representative consumers, incumbent firms and the innovation sector following Cong et al. (2022) and solve for the optimal allocation of labor. Subsequently, we add the data intermediary sector to the model and consider the optimal allocation of labor in the new context. Finally, we compare the growth rate of innovation and efficiency in the two economies in the presence and absence of data intermediary.

3.1 Economy without data intermediary

3.1.1 Representative consumers

Homogeneous consumers live in a continuous-time economy. We assume that the size of consumer population is a constant to L in order to highlight endogenous economic growth that is not driven by population growth. Without loss of generality, we assume each consumer's instantaneous utility is:

$$u\left(c_t\right) = lnc_t \tag{1}$$

where $c_t = \left(\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{\sigma}{\sigma-1}}$ is the consumption index, i.e., the CES aggregate of the consumption of the differentiated varieties. $\sigma > 1$ is the elasticity of substitution. c_{it} is the consumption level of variety i at time t. N_t is the total number of varieties of products.

Each consumer owns a unit of labor, which he supplies inelastically in exchange for a wage w_t. They also hold assets a_t that pay a return r_t, which are claims on the value of

the monopolistically competitive firms. The representative consumer solves,

$$\max_{\{c_{it}\}} \int_0^\infty e^{-\rho t} u\left(c_t\right) dt \tag{2}$$

subject to

$$c_t = \left(\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}$$
$$\dot{a}_t = r_t a_t + w_t - \int_0^{N_t} p_{it} c_{it} di$$

Here ρ captures the consumers' time preference and p_{it} is the price of variety i at time t.

Then, the Euler equation for consumption from Hamiltonian is simply:

$$\frac{\dot{c}_t}{c_t} = r_t - \rho \tag{3}$$

The rate of change in consumption is positively correlated with interest rates and negatively correlated with the time discount rate. Higher current interest rates lead investors to reduce current consumption for saving, increasing future consumption and the consumption growth rate. Conversely, a higher discount rate discounts future consumption more heavily, making investors impatient and prioritize current consumption over future consumption, thus lowering the consumption growth rate.

3.1.2 Incumbent firms

Each firm's output on digital product i is determined by the quality of the labor invested in the product L_{it} and the idea of quality A:

$$Y_{it} = AL_{it} \tag{4}$$

Similar to the consumption index shown in (1), the aggregate output Y_t is given as follows:

$$Y_t = \left(\int_0^{N_t} Y_{it}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} \tag{5}$$

With the first order conditions from the consumers' problem, each firm takes demand for its variety as given, the price of product p_{it} is endogenously determined by the economy:

$$p_{it} = \left(\frac{c_t}{c_{it}} \right)^{\frac{1}{\sigma}} = \left(\frac{Y_t}{Y_{it}} \right)^{\frac{1}{\sigma}} \tag{6}$$

Letting V_{it} denote the market value of firm i, the incumbent firm's profit maximization

problem is:

$$r_t V_{it} = \max_{\{L_{it}\}} \left(\frac{Y_t}{Y_{it}} \right)^{\frac{1}{\sigma}} Y_{it} - w_t L_{it} + \dot{V}_{it} \tag{7}$$

For simplicity, we define:

$$L_{pt} = \int_0^{N_t} L_{it} di$$

3.1.3 Innovation sector

A new digital variety can be designed and created by innovation sector as in Cong et al. (2021) and Cong et al. (2022). Potential entrants enter by conducting R&D using labor L_{et}. We specify the evolution of the aggregate innovation possibility frontier, that is, the number of varieties:

$$\dot{N}_t = \eta N_t L_{et} \tag{8}$$

where $\eta > 0$ is an efficiency term of innovation, L_{et} is labor employed in the innovation sector.

The potential entrants maximize their expected profit by choosing the units of labor to employ,

$$\max_{\{L_{et}\}} \dot{N}_t V_t - w_t L_{et} \tag{9}$$

The FOC with respect to labor is then,

$$\eta N_t V_t = w_t \tag{10}$$

3.1.4 Optimal allocation

We define the proportion of labor invested in innovation sector and incumbent firms as,

$$l_{et} = \frac{L_{et}}{L_t} \quad \text{and} \quad l_{pt} = \frac{L_{pt}}{L_t}$$

We can now state the social planner's problem concisely. The key allocations that need to be determined are how to allocate labor between production and innovation. The optimal allocation solves,

$$\max_{\{l_{pt}\}} \int_0^{\infty} e^{-\rho t} ln c_t dt \tag{11}$$

subject to

$$\begin{aligned} c_t &= \frac{Y_t}{L} \\ Y_t &= A N_t^{\frac{1}{\sigma-1}} l_{pt} L \\ \dot{N}_t &= \eta N_t l_{et} L \\ l_{pt} + l_{et} &= 1 \end{aligned}$$

The Hamiltonian for the optimal allocation is,

$$\mathcal{H}\left(l_{pt}, N_t, \lambda_t\right) = ln\left(AN_t^{\frac{1}{\sigma-1}} l_{pt}\right) + \lambda_t \eta N_t l_{et} L \tag{12}$$

We then have following FOCs,

$$\frac{\partial \mathcal{H}}{\partial l_{pt}} = \frac{1}{l_{pt}} - \lambda_t \eta N_t L = 0 \tag{13}$$

$$\frac{\partial \mathcal{H}}{\partial N_t} = -\dot{\lambda}_t + \rho\lambda_t = \frac{1}{\sigma-1}\frac{1}{N_t} + \lambda_t \eta l_{et} L = 0 \tag{14}$$

Finally, along the BGP, the rate of innovation in equilibrium is,

$$\frac{\dot{N}_t}{N_t} = \eta L - \rho\left(\sigma - 1\right) \tag{15}$$

In this economy, the rate of innovation in an economy is determined by a combination of exogenous innovation efficiency, consumers' time preferences, and elasticity of substitution. When consumers discount future consumption less, the increase in utility from future product variety is higher, so the innovation department will have a higher innovation growth rate. When consumers have a lower consumption substitution elasticity, the utility from increased product variety is harder to be replaced by more consumption of original products, so the innovation department will have a higher innovation growth rate.

3.2 Data Economy

3.2.1 Representative consumers

In economy where data generation exists, each consumer produces data as by-products of consumption (e.g., Veldkamp 2005; Jones and Tonetti 2020a; Cong et al. 2021 and Cong et al. 2022). They can sell data to a data intermediary and choose how much data to sell to balance the gain in income versus the cost to privacy. Each consumer's instantaneous utility is:

$$u\left(c_t, d_t\right) = lnc_t - d_t^{\kappa} \tag{16}$$

Note that κ captures the privacy costs, such as expected losses from leakage, the extra costs when a platform uses consumer data for price discrimination, the hacking of a centralized database. The disutility increases with respect to the quantity of data sold.

The representative household solves,

$$\max_{\{c_{it}, d_t\}} \int_0^{\infty} e^{-\rho t} L_0 u\left(c_t, d_t\right) dt \tag{17}$$

subject to

$$c_t = \left(\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}$$

$$\dot{a}_t = r_t a_t + w_t + p_{dt} d_t - \int_0^{N_t} p_{it} c_{it} di$$

where p_{dt} is the price of data at time t.

The Euler equations for consumption and data sales are simply,

$$\frac{\dot{c}_t}{c_t} = r_t - \rho \tag{18}$$

$$\frac{\dot{p}_{dt}}{p_{dt}} - (\kappa - 1) \frac{\dot{d}_t}{d_t} = r_t - \rho \tag{19}$$

3.2.2 Incumbent firms

Each incumbent firms produces a distinct variety through buying data sets D_{it} from data intermediary and hiring labor L_{it}. Each firm takes the price of data and the wage of labor as given. The output function is simply:

$$Y_{it} = A_i L_{it} = D_{it}^{\alpha} L_{it}, \quad with \;\; \alpha \in (0,1) \tag{20}$$

where A_i is quality of ideas, D_{it} is the amount of data used in producing the quality of variety i, L_{it} is labor. Parameter $\alpha \in (0,1)$ captures the importance of data in production.

Similar to eq(5) in the prior subsection, we define the aggregate output Y_t as follows,

$$Y_t = \left(\int_0^{N_t} Y_{it}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} \tag{21}$$

Letting V_{it} denote the market value of firm i, the incumbent firm's problem is

$$r_t V_{it} = \max_{\{L_{it}, D_{it}\}} \left(\frac{Y_t}{Y_{it}} \right)^{\frac{1}{\sigma}} Y_{it} - w_t L_{it} - p_{p,it}^d D_{it} + \dot{V}_{it} \tag{22}$$

where the price of data sets faced by incumbent firms is denoted as $p_{p,it}^d$.

FOCs of incumbent firm's problem are,

$$\left(1 - \frac{1}{\sigma}\right) \left(\frac{Y_t}{Y_{it}} \right)^{\frac{1}{\sigma}} \frac{Y_{it}}{L_{it}} = w_t \tag{23}$$

$$\alpha \left(1 - \frac{1}{\sigma}\right) \left(\frac{Y_t}{Y_{it}} \right)^{\frac{1}{\sigma}} \frac{Y_{it}}{D_{it}} = p_{p,it}^d \tag{24}$$

Since the price, outputs, and profits are equal for all incumbent firms, we simplify

notations by writing $V_{it} = V_t$ and $D_{it} = D_t$ for all i and t in the remainder of this paper.

3.2.3 Innovation sector

In economy that includes data production, potential entrants conduct R&D using labor L_{et}, and data D_t purchased from data intermediary. We specify the innovation sector production function and the dynamics of the innovation frontier as:

$$\dot{N}_t = \eta N_t L_{et}^{1-\beta} D_t^{\beta} \tag{25}$$

where $\eta > 0$ is an efficiency term of innovation. $\beta \in (0, 1)$ represents the relative contribution of data and labor in innovating process.

The maximization problem for potential entrant is given by:

$$\max_{\{L_{et}, D_t\}} \dot{N}_t V_t - w_t L_{et} - p_{e,it}^d D_t \tag{26}$$

The FOCs with respect to labor and data are then,

$$(1-\beta)\,\eta N_t L_{et}^{-\beta} D_t^{\beta} V_t = w_t \tag{27}$$

$$\beta \eta N_t L_{et}^{1-\beta} D_t^{\beta-1} V_t = p_{e,it}^d \tag{28}$$

3.2.4 Data intermediary

We introduce a data intermediary industry with competitive entry following Jones and Tonetti (2020a) and Cong et al. (2022). The data intermediary collects data from consumers at a given price p_{dt} and integrates data into a data set D_t. The intermediary maximizes profits by choosing the quantity of data to buy from consumers and the discriminatory price at which it sells data sets to incumbent firms. The data intermediary makes zero profit in the presence of free entry. Thus, we characterize data intermediation through a cost minimization problem and a zero-profit condition. The intermediary solves:

$$\min_{\{d_t\}} p_{dt} d_t L \tag{29}$$

subject to

$$D_t \leq d_t L$$

and the zero-profit condition:

$$\int_0^{N_t} p_{p,it}^d D_t di + p_{e,it}^d D_t = p_{dt} d_t L$$

The functions of the data intermediary in the economy are combining all consumers' data and selling them to production firms and entrant innovators simultaneously. The price

of selling data to firms $p_{p,it}$, the price of selling data to the innovation sector $p_{e,it}$ and the price of collecting data from consumers p_{dt} are all endogenously determined in the economy.

3.2.5 Social planner's problem

The social planner needs to determine the labor allocation and the quantity of data that consumers contribute, solving,

$$\max_{\{l_{pt},d_t\}} \int_0^\infty e^{-\rho t}\left(lnc_t - d_t^\kappa\right)dt \tag{30}$$

subject to

$$\begin{aligned} c_t &= \frac{Y_t}{L} \\ Y_t &= N_t^{\frac{1}{\sigma-1}} l_{pt} L^{1+\alpha} d_t^\alpha \\ \dot{N}_t &= \eta N_t l_{et}^{1-\beta} L d_t^\beta \\ l_{pt} + l_{et} &= 1 \end{aligned}$$

Growth rate of innovation in data economy's equilibrium $\frac{\dot{N}_t}{N_t}$ solves the following systems,[6]

$$\begin{cases} \frac{\dot{N}_t}{N_t} = \eta L \left[\frac{\frac{1-\beta}{\beta}\left(\kappa d^k - \alpha\right)}{1+\frac{1-\beta}{\beta}\left(\kappa d^k - \alpha\right)}\right]^{1-\beta} d^\beta \\ \frac{\dot{N}_t}{N_t} = \frac{(\sigma-1)\rho}{\beta}\left(\kappa d^k - \alpha\right) \end{cases} \tag{31}$$

3.3 Simulation

We simulate the two types of economies to compare the optimal rate of innovation in the economy under different scenarios of the presence or absence of data intermediation. In Section 3.1, innovation activities depends only on the amount of labor invested, while in Section 3.2, the rate of innovation is influenced by the amount of data collected from consumers by the data intermediary. We compare the results derived in these two subsections. Table 2 summarizes the model parameters. The parameters are determined following Cong et al. (2021) and Cong et al. (2022). The variable we focus on is the growth rate of product variety $\frac{\dot{N}_t}{N_t}$ and total factor productivity (TFP).

Figure 1 shows the evolution of $\frac{\dot{N}_t}{N_t}$ as a function of the parameters in the two economies. It can be seen that when the parameters are within a rational range, the growth rate of innovation is higher in the economy that includes data intermediaries than in the one that does not. Subfigure (e) shows the rate of innovation in the two economies

[6]See Appendix B for the proof process.

for different consumer discount rates ρ. The more consumers discount future consumption, the lower the economy's equilibrium rate of innovation. When ρ is around 0.1, the rate of innovation in the two economies reaches a state of almost no difference. Subfigure (f) shows the innovation rate in the two economies for different innovation efficiency η. We can see that the gap between the growth rate of innovation in the economy with data intermediary and the one without data intermediary increases as the innovation sector becomes more efficient in factor utilization. Intuitively, the more innovation-driven an economy is, the more pronounced the role of the presence of data intermediaries in facilitating endogenous economic growth from innovation.

Figure 2 gives the evolution of TFP as a function of the parameters in the two economies. In economy with data intermediary, the TFP of firms grows intuitively with the importance and innovation efficiency of data elements in the production and innovation sectors. An increase in consumers' subjective discount rate means that they attach less importance to future earnings. Enterprises will reduce investment in innovation and pay more attention to current production, thereby reducing TFP. An increase in the degree of privacy concern will limit the circulation and use of data, affect innovation efficiency, and thus reduce TFP. An increase in consumption substitution elasticity means that consumers can more easily substitute between different products, which will lead to greater competitive pressure on enterprises and reduce investment in innovation, thereby reducing TFP. Similar to the innovation growth rate, within a reasonable parameter range, the TFP of an economy with the data intermediary is higher than that of an economy without data intermediary.

The results of the theoretical model show that the introduction of a data intermediary can increase the level of innovation and thus TFP. The relevant corollary clearly shows and emphasizes the important significance of data marketization. Data marketization is an important driving force for digital innovation. It is through the process of data marketization that valuable data resources are effectively allocated and utilized. This not only promotes innovation but also changes the production efficiency of enterprises, highlighting its profound significance in the era of data economy.

3.4 Discussion

In addition to the discussion of equilibrium innovation decisions in the previous sections, there are some factors that affect the speed of innovation and growth that are not included in the endogenous growth model in this paper. The creation of a data market addresses crucial issues such as standardization, clarity of ownership, and synergy with other factors of production, and thus may be important in driving digital innovation.

First, the primary objective of constructing a data market is to establish a public, transparent, and active data trading market where data collection standards and cir-

culation standards are unified. Consequently, the construction of a data market will accelerate the circulation of data, reduce the cost of data collection and acquisition, and facilitate the collection and transaction of innovative data by firms. Transaction cost theory posits that reducing transaction costs is crucial for the survival and development of firms (Coase 1937). For firms, lowering data transaction costs is a vital component in promoting digital innovation. Therefore, the construction of a data market can expedite the flow of data in the market, reduce the costs associated with data collection and acquisition for firms, and accelerate their research and development processes, thereby facilitating digital innovation.

Second, research indicates that issues surrounding data ownership not only affect the development, utilization, and circulation of data but also influence the innovation and development of the digital industry (Kerber 2021; Cong et al. 2021; Eckardt and Kerber 2024). When data ownership is uncertain, the compliance of data operations faces severe challenges during repeated mining and utilization processes. In cases of data rights infringement, it becomes difficult to identify the violators, leading to conflicts of interest among data users and obstructing the exploitation and realization of data value. Consequently, without clear data ownership, digital firms are unable to actively and creatively explore data value, nor can they transform various accumulated data into assets for innovation activities, hindering the progress of innovation and research and development activities. The construction of a data market establishes the rights and responsibilities of data trading entities, regulates the behavior of digital economy entities, clarifies the ownership and usage rights of data, defines the scope of tradable data, and providing legal and institutional safeguards for data transactions. Once data ownership is clarified, data owners could actively and creatively develop data value, customize different types of data, and facilitate the acquisition and use of digital industry data, thus promoting digital innovation activities and accelerating the digital innovation process of firms.

Lastly, the combination of data with other production factors in the production process is crucial for leveraging the value of data and promoting digital innovation. Kivimaa and Kern (2016) shows that data must be integrated with other production factors such as capital and labor to create new production functions through new combinations of production conditions. This integration can foster new industrial patterns based on industrial digitalization and digital industrialization, driving digital innovation. The construction of a data market can break the *data islands* and *data monopolies* that exist in reality, facilitating the entry of various types of data into the market and enhancing the synergy of data. Zhang et al. (2023) highlights that external connections, such as cross-institutional collaborations and international partnerships, are essential in breaking the constraints of knowledge path dependence. These connections bring new perspectives and resources, which are vital for fostering radical inventions. On one hand, the marketization of data can provide an information-sharing platform for collaborative innovation among firms,

improving the efficiency of innovation cooperation and generating a synergistic effect. On the other hand, the marketization of data accelerates the coordination and interaction among production factors within firms, with the deep integration of data with labor, knowledge, and management elements enhancing the quality of human capital, organizational learning capacity, and management efficiency. This multi-factor-driven approach promotes the integration of traditional and digital technologies, laying a micro-foundation for digital innovation.

4 Data, sample, and descriptive statistics

To conduct a thorough empirical investigation into the impact of data trading platforms on corporate innovation, we compile a unique dataset comprising detailed firm-level financial data and patent applications information.

4.1 Data sources and sample selection

For corporate innovation data, we collect a comprehensive dataset of detailed patent applications of listed companies up to the year 2022 filed with the Chinese National Intellectual Property Administration (CNIPA). The dataset includes the patent code of the patent, the name of the company that filed the patent and the time of filing. Among them, the patent code contains the classification information of the patent. We integrate patent data with company level financial data from CSMAR. In order to encompass patent applications of subsidiaries of publicly traded companies, we match the two datasets together using a matching algorithm based on company names. The final merged dataset contains comprehensive micro-level data on firms' annual financial performance and patent filings.

We focused on A-share firms listed on the Shanghai and Shenzhen stock exchanges, compiling panel data covering the period from 2008 to 2022. To ensure the representativity of our data, we implement the following criteria:(1) exclude ST-classified firms; (2) exclude firms in the financial industry; (3) exclude firms with less than one year since their initial public offering (IPO); (4) exclude firms with total assets less than total liabilities; (5) winsorize all variables except the dummy variables at the 1% level at both tails. After processing, a total of 24,418 observations of firm-year samples from 2008 to 2022 were obtained for this study.

4.2 Variable definitions

4.2.1 Explained variable

Digital innovation. We measure the digital innovation output of firms in a given year by the number of digital patents they file annually. Digital patents and four specific types of

digital patents are defined based on the International Patent Classification (IPC) codes as classified in the "Reference Table of Core Industries in the Digital Economy and International Patent Classification" issued by the CNIPA in 2023. A patent is classified as a digital patent for a specific industry if its patent code contains the industry code of a digital economy industry. The digital economy industries are divided into four major categories: Digital Product Manufacturing, Digital Product Service, Digital Technology Application, and Digital Factor Drive.

Digital Product Manufacturing includes the manufacturing of computers, communications and radar equipment, digital media devices, smart devices, electronic components and devices, and other digital products. Digital product services refer to patents related to the maintenance of digital products, such as computer and peripheral equipment repair and communication equipment repair. Digital technology applications consist of software development, telecommunications, broadcasting, satellite transmission, internet-based services, IT services, and other sectors that utilize digital technologies for various purposes. Digital factor-driven industries include patents related to internet platforms, information infrastructure construction, and the research and experimental development of digital technologies. The reference table creates a connection between the International Patent Classification and the core industries of the digital economy.

4.2.2 Explanatory variable

Treatment. This variable measures the establishment of data trading platforms in each province of China. If a province where firm i is located establishes a data trading platform in year t, the *Treatment* variable is assigned a value of 1 in that year and subsequent years; otherwise, it is assigned a value of 0. The treated province-level regions and treatment times are shown in Table 1.The first column of this table reports the treated province-level regions where the data trading platforms are established. The second column indicates the exchange location and the third column indicates the year the data trading platform was established. From this table, we observe the sequential establishment of data trading platforms across provinces in China, spanning from 2014 to 2022.

4.2.3 Control variables

First, we introduce some unique control variables in this paper. One is *Digital Others*, which serves as a measure for the level of a firm's innovation output other than its digital innovation. It is defined as the number of all patent applications minus the digital patents filed by a firm. Another is *Gov Attention*, which gauges the degree of provincial government attention to the digital economy, thereby excluding policy attention effects at the provincial level other than those of the data trading platforms. Following Peng et al. (2024), we constructed a word list containing vocabulary specialized in digital

technologies and digital applications. Then, we measured the extent of the government's attention to digitalization by taking the natural logarithm of the frequency of occurrences of the words in the list in the government's annual work reports.

Second, we select the following variables as control variables based on the literature. *Age* is calculated as the natural logarithm of one plus the number of years since a firm's establishment. *Growth* represents the annual growth rate of a firm's operating revenue. *Leverage* is defined as the ratio of a firm's total liabilities to total assets. *Size* is the natural logarithm of a firm's total assets for the year. *Employee* is the natural logarithm of a firm's number of employees. *Market Value* is the ratio of a firm's total market value to total assets. *Profitability* is the ratio of a firm's net profit to operating revenue. *Liquidity* is the ratio of a firm's current assets to total assets. *Cash Flow* is the ratio of a firm's net cash flow from operating activities to total assets. *Board Size* is the natural logarithm of a firm's number of board members. *Firm R&D* is the ratio of a firm's research and development expenditure to operating revenue. *Ownership* is a dummy variable, taking a value of 1 for state-owned firms and 0 otherwise.

4.3 Descriptive statistics

After data cleaning and processing, Figure 3 presents the trend of digital patent applications by all publicly listed companies in China. Different colors represent various types of digital patents. It is clear that the majority of digital patent applications are in the category of Digital Product Manufacturing. Moreover, the number of digital patent filings has increased exponentially since 2018. This highlights the significance of manufacturing industry in China as well as the rapid development of the digital economy in the country.

The descriptive statistics of our sample companies are shown in Table 3. All potentially unbounded variables are winsorized at the 1% extremes. As can be observed from Table 3, on average, a firm files 11 digital patent applications. During the sample period, some highly active firms apply for up to 162 digital patents. Additionally, on average, firms in the sample file 18 patents other than digital patents. Some extremely active firms in the sample period apply for as many as 376 non-digital patents. The non-negative integer nature of our dependent variable *Digital innovation* necessitates the use of count models, such as the Poisson model or the negative binomial regression model (Silva and Tenreyro 2006; Cohn et al. 2022). The data distribution reveals that the variance of the patent count greatly exceeds its mean, indicating an overdispersion problem. Since the Poisson model assumes that the variance equals the mean, which is clearly not applicable here, the negative binomial regression model is more appropriate (Zhang et al. 2023). Therefore, for the regressions in this paper where the number of patents is used as the dependent variable, the negative binomial regression model is employed.

5 Data Marketization and Corporate Innovation

Data trading platforms enhance the level of digital innovation by increasing the availability of data, an important element of digital innovation. To explore this hypothesis, we perform a difference-in-differences (DiD) analysis in this section. We utilize the staggered establishment of data trading platforms as a quasi-experiment to establish causality, detailed in Section 5.1 and Section 5.2 presents the results of robustness tests. Section 5.3 presents an analysis of heterogeneity based on geographic distance.

One might wonder why we chose enterprises in the province where the data trading platform is located as the treatment group instead of those in close proximity to the data trading platform. The reason is that interregional trading activities in China are fragmented along provincial lines, especially when it comes to the trading of intangible assets like intellectual property. This is due to China's province-based intellectual property system and the difficulty of resolving legal disputes across provincial borders (Bai et al. 2004; Poncet 2005). According to Singh and Marx (2013), state borders independently affect knowledge diffusion, beyond what can be explained by geographic proximity, such as metropolitan collocation or shorter within-region distances. Hunt et al. (2024) using U.S. data find that commuting zones located an additional 200 kilometers from the nearest AI hotspot experience a 17% lower growth in the share of AI job vacancies, partly due to the presence of state borders between certain commuting zones and their nearest hotspot, which may hinder migration and the flow of tacit knowledge. Therefore, local governments, as part of opening up channels for data elements, will encourage data trading platforms to serve enterprises in their own provinces more. There is an old Chinese saying, "The pavilion closest to the water enjoys the moonlight first". On the surface, this means those who are geographically close gain the first opportunity. However, its metaphorical meaning is that those close to authority power gain the first chance.

5.1 Baseline results

Over time, data trading platforms were gradually established in different regions of China, impacting different firms at different times. The staggered establishment of data trading platforms offers an advantage for our analysis. Using a multi-period DiD approach helps mitigate the typical identification problems that arise in studies based on a single shock. In single-shock scenarios, omitted variables that occur simultaneously with the shock can influence firms' innovation strategies, making it difficult to isolate the true impact of the shock. Therefore, the staggered establishment of data trading platforms in China provides a unique and ideal context for addressing endogeneity issues and establishing causality (Wang et al. 2023).

Exploiting the staggered establishments of data trading platforms as a quasi-experiment (Beck et al. 2010), we draw upon the study by Baker et al. (2022) and employ the

multi-period DiD method. This approach controls for time-constant individual differences between areas where the policy is implemented and those where it is not, accounts for variations before and after the policy implementation, and thereby more effectively identifies the net impact of construction on firm digital innovation.

$$Digital\ Innovation_{i,t} = exp\left(\alpha + \beta\, Treatment_{i,t} + \sum \gamma_k X_{k,i,t} + \theta\, Year_t + \delta\, Firm_i + \epsilon_{i,t}\right), \tag{32}$$

The subscript i represents firms, while the subscript t represents years. The dependent variable is digital patents applications as defined in Section 4.2.1. *Treatment* indicates whether a data trading platform already exists in the province where the enterprise is located, as defined in Section 4.2.2. $X_{k,i,t}$ is a vector of control variables including standard firm characteristics as defined in Section 4.2.3. $Year_t$ and $Firm_i$ represent the time varying aggregate shocks and time-invariant firm heterogeneity. $\epsilon_{i,t}$ is the error term. We follow Zhang et al. (2023) to employ a binomial negative regression model to estimate the coefficients. The standard errors are robust standard errors adjusted using the OIM method. β represents the treatment effects of data trading platforms and is the primary focus of the regression analysis. We report the results of our baseline DiD estimations in Table 4. The first column in Table 4 reports the estimation results without control variables, and we add the control variables in the second column in this table.

Our results show that the establishment of data trading platforms exerts a significant positive impact on firms' digital innovation. Specifically, in column (1), the coefficient of the treatment indicator is 0.135 and is statistically significant at the 1% level. This indicates that the presence of data trading platforms has increased the number of digital patent applications filed by companies by a factor of 0.144 [7]. Table 4's column (2) presents the results when control variables are included. The coefficient of the treatment indicator remains positive and significant, although with a smaller magnitude (0.061) and significant at the 5% level.

Our control variables also provide interesting insights. For example, *Digital Others* has a positive and significant coefficient, suggesting that firms with a higher number of other types of patents are also more likely to engage in digital innovation. In contrast, variables such as *Age* and *Size* have negative and significant coefficients, suggesting that older and larger firms are less likely to file digital patents. Meanwhile, the coefficient of *Gov Attention* is insignificant, indicating that, except for the establishment of data trading platforms, the local government's attention to the digital economy has a limited impact on firms' digital innovation.

We further categorize digital innovation into four types: Digital Product Manufacturing, Digital Product Service, Digital Technology Application, and Digital Factor Drive.

[7] $e^{0.135} - 1$.

Except for Digital Factor Drive patents, the other three types show significant positive effects, which may be due to the relatively small sample size of Digital Factor Drive patents. Among the digital patents produced by a firm, the majority belong to the Digital Product Manufacturing category, as shown in Figure 3, where the red-colored Digital Factor Drive is nearly absent.

5.2 Robustness checks

Although the staggered establishment of data trading platforms has the advantage of multiple impacts, our DiD analysis still faces three main issues. The first concern with our DiD specification is the issue of reverse causality. Specifically, a data trading platform might be set up in provinces where digital patenting activities are already robust. This occurs because regions with higher rates of digital patent filings indicate greater demand for data trading services, leading to the establishment of platforms to cater to this need. Therefore, the direction of causality may run from digital patent activity to the creation of data trading platforms, rather than the platforms driving innovation. The second issue is that the overall innovation ability of enterprises is improving over time. As a part of the overall innovation activity, digital innovation also improves over time, independent of the establishment of data trading platforms. The third issue is also the reverse causality problem. It may be that only provinces where certain innovation elements are enriching will establish data trading platforms, and the growth of digital innovation is not due to the establishment of data trading platforms but due to the enrichment of this innovation element. To strengthen our identification strategy, we conducted three dynamic DiD analyses in Section 5.2.1 and Section 5.2.2 to address these issues.

Additionally, we conduct two supplementary tests to assess the robustness and validity of our findings. In Section 5.2.3, we introduce a placebo test in which we randomly assign treatment and control statuses to the observations within our sample. Section 5.2.4 controls for other potentially related innovation policies of governments.

5.2.1 Firm-level dynamic DiD analysis

A potential problem with our DiD specification is the possibility of reverse causality, where data trading platforms might be established in provinces with already high levels of digital innovation. This is because regions with higher levels of digital development have a greater demand for data trading, leading to the establishment of a data trading platform to meet this demand. To address this concern, we investigate the dynamic treatment effects of the establishment of data trading platforms. If data trading platforms are established in response to firms' higher levels of digital innovation, we anticipate observing significant differences in digital innovation between the treatment and control groups prior to the inception of these platforms. To test this, we replace the treatment indicator in

Table 4 with a series of dummies representing the years around the establishment of data trading platforms. The results of this dynamic DiD analysis are shown in Figure 4. In Figure 4, *Treatment(0)* corresponds to the year in which the data trading platform is established. *Treatment(−τ)* and *Treatment(τ)* represent the years before and after the establishment of the data trading platforms. We set *Treatment(-1)* as the baseline. Note that the two-way fixed effects estimation of multi-period DiD is a weighted average of several different treatment effects, and the weight may be negative. In such a case, the average treatment effect obtained after weighted averaging may be opposite in direction to the true average treatment effect (Baker et al. (2022)). We also perform a series of heterogeneity robust DiD estimations following Butts and Gardner (2021) and Sun and Abraham (2021).

If the hypothesis that demand drives the establishment of data trading platforms is correct, we would expect to see a significant difference in digital innovation applications between the treatment and control groups even prior to the establishment of data trading platforms. However, as shown in Figure 4a, neither *Treatment(−2)* nor *Treatment(−3)* is statistically significant, and no pre-existing trends are evident. In contrast, the treatment indicators become statistically and economically significant after the data trading platforms are established (significant at the 5% level). Consequently, the findings in Figure 4a refute the demand-driven hypothesis and dismiss the reverse causality argument.

To address the second concern that other kinds of patents, not just digital innovation, has increased, we test for the number of green patent applications. Regarding green patents, we follow the method proposed by Haščič and Migotto (2015). Since Chinese patents are only classified under the International Patent Classification (IPC) system and lack the Cooperative Patent Classification (CPC) codes, we can only classify them based on the World Intellectual Property Organization (WIPO)'s definition of green patents using the IPC codes. The model setting is the same, except that we replace *Digital Others* with *Green Others* to control for the firm's overall level of innovation. As shown in Figure 4b, there is no significant difference in green patent applications between the treatment and control groups, either before or after the establishment of data trading platforms. This suggests that the establishment of data trading platforms specifically promotes digital innovation. It is consistent with our narrative that data resources are directed toward data-related innovation, rather than a general increase in innovation resources.

5.2.2 Province-level dynamic DiD analysis

Although no pre-existing trends were observed in the firm-level analysis, it is still possible that the establishment of data trading platforms might be associated with province-level characteristics, which could, in turn, potentially affect the local level of innovation (Li

2009). To address this concern, we conducted a dynamic DiD analysis of potential factors affecting digital innovation at the province level, as presented in Table 5.

The empirical setup for the regressions in Table 5 mirrors that of 5.2.1, but is conducted at the province-year level. The dependent variable in regression (1) is *Real GDPpc*, defined as per capita real GDP indexed to 2000 (one hundred thousand RMB). The dependent variable in regression (2) is *Social Consumption*, defined as total retail sales of consumer goods divided by regional GDP. In regression (3), *Fiscal Expenditure* is defined as province-level fiscal expenditures. These three variables represent the initial economic conditions of the provinces. In regressions (4) to (8), the dependent variables are *Industrialization level*, *Province R&D*, *Province Innovation*, *Informatization Level*, and *Technological Development*. Detailed definitions of the variables can be found in the Appendix A1. These variables represent provincial levels of industrialization, R&D spending, innovation, informatization, and technological development. All regressions control for population at the province level. Year fixed effects are included to capture aggregate time trends, and province fixed effects are included to control for any unobserved heterogeneity that remains constant over time at the provincial level.

Before and after the establishment of the data trading platform, there were no significant differences between the treated and control provinces in terms of regional economic conditions, industrialization, and technology levels. This lack of a prior trends at the province level refutes the reverse causality argument and reduces concerns that the establishment of data trading platforms might be associated with province-level characteristics. Moreover, the fact that provincial macroeconomic variables do not grow significantly after the establishment of the data trading platform suggests that the significant growth in digital innovation by firms after the establishment of the data trading platform cannot be attributed to growth in provincial macroeconomic variables.

5.2.3 Placebo test

To ensure the robustness of our results, we conduct a placebo test by utilizing pseudo-treatment individuals and pseudo-treatment timing simultaneously. Specifically, based on the earliest and latest establishment times of data trading platforms in the sample, pseudo-treatment timing for each province is randomly drawn from a uniform distribution within this interval for the Two-Way Fixed Effects (TWFE) estimation, repeated 500 times to derive the distribution of placebo effects. If the results in Table 4 are genuinely due to the establishment of data trading platforms and not random or influenced by other unobserved factors, we should not see similar outcomes in this artificially treated sample.

We re-estimate our baseline results using these pseudo-treated samples. Due to the naturally high correlation between digital patents and other patents, we do not include *Digital Others* variable in our simulation. Figure 5 shows the empirical distribution of the

estimates of the key regression coefficients ("Treatment"). The figure compares the true coefficient estimate with its empirical distribution and kernel density.The actual positive coefficient estimate is above the 99^{th} percentile of the distribution. Therefore, the results from this placebo test support the validity of our findings, suggesting they are unlikely to be caused by random chance.

5.2.4 Other digital policies

There may be concerns that our results are influenced by government digitization policies. The establishment of a data trading platform may be part of a government digitization policy, and an increase in the level of digital innovation in firms may not result from the establishment of a data trading platform, but from a government digitization policy. We try to rule out the effect of government digitization policies in two ways. First, the government's annual attention to digital issues is already included as a control variable in our main regression. Second, to control the impact of digital transformation or digital innovation policies in different provinces, we collected all policies at the provincial level across the country. We set up the first indicator called *digital policy* (1^{st}), which is used to capture the moment when terms like "digital transformation", "digital innovation", or "digital technology" first appear in provincial policy documents. The second indicator is the time of the first policy release related to the "Digital Economy Development 14th Five-Year Plan" of provincial policies, defined as *Digital Policy* (14^{th}*5-year)*.[8]

We reexamined the impact of the establishment of a data trading platform on firms' digital innovation, controlling for digital policies at the provincial level. As shown in Table 6, columns (1) and (2) show the results after controlling for *Digital Policy*(14^{th}*5-year)*, and columns (3) and (4) show the results after controlling for *Digital Policy*(1^{st}*)*. The results suggest that these policies do not have a significant impact on the development of digital innovation. Furthermore, the coefficient for data trading platforms remains significantly positive after including the policy variables. This suggests that the growth in firms' digital innovation is unlikely to have been brought about by other government digitization policies.

5.3 Distance Between Firms and data trading platforms

In this section, we examine the relationship between variations in the distance of firms within a province to the city where the trading platform is located and their innovation

[8]The "14th Five-Year Plan for Digital Economy Development" was issued by the central government in October 2021. It outlines the overall requirements, main tasks, key projects, and safeguard measures for the development of China's digital economy, providing an action guide for promoting digital economy development across regions and departments during the 14th Five-Year Plan period. Following this policy, provinces quickly formulated their own "14th Five-Year Plan" policies at the provincial level. For instance, in October 2021, Shanghai issued the "14th Five-Year Plan for Comprehensive Promotion of Urban Digital Transformation".

output. Our focus is on determining the differences in innovation growth across regions before and after the establishment of data trading platforms.

The most established theory for examining the impact of geographic distance on innovation spillovers is the learning region theory. This theory posits that tacit knowledge is not easily disseminated (Asheim 1996; Maskell and Malmberg 1999) because it represents a kind of experiential understanding of "knowing how to do things" that is characterized by context-specific attributes. The transmission of such knowledge occurs primarily through face-to-face interactions that rely on shared language and cultural codes. Tacit knowledge spillovers diminish as the spatial distance between interacting parties increases, eventually reaching a limit at which they almost disappear. Consequently, the knowledge embedded in a region, along with its intangible social capital and institutional assets (Maskell and Malmberg 1999), becomes less accessible to distant potential imitators, thereby maintaining the initial advantage of the leading region.

A substantial body of empirical research supports these theoretical views, showing that knowledge spillovers have regional characteristics and tend to decline with increasing distance. First, empirical studies based on patent citation data in Europe and the United States have found a high proportion of citations within the same region (Jaffe et al. 1993), with most cited patents originating from the same country or geographically close countries (Maurseth and Verspagen 2002; Paci and Usai 2009). Second, research on alliances among U.S. firms has also established the link between knowledge spillovers and geographic proximity, showing that closer geographic proximity significantly increases knowledge spillovers between firms (Arita and McCann 2000; Gomes-Casseres et al. 2006). Karlsson et al. (2004) propose that each point in geographic space has a certain "potential" accessibility to critical factors like local institutions, culture, information, personal contacts, and markets. By combining these accesses, certain locations emerge as more conducive to innovation than others.

The basic prediction of the literature is that the closer firms are to the factors needed to innovate, the higher the level of innovation. For the first data trading platform in each province, it is usually set up in a key city like the provincial capital or an economically developed one. Therefore, before discussing the differences in the impact of data trading platforms on enterprises at different distances, we first explore the relationship between a province's firms' innovation and their distance from the city where the first data trading platform is located. As shown in columns (1) and (2) of Table 7, there is a significant negative correlation, indicating that firms that are farther away from the city where that data trading platform is located prior to its establishment have a lower level of innovation, consistent with Maskell and Malmberg (1999), Arita and McCann (2000) and Gomes-Casseres et al. (2006). Subsequently, we add the interaction terms of *Treatment* and *Distance* to the regression to measure the heterogeneity of the impact of data trading platform establishment on firms' digital innovation. In columns (3) and (4), the

interaction term of *Treatment*Distance* is significantly positive, indicating that firms located farther from the city where the data trading platform is established achieve greater growth in their innovation level. The implication of these results is quite straightforward. Prior to the establishment of data trading platforms, data and knowledge exchange among firms could only be accomplished through face-to-face interactions. Consequently, firms in more central cities had more convenient information flow and easier access to innovation factors, which led to higher levels of digital innovation. However, firms in more remote cities encountered difficulties in utilizing their data, thereby reducing their data-driven innovation potential. With the establishment of data trading platforms, the platform network is optimized. Enterprises in remote areas can obtain data more conveniently. As a result, compared with enterprises in areas where data trading platform located, those in remote areas experience more significant growth in digital innovation.

6 Data trading platforms enhance firm's TFP through digital innovation

A considerable body of literature has focused on the impact of factor market development on total factor productivity (TFP). The existing research is primarily concerned with the losses in resource efficiency caused by distortions in factor markets. Studies have found that discrepancies in resource use efficiency between poor and developed countries lead to differences in TFP across countries. In China, the relatively late development of factor markets has led to price distortions, resulting in resource misallocation effects that significantly hinder improvements in firm or industry productivity (Hsieh and Klenow 2009). According to our endogenous growth model, the construction of data trading platforms enables the accelerated flow of data and the analysis of large amounts of information. It helps better realize the application value of data and improves the efficiency of resource utilization by promoting enterprise innovation. In summary, data transcends the competitive and scarce nature of traditional factors and has a distinctly different impact on productivity compared to traditional factors (Jones and Tonetti 2020a; Cong et al. 2021; Cong et al. 2022).

In this subsection, we examine whether data trading platforms have improved firms' total factor productivity through digital innovation. Considering that firms may timely adjust their factor inputs based on observable production efficiency, there exists a strong endogenous relationship between firm-level total factor productivity and factor inputs. This endogeneity often results in simultaneity bias when estimating production functions using ordinary least squares (OLS) methods. Therefore, this study uses the LP (Levinsohn and Petrin 2003) method in addition to the OP (Olley and Pakes 1996) method to calculate the total factor productivity of the firms. To examine the impact of the

development of the data market on firms' total factor productivity, we adopt a staggered DiD approach. The specific model is set as follows:

$$TFP_{i,t} = \alpha + \beta\, Treatment_{i,t} + \sum \gamma_k X_{k,i,t} + \theta\, Year_t + \delta\, Firm_i + \epsilon_{i,t}, \tag{33}$$

To test the impact of digital innovation on firms' total factor productivity, we regress firms' TFP on digital innovation with a lag of one period because there is a lag in the impact of innovation on firm efficiency (Babina et al. (2024)):

$$TFP_{i,t} = \alpha + \beta\, Digital\ Innovation_{i,t-1} + \sum \gamma_k X_{k,i,t} + \theta\, Year_t + \delta\, Firm_i + \epsilon_{i,t}, \tag{34}$$

The regression results are presented in Table 8. Looking at the estimation results in Table 8A, it can be observed that irrespective of whether control variables are included, the coefficients of the explanatory variable *Treatment* for total factor productivity (*TFP*) estimated using both methods are significantly positive. This clearly suggests that the establishment of data trading platforms remarkably promotes the growth of corporate productivity. Turning to Table 8B, it becomes evident that the regression coefficients of firms' total factor productivity (*TFP_OP*, *TFP_LP*) on digital innovation are positive and significant at the 1% level.[9] This suggests that, on average, firms that engage in digital innovation have higher total factor productivity. The results presented above convincingly demonstrate that the construction of data trading platforms facilitates the flow of valuable data resources and incentivizes firms to engage in more digital innovation activities. This increase in digital innovation, in turn, raises total factor productivity.

7 Conclusion

Our study examines the impact of data marketization on firm innovation and productivity in China. By constructing endogenous growth models and conducting empirical analyses, we find that data trading platforms significantly enhance firms' digital innovation and total factor productivity (TFP).

Theoretically, the introduction of data intermediaries in our model shows increased innovation rates and TFP compared to economies without them. Empirically, using the staggered establishment of data trading platforms in China as a quasi-experiment, we find that these platforms lead to a significant increase in firms' digital patent filings. Additionally, we find that data trading platforms have a more pronounced impact on firms in remote areas, reducing the innovation gap between regions.

[9]The coefficients are very small because "digital innovation" is measured using the absolute number of patents related to digital technology. If logarithmic transformation is applied, it will increase the scale of the coefficients.

China's data trading platforms, as core infrastructure projects, to some extent, reflect the effectiveness of the data market and impose exogenous effects on enterprises as market participants. The analysis of these effects enriches theoretical research on innovation incentives in the context of data factor marketization and provides general insights for the development of the digital economy. The development of a unified and well-functioning data market is an important means to overcome the challenges associated with the flow and use of data, thereby enhancing the growth and innovation capabilities of the digital economy.

References

Acemoglu, Daron and Pascual Restrepo, "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American economic review*, 2018, *108* (6), 1488–1542.

Ahn, Sang-Jin, "Three characteristics of technology competition by IoT-driven digitization," *Technological Forecasting and Social Change*, 2020, *157*, 120062.

Akcigit, Ufuk and Qingmin Liu, "The role of information in innovation and competition," *Journal of the European Economic Association*, 2016, *14* (4), 828–870.

Annarelli, Alessandro, Cinzia Battistella, Fabio Nonino, Vinit Parida, and Elena Pessot, "Literature review on digitalization capabilities: Co-citation analysis of antecedents, conceptualization and consequences," *Technological Forecasting and Social Change*, 2021, *166*, 120635.

Ardito, Lorenzo, Angelo Natalicchio, Francesco Paolo Appio, and Antonio Messeni Petruzzelli, "The role of scientific knowledge within inventing teams and the moderating effects of team internationalization and team experience: Empirical tests into the aerospace sector," *Journal of Business Research*, 2021, *128*, 701–710.

Arita, Tomokazu and Philip McCann, "Industrial alliances and firm location behaviour: some evidence from the US semiconductor industry," *Applied Economics*, 2000, *32* (11), 1391–1403.

Asheim, Bjrn T., "Industrial districts as 'learning regions': A condition for prosperity," *European Planning Studies*, 1996, *4* (4), 379–400.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson, "Artificial intelligence, firm growth, and product innovation," *Journal of Financial Economics*, 2024, *151*, 103745.

Bai, Chong-En, Yingjuan Du, Zhigang Tao, and Sarah Y Tong, "Local protectionism and regional specialization: evidence from China's industries," *Journal of international economics*, 2004, *63* (2), 397–417.

Baker, Andrew C, David F Larcker, and Charles CY Wang, "How much should we trust staggered difference-in-differences estimates?," *Journal of Financial Economics*, 2022, *144* (2), 370–395.

Beck, Thorsten, Ross Levine, and Alexey Levkov, "Big Bad Banks? The Winners and Losers from Bank Deregulation in the United States," *The Journal of Finance*, 2010, *65* (5), 1637–1667.

Beltagui, Ahmad, Ainurul Rosli, and Marina Candi, "Exaptation in a digital innovation ecosystem: The disruptive impacts of 3D printing," *Research policy*, 2020, *49* (1), 103833.

Benassi, Mario, Elena Grinza, and Francesco Rentocchini, "The rush for patents in the Fourth Industrial Revolution," *Journal of Industrial and Business Economics*, 2020, *47*, 559–588.

Bergemann, Dirk and Alessandro Bonatti, "Markets for information: An introduction," *Annual Review of Economics*, 2019, *11* (1), 85–107.

Bertello, Alberto, Alberto Ferraris, Stefano Bresciani, and Paola De Bernardi, "Big data analytics (BDA) and degree of internationalization: the interplay between governance of BDA infrastructure and BDA capabilities," *Journal of Management and Governance*, 2021, *25* (4), 1035–1055.

Bolton, Patrick, Marcin T Kacperczyk, and Moritz Wiedemann, "The co2 question: Technical progress and the climate crisis," *Available at SSRN*, 2023, *4212567*.

Boudreau, Kevin, "Open platform strategies and innovation: Granting access vs. devolving control," *Management science*, 2010, *56* (10), 1849–1872.

Butts, Kyle and John Gardner, "{did2s}: Two-stage difference-in-differences," *arXiv preprint arXiv:2109.05913*, 2021.

Cheng, Cong and Limin Wang, "How companies configure digital innovation attributes for business model innovation? A configurational view," *Technovation*, 2022, *112*, 102398.

Coase, R. H., "The Nature of the Firm," *Economica*, 1937, *4* (16), 386–405.

Cohen, Lauren, Umit G Gurun, and Quoc H Nguyen, "The ESG-innovation disconnect: Evidence from green patenting," Technical Report, National Bureau of Economic Research 2020.

Cohn, Jonathan B, Zack Liu, and Malcolm I Wardlaw, "Count (and count-like) data in finance," *Journal of Financial Economics*, 2022, *146* (2), 529–551.

Cong, Lin William, Danxia Xie, and Longtian Zhang, "Knowledge Accumulation, Privacy, and Growth in a Data Economy," *Management Science*, 2021, *67* (10), 6480–6492.

_ , Wenshi Wei, Danxia Xie, and Longtian Zhang, "Endogenous growth under multiple uses of data," *Journal of Economic Dynamics and Control*, 2022, *141*, 104395.

Corvello, Vincenzo, Jaroslav Belas, Carlo Giglio, Gianpaolo Iazzolino, and Ciro Troise, "The impact of business owners' individual characteristics on patenting in the context of digital innovation," *Journal of Business Research*, 2023, *155*, 113397.

Eckardt, Martina and Wolfgang Kerber, "Property rights theory, bundles of rights on IoT data, and the EU Data Act," *European Journal of Law and Economics*, 2024, *57* (1), 113–143.

Farboodi, Maryam and Laura Veldkamp, "A model of the data economy," Technical Report, National Bureau of Economic Research Cambridge, MA, USA 2021.

_ **and** _ , "Data and markets," *Annual Review of Economics*, 2023, *15* (1), 23–40.

Favot, Marinella, Leyla Vesnic, Riccardo Priore, Andrea Bincoletto, and Fabio Morea, "Green patents and green codes: How different methodologies lead to different results," *Resources, Conservation & Recycling Advances*, 2023, *18*, 200132.

Glinkina, Olga Vladimirovna, Svetlana Aleksandrovna Ganina, Anna Viktorovna Maslennikova, Tatyana Anatolevna Solostina, and Marina ViktorovnaSoloveva, "Digital changes in the economy: advanced opportunities for digital innovation," *Int. J. Manag*, 2020, *11* (3).

Goldfarb, Avi and Catherine Tucker, "Digital economics," *Journal of economic literature*, 2019, *57* (1), 3–43.

Gomes-Casseres, Benjamin, John Hagedoorn, and Adam B. Jaffe, "Do alliances promote knowledge flows?," *Journal of Financial Economics*, 2006, *80* (1), 5–33.

Han, Pengfei, Chunrui Liu, and Xuan Tian, "Does Trading Spur Specialization? Evidence from Patenting," 2024.

Haščič, Ivan and Mauro Migotto, "Measuring environmental innovation using patent data," 2015.

Hege, Ulrich, Kai Li, and Yifei Zhang, "Climate Innovation and Carbon Emissions: Evidence from Supply Chain Networks," *Available at SSRN 4557447*, 2023.

_ **, Sebastien Pouget, and Yifei Zhang**, "Climate Patents and Financial Markets," 2024.

Hess, Thomas, Christian Matt, Alexander Benlian, and Florian Wiesböck, "Options for formulating a digital transformation strategy.," *MIS Quarterly Executive*, 2016, *15* (2).

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Hunt, Jennifer, Iain M Cockburn, and James Bessen, "Is Distance from Innovation a Barrier to the Adoption of Artificial Intelligence?," Technical Report, National Bureau of Economic Research 2024.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, 1993, *108* (3), 577–598.

Johansson, Sara, Malin Kullström, Jennie Björk, Anna Karlsson, and Susanne Nilsson, "Digital production innovation projects–The applicability of managerial controls under high levels of complexity and uncertainty," *Journal of manufacturing technology management*, 2020, *32* (3), 772–794.

Jones, Charles I. and Christopher Tonetti, "Nonrivalry and the Economics of Data," *American Economic Review*, 2020, *110* (9), 2819–2858.

Jones, Charles I and Christopher Tonetti, "Nonrivalry and the Economics of Data," *American Economic Review*, 2020, *110* (9), 2819–2858.

Karlsson, Charlie, Per Flensburg, and Sven-Åke Hörte, *Knowledge Spillovers and Knowledge Management*, Edward Elgar Publishing, 2004.

Kerber, Wolfgang, "Specifying and Assigning'Bundles of Rights' on Data: An Economic Perspective," *Available at SSRN 3847620*, 2021.

Kivimaa, Paula and Florian Kern, "Creative destruction or mere niche support? Innovation policy mixes for sustainability transitions," *Research Policy*, February 2016, *45* (1), 205–217.

Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 2003, *70* (2), 317–341.

Li, Quan, Huimin Chen, Yang Chen, Tong Xiao, and Li Wang, "Digital economy, financing constraints, and corporate innovation," *Pacific-Basin Finance Journal*, 2023, *80*, 102081.

Li, Xibao, "China's regional innovation capacity in transition: An empirical approach," *Research policy*, 2009, *38* (2), 338–357.

Maskell, P and A Malmberg, "Localised learning and industrial competitiveness," *Cambridge Journal of Economics*, 1999, *23* (2), 167–185.

Maurseth, Per Botolf and Bart Verspagen, "Knowledge Spillovers in Europe: A Patent Citations Analysis," *The Scandinavian Journal of Economics*, 2002, *104* (4), 531–545.

Müller, Oliver, Maria Fay, and Jan Vom Brocke, "The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics," *Journal of management information systems*, 2018, *35* (2), 488–509.

Nambisan, Satish, Kalle Lyytinen, Ann Majchrzak, and Michael Song, "Digital innovation management," *MIS quarterly*, 2017, *41* (1), 223–238.

Olley, G Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–1297.

Opland, Leif Erik, Ilias O Pappas, Jostein Engesmo, and Letizia Jaccheri, "Employee-driven digital innovation: A systematic review and a research agenda," *Journal of Business Research*, 2022, *143*, 255–271.

Paci, Raffaele and Stefano Usai, "Knowledge flows across European regions," *The Annals of Regional Science*, 2009, *43* (3), 669–690.

Pan, Xia, Wenyin Cheng, Yuning Gao, Tomas Balezentis, and Zhiyang Shen, "Is environmental regulation effective in promoting the quantity and quality of green innovation?," *Environmental Science and Pollution Research*, 2021, *28*, 6232–6241.

Peng, Zihao, Yiqi Huang, Luanhong Liu, Wanli Xu, and Xihong Qian, "How government digital attention alleviates enterprise financing constraints: An enterprise digitalization perspective," *Finance Research Letters*, 2024, *67*, 105883.

Poncet, Sandra, "A fragmented China: Measure and determinants of Chinese domestic market disintegration," *Review of international Economics*, 2005, *13* (3), 409–430.

Romer, Paul M, "Endogenous technological change," *Journal of political Economy*, 1990, *98* (5, Part 2), S71–S102.

Rosenkopf, Lori and Paul Almeida, "Overcoming local search through alliances and mobility," *Management science*, 2003, *49* (6), 751–766.

Scott, Susan V, John Van Reenen, and Markos Zachariadis, "The long-term effect of digital innovation on bank performance: An empirical study of SWIFT adoption in financial services," *Research Policy*, 2017, *46* (5), 984–1004.

Silva, JMC Santos and Silvana Tenreyro, "The log of gravity," *The Review of Economics and statistics*, 2006, *88* (4), 641–658.

Singh, J. and M. Marx, "Geographic constraints on knowledge spillovers: political borders vs. spatial proximity," *Management Science*, 2013, *59*, 2056–2078.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of econometrics*, 2021, *225* (2), 175–199.

Veldkamp, Laura, "Slow boom, sudden crash," *Journal of Economic theory*, 2005, *124* (2), 230–257.

_ , "Valuing data as an asset," *Review of Finance*, 2023, *27* (5), 1545–1562.

_ **and Cindy Chung**, "Data and the aggregate economy," *Journal of Economic Literature*, 2024, *62* (2), 458–484.

Verhoef, Peter C, Thijs Broekhuizen, Yakov Bart, Abhi Bhattacharya, John Qi Dong, Nicolai Fabian, and Michael Haenlein, "Digital transformation: A multidisciplinary reflection and research agenda," *Journal of business research*, 2021, *122*, 889–901.

Wamba, Samuel Fosso, Angappa Gunasekaran, Shahriar Akter, Steven Ji fan Ren, Rameshwar Dubey, and Stephen J Childe, "Big data analytics and firm performance: Effects of dynamic capabilities," *Journal of business research*, 2017, *70*, 356–365.

Wang, Dingqing, Hongwei Liao, Aiguo Liu, and Dongdong Li, "Natural resource saving effects of data factor marketization: Implications for green recovery," *Resources Policy*, 2023, *85*, 104019.

Wang, Juite and Chih-Chi Hsu, "A topic-based patent analytics approach for exploring technological trends in smart manufacturing," *Journal of Manufacturing Technology Management*, 2021, *32* (1), 110–135.

Wei, Shaobo, Dabao Xu, and Hua Liu, "The effects of information technology capability and knowledge base on digital innovation: the moderating role of institutional environments," *European Journal of Innovation Management*, 2022, *25* (3), 720–740.

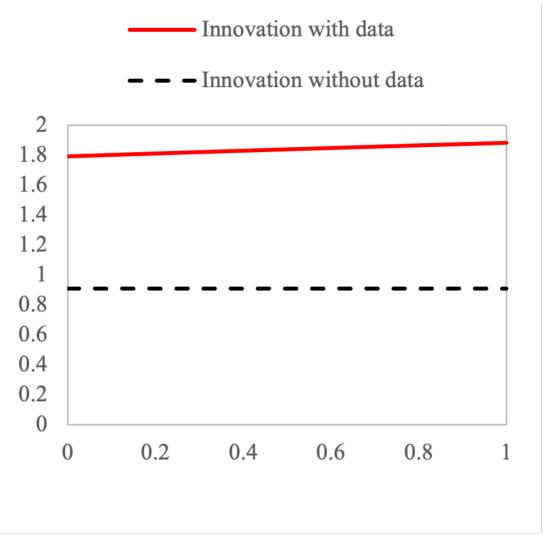
Yang, Chih-Hai, "How artificial intelligence technology affects productivity and employment: firm-level evidence from Taiwan," *Research Policy*, 2022, *51* (6), 104536.

Zhang, Ningning, Dingyi You, Le Tang, and Ke Wen, "Knowledge path dependence, external connection, and radical inventions: Evidence from Chinese Academy of Sciences," *Research Policy*, 2023, *52* (4), 104738.

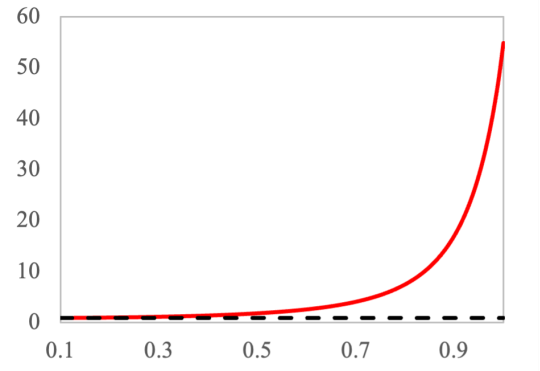
Figures

Figure 1: Innovations for Different Parameters

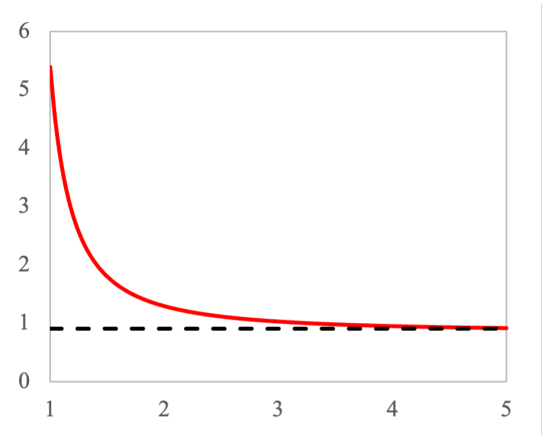
This figure illustrates the growth rate of innovation of two economies under different parameter settings. The definitions and value references of parameters α, β, κ, σ, ρ, and η are shown in Table 2. Each subfigure shows the relationship between the innovation rate of the economy and its corresponding parameter, while other parameters remain unchanged.



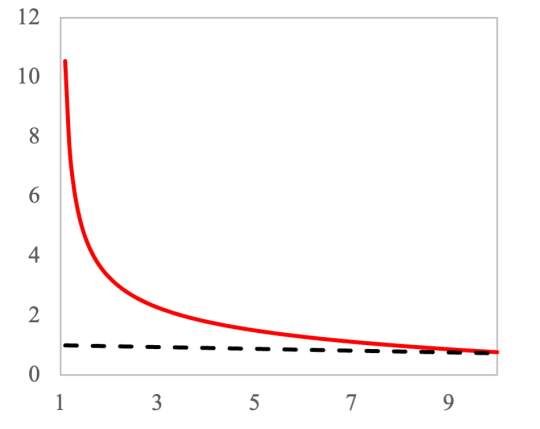
(a) α



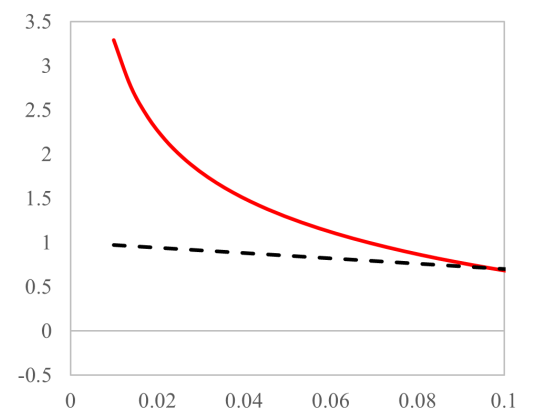
(b) β



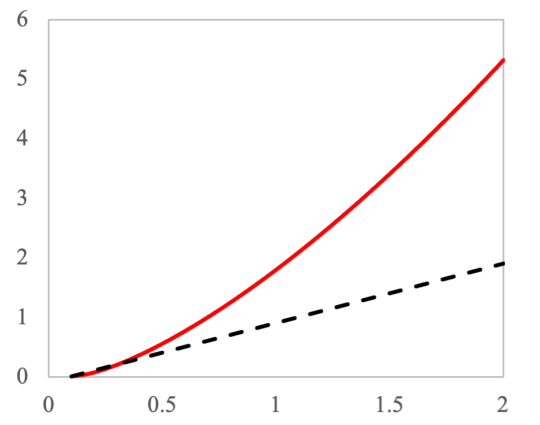
(c) κ



(d) σ



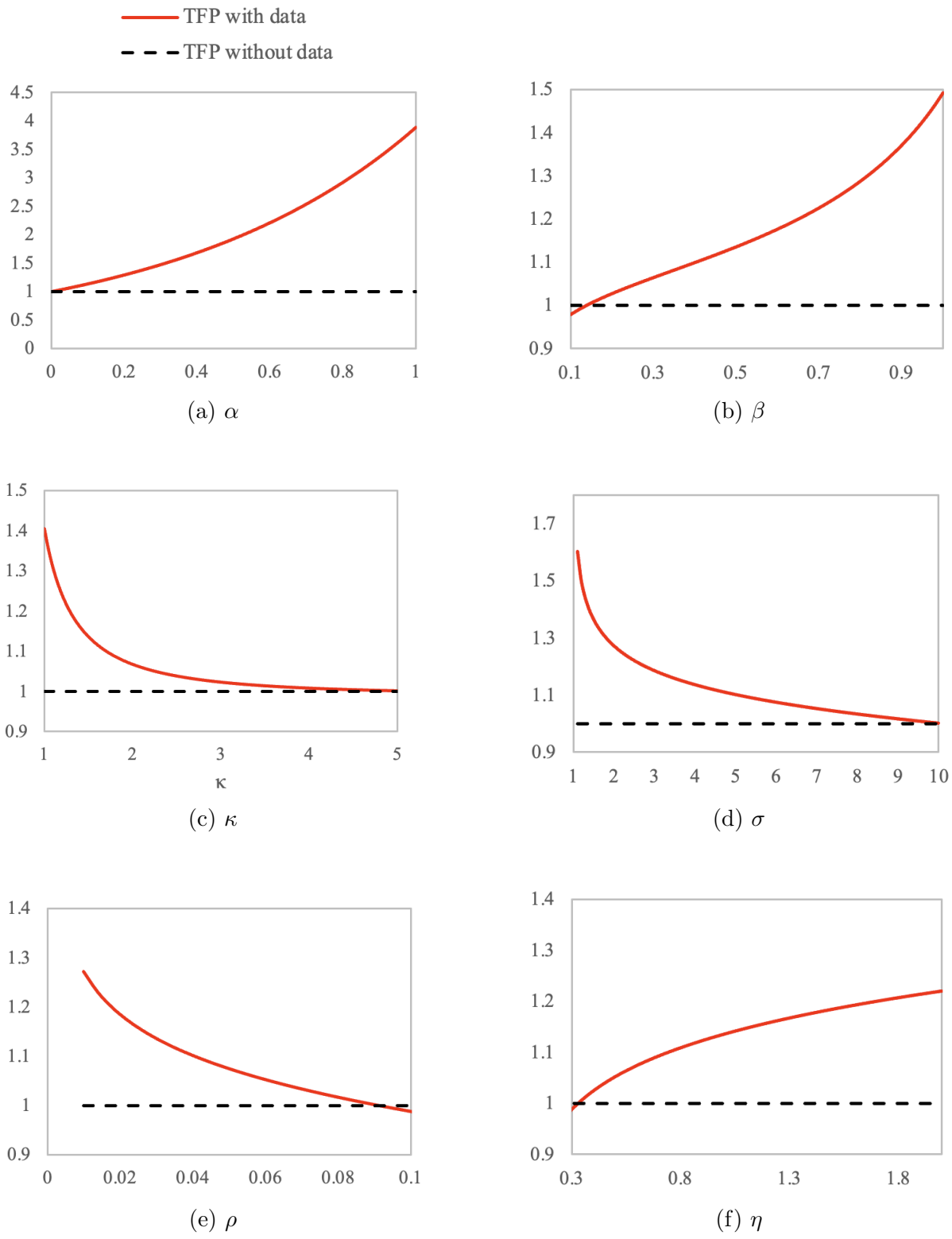
(e) ρ



(f) η

Figure 2: TFP for Different Parameters

This figure illustrates the total factor productivity （TFP） of two economies under different parameter settings. The definitions and value references of parameters α, β, κ, σ, ρ, and η are shown in Table 2. Each subfigure shows the relationship between the total factor productivity of the economy and its corresponding parameter while other parameters remain unchanged.



(a) α (b) β

(c) κ (d) σ

(e) ρ (f) η

Figure 3: Trends in Digital Patent Applications (2008-2022)

This figure illustrates the annual counts of four categories of digital patent applications. Blue bars represent patents related to the digital product manufacturing industry, which primarily includes computer manufacturing, communication equipment manufacturing, and electronic component manufacturing. Orange bars correspond to patents in the digital product service industry, covering activities such as wholesale, retail, leasing, and maintenance of digital products. Green bars denote patents in the digital technology application industry, focusing on software development, Internet services, and information technology services. Finally, red bars show patents in the data element-driven industry, which involves property rights transactions of data resources and the construction of information infrastructure.

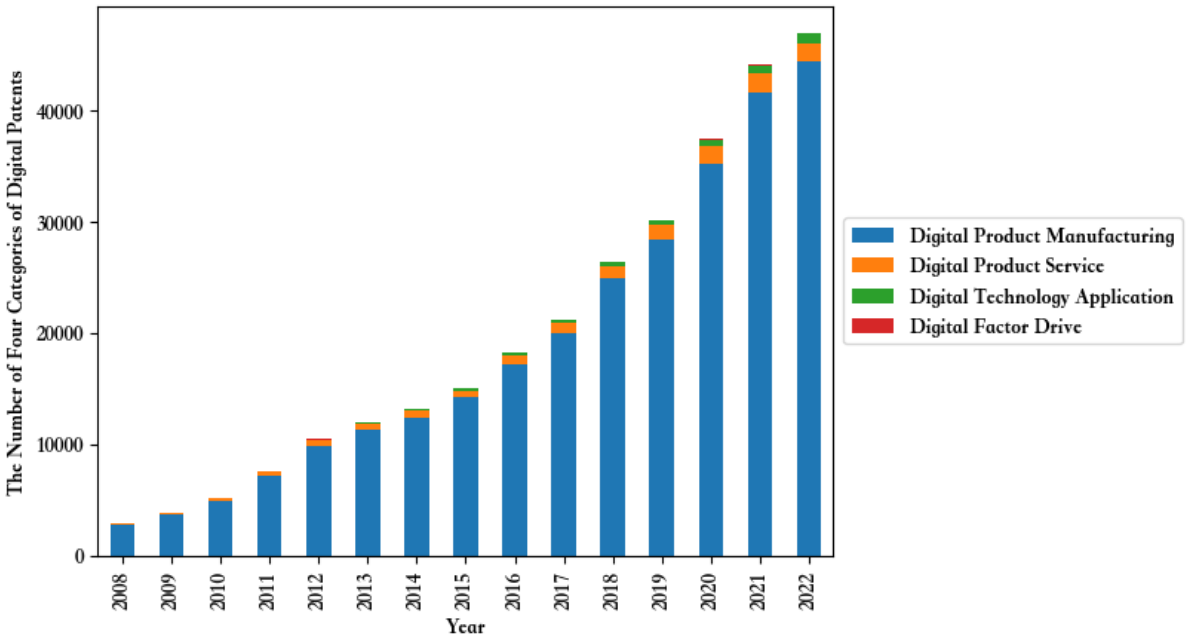
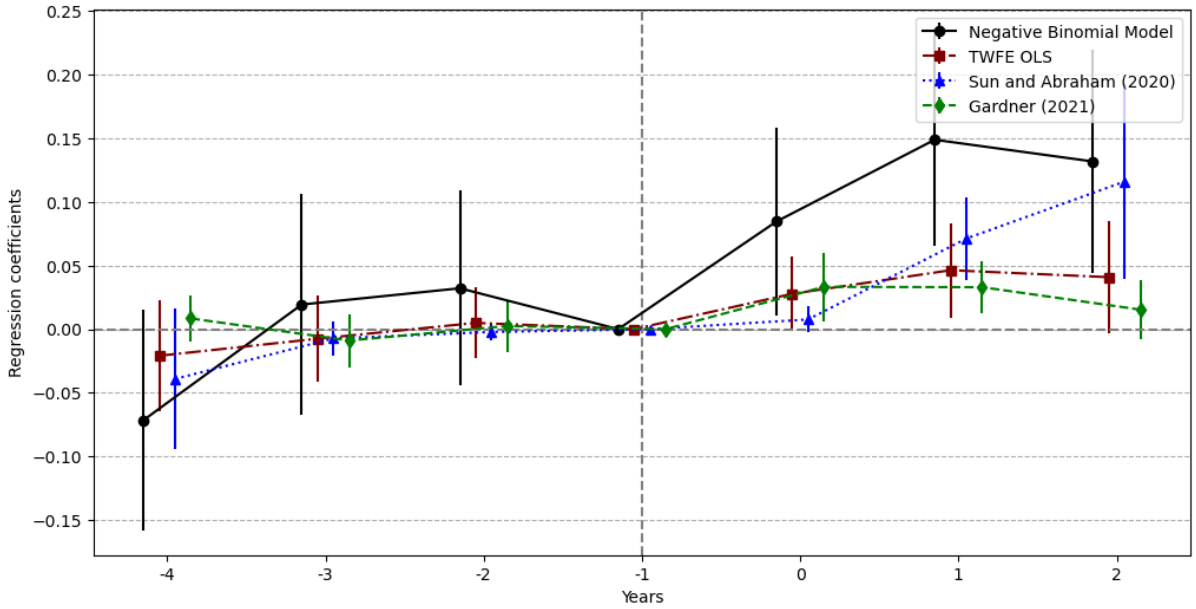
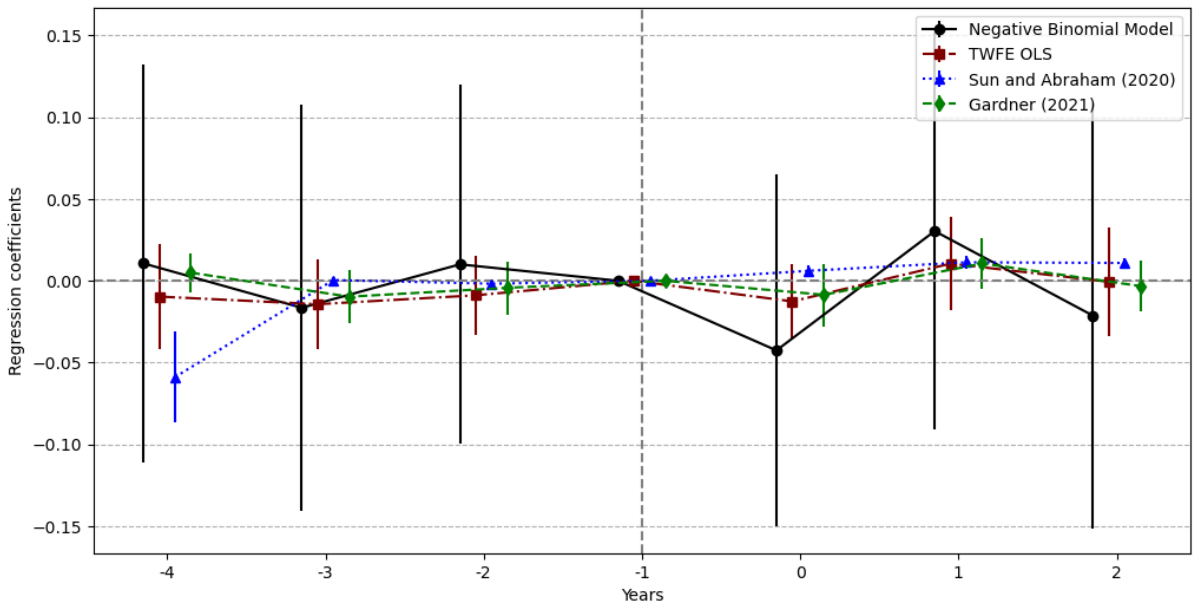


Figure 4: Dynamic DiD analysis at the firm level

This figure shows the coefficients and confidence intervals of the dynamic difference-in-differences (DiD) analysis comparing digital patents with green patents. The horizontal axis represents the years before and after the establishment of data trading platforms. *-1* represents the baseline year, *0* is the year when the data trading platform was established, *-4* represents the fourth year before the establishment of the data trading platform and earlier years, *-3* represents the third year before the establishment of the data trading platform, and so on. The coefficients of digital patents are represented in the first plot, while those of green patents are shown in the second plot. The control variables are delineated in Section 4.2.3. The regression includes firm fixed effects and year fixed effects.



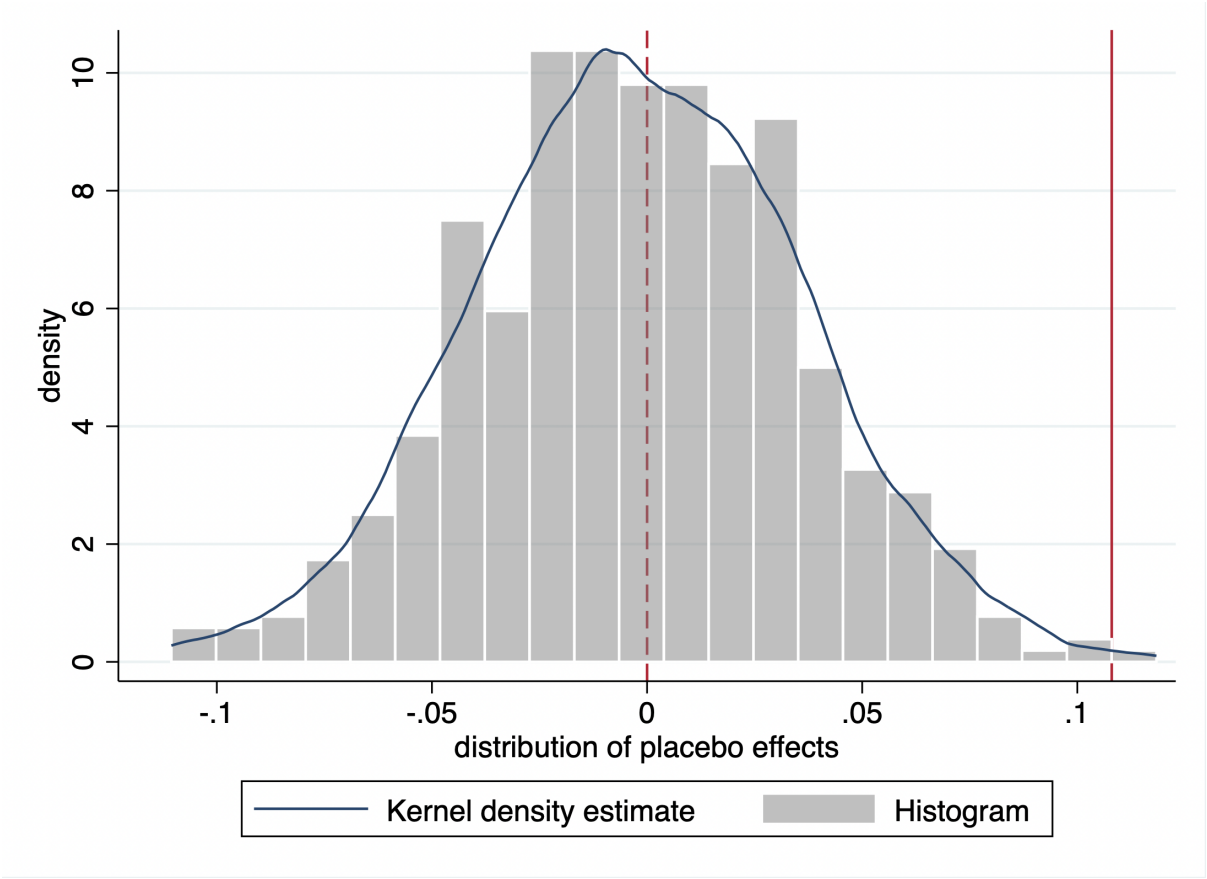
(a) Digital Patents



(b) Green Patents

Figure 5: Placebo Test

In this figure, we conducted a placebo test for Model 2 in Table 4. This figure represents the unrestricted mixed placebo test, utilizing pseudo-treatment individuals and pseudo-treatment timing simultaneously. Specifically, based on the earliest and latest establishment times of data trading platforms in the sample, pseudo-treatment timing for each province is randomly drawn from a uniform distribution within this interval for TWFE estimation, repeated 500 times to derive the distribution of placebo effects. We plot the kernel density of the coefficient estimates, and the true coefficient estimate is marked by a red vertical line. The two-sided P-value is 0.006, and the right-sided P-value is 0.002, both of which are statistically significant at the 1% level.



Tables

Table 1: Treated Province-level Regions and Treatment Time

This table shows the location and establishment year of the data trading platform in each province. The first column lists the provinces where the data trading platforms were established. The second column indicates the specific locations, and the third column provides the starting year of the treatment event.

Treated Province-level Regions	Location	Treatment Starting Year
Beijing	Beijing	2014
Chongqing	Chongqing	2015
Guizhou	Guiyang	2015
Hubei	Wuhan	2015
Jiangsu	Yancheng	2015
Hebei	Shijiazhuang	2015
Heilongjiang	Harbin	2016
Shaanxi	Xi'an	2016
Shanghai	Shanghai	2016
Zhejiang	Hangzhou	2016
Guangdong	Guangzhou	2016
Xinjiang	Urumqi	2016
Gansu	Lanzhou	2016
Henan	Zhengzhou	2017
Anhui	Huainan	2017
Shandong	Qingdao	2017
Jilin	Changchun	2018
Shanxi	Taiyuan	2020
Guangxi	Nanning	2020
Tianjin	Tianjin	2021
Hainan	Haikou	2021
Sichuan	Deyang	2021
Hunan	Changsha	2022
Fujian	Fuzhou	2022

Table 2: Baseline Parameters for Quantitative Analyses

This table provides the values of the parameters for the simulation part of Section 3. The general values of the parameters are given in Cong et al. (2021) and Cong et al. (2022).

Variable	Description	Value
α	Contribution of data in production sector	0.1
β	Contribution of data in innovation sector	0.5
κ	Severity of consumers' privacy concern	1.5
σ	Reciprocal of elasticity of intertemporal substitution	4
ρ	Subjective discount factor	0.03
η	Innovation efficiency	1
L	Population level	1
A	Fixed quality of ideas	1

Table 3: Descriptive Statistics

This table reports the summary statistics of the main variables as defined in Table A1. Because not all provinces have established data trading platforms, the number of observations for *Distance* is slightly reduced. All the potentially unbounded variables are winsorized at the 1% extremes.

VarName	Obs	Mean	SD	Min	Median	Max
Digital Innovation	24418	11.282	28.252	0.000	1.000	162.000
DPM	24418	10.484	26.453	0.000	1.000	152.000
DPS	24418	0.158	0.588	0.000	0.000	4.000
DTA	24418	0.439	1.462	0.000	0.000	9.000
DFD	24418	0.019	0.136	0.000	0.000	1.000
Digital Others	24418	17.560	44.782	0.000	0.000	376.000
Green Innovation	24418	1.578	4.274	0.000	0.000	25.000
Green Others	24418	27.466	63.747	0.000	4.000	377.000
Gov Attention	24418	0.644	0.590	0.000	0.693	2.079
Distance	20502	2.726	2.400	0.000	4.241	6.722
TFP_LP	24418	8.295	1.014	1.562	8.197	12.691
TFP_OP	24418	6.655	0.832	1.605	6.560	11.174
Age	24418	2.852	0.352	1.386	2.890	3.497
Growth	24418	0.168	0.351	-0.602	0.118	2.605
Leverage	24418	0.417	0.197	0.059	0.411	0.887
Size	24418	22.226	1.291	19.242	22.022	26.057
Employee	24418	7.822	1.240	2.833	7.712	13.253
Market Value	24418	2.052	1.260	0.858	1.646	8.204
Profitability	24418	0.069	0.161	-1.052	0.069	0.510
Liquidity	24418	2.434	2.282	0.306	1.681	13.974
Cash Flow	24418	0.051	0.067	-0.177	0.049	0.261
Board Size	24418	2.129	0.199	1.609	2.197	2.708
Ownership	24418	0.350	0.477	0.000	0.000	1.000
Firm R&D	24418	4.543	4.754	0.000	3.590	24.990
Real GDPpc	465	0.108	0.052	0.049	0.089	0.288
Social Consumption	465	0.372	0.066	0.224	0.374	0.523
Fiscal Expenditure	465	0.268	0.188	0.105	0.220	1.194
Industrialization Level	465	0.323	0.094	0.078	0.335	0.508
Province R&D	465	0.016	0.011	0.002	0.013	0.060
Province Innovation	465	9.220	1.662	4.522	9.387	12.222
Informationization Level	465	0.062	0.048	0.015	0.041	0.237
Technological Development	465	225.056	134.210	1.000	225.000	453.000
Population	465	451.536	684.886	2.378	277.941	3925.870

Table 4: Data Trading Platforms and Digital Innovation

This table reports the main estimation results on the effects of data trading platforms. A negative binomial model is used. The variables are defined in Table A1. Robust standard errors, adjusted using the OIM method, are reported in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	*Digital Innovation*		*DPM*	*DPS*	*DTA*	*DFD*
Treatment	0.135***	0.061**	0.057**	0.151*	0.204***	-0.421*
	(0.031)	(0.029)	(0.029)	(0.091)	(0.067)	(0.223)
Gov Attention		0.015	0.019	-0.078*	-0.022	0.060
		(0.015)	(0.015)	(0.045)	(0.035)	(0.113)
Digital Others		0.008***	0.008***	0.006***	0.006***	0.007***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Age		-0.128**	-0.104*	0.063	-0.112	-1.129
		(0.054)	(0.055)	(0.284)	(0.166)	(0.832)
Growth		-0.012	-0.024	-0.073	0.084	-0.071
		(0.023)	(0.023)	(0.082)	(0.056)	(0.190)
Leverage		-0.233***	-0.241***	-0.715**	-0.418*	0.640
		(0.086)	(0.087)	(0.329)	(0.222)	(0.804)
Size		-0.047**	-0.042**	0.179**	0.119**	0.483**
		(0.020)	(0.021)	(0.079)	(0.050)	(0.207)
Employee		0.118***	0.115***	0.230***	0.107**	0.007
		(0.019)	(0.019)	(0.076)	(0.046)	(0.191)
Market Value		-0.014*	-0.013	0.018	0.006	0.120*
		(0.008)	(0.008)	(0.029)	(0.018)	(0.066)
Profitability		0.350***	0.351***	0.278	0.301**	0.224
		(0.063)	(0.063)	(0.218)	(0.142)	(0.539)
Liquidity		0.003	0.001	-0.003	0.006	0.010
		(0.006)	(0.006)	(0.027)	(0.014)	(0.060)
Cash Flow		-0.187	-0.153	-0.448	-0.671**	0.867
		(0.129)	(0.130)	(0.433)	(0.306)	(1.022)
Board Size		-0.025	-0.021	-0.066	0.039	0.230
		(0.057)	(0.058)	(0.191)	(0.137)	(0.498)
Ownership		-0.071**	-0.069**	0.254*	-0.099	-0.200
		(0.034)	(0.034)	(0.140)	(0.093)	(0.490)
Firm R&D		0.024***	0.024***	0.024**	0.027***	0.040
		(0.003)	(0.003)	(0.010)	(0.006)	(0.029)
Constant	-1.241***	-0.647*	-0.761**	-7.508***	-4.166***	4.549
	(0.070)	(0.362)	(0.370)	(1.563)	(0.939)	(622.966)
Observations	24,418	24,418	24,073	8,440	9,910	3,327
Year fixed-effect	YES	YES	YES	YES	YES	YES
Firm fixed-effect	YES	YES	YES	YES	YES	YES
Control variables	NO	YES	YES	YES	YES	YES
Log-Likelihood Full Model	-43835	-42416	-41062	-4166	-7719	-851.9
Wald Chi2	3434	8865	8633	1029	1017	131.1
Prob >Chi2	0	0	0	0	0	0

Table 5: Dynamic DiD analysis at the province level

We report the results of the dynamic DiD analysis at the province level in this table. *Treatment(0)* is defined with respect to the year when the data trading platform is established. *Treatment(−τ)* and *Treatment(τ)* correspond to τ years before and after the establishment of data trading platforms, respectively. The dependent variables are defined in Table A1. Columns (1) to (4) mainly comprise economic variables at the provincial level, while columns (5) to (8) mainly consist of variables related to technology and innovation. The control variables are described in Section 4.2.3. The regression includes year and province fixed effects. Standard errors are clustered at the province level and reported in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	*Real GDPpc*	*Social consumption level*	*Fiscal expenditure*	*Industrialization level*	*Province R&D*	*Province innovation*	*Informationization level*	*Technological development level*
Treatment(-4+)	-0.000	0.016	-0.006	0.022	-0.001	0.143	0.000	-51.239**
	(0.001)	(0.017)	(0.009)	(0.015)	(0.001)	(0.135)	(0.004)	(18.762)
Treatment(-3)	0.001	0.001	-0.011*	0.010	-0.000	0.162	-0.005	-31.696**
	(0.001)	(0.013)	(0.006)	(0.008)	(0.000)	(0.101)	(0.004)	(14.147)
Treatment(-2)	0.001	0.003	-0.010**	0.008*	-0.000	0.085	-0.007*	-11.349
	(0.001)	(0.009)	(0.004)	(0.004)	(0.000)	(0.069)	(0.004)	(7.668)
Treatment(0)	-0.000	0.010	0.002	-0.005	0.000	-0.017	0.002	-8.215
	(0.001)	(0.009)	(0.003)	(0.004)	(0.000)	(0.051)	(0.003)	(9.186)
Treatment(1)	0.001	0.012	0.002	-0.003	-0.001**	-0.020	-0.001	3.421
	(0.001)	(0.012)	(0.007)	(0.009)	(0.001)	(0.073)	(0.004)	(15.739)
Treatment(2)	0.002	0.007	-0.003	-0.006	-0.001	-0.059	-0.007	-1.578
	(0.002)	(0.015)	(0.010)	(0.013)	(0.001)	(0.074)	(0.005)	(20.642)
Constant	0.127***	0.324***	0.297***	0.348***	0.004**	9.512***	0.045***	307.025***
	(0.009)	(0.049)	(0.016)	(0.032)	(0.002)	(0.236)	(0.006)	(37.155)
Observations	395	395	395	395	395	395	395	395
R-squared	0.995	0.733	0.987	0.910	0.976	0.955	0.925	0.872
Year fixed-effect	YES	YES	YES	YES	YES	YES	YES	YES
Province fixed-effect	YES	YES	YES	YES	YES	YES	YES	YES
Control variable	YES	YES	YES	YES	YES	YES	YES	YES

Table 6: Digital Policies and Digital Innovation

We examine the relationship between digital policies and digital innovation in this table. A negative binomial model is used. Columns (1) and (2) present results for *Digital Policy(14^{th}5-year)*, while columns (3) and (4) present results for *Digital Policy(1^{st})*. The odd-numbered columns exclude control variables, whereas the even-numbered columns include them. The variables are defined in Table A1, and the control variables are described in Section 4.2.3. Robust standard errors, adjusted using the OIM method, are reported in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

	(1)	(2)	(3)	(4)
VARIABLES	*Digital Innovation*			
Treatment	0.135***	0.061**	0.133***	0.062**
	(0.031)	(0.029)	(0.031)	(0.029)
Digital Policy(14^{th}5-year)	-0.031	-0.041		
	(0.032)	(0.030)		
Digital Policy(1^{st})			0.035	-0.008
			(0.027)	(0.025)
Constant	-1.241***	-0.650*	-1.241***	-0.644*
	(0.070)	(0.362)	(0.070)	(0.363)
Observations	24,418	24,418	24,418	24,418
Year fixed-effect	YES	YES	YES	YES
Firm fixed-effect	YES	YES	YES	YES
Control variables	NO	YES	NO	YES
Log-Likelihood Full Model	-43835	-42415	-43834	-42416
Wald Chi2	3436	8868	3435	8865
Prob >Chi2	0	0	0	0

Table 7: Treatment Based on Distance Between Firms and Data Trading Platforms

This table illustrates the relationship between the proximity of firms to data trading platforms and digital innovation. A negative binomial model is used. Columns 1 and 2 cover the sample period before the establishment of the data trading platform, spanning the years from 2008 to 2013. Columns 3 and 4 include the entire sample period, spanning from 2008 to 2022. The variables are defined in Table A1, and the control variables are described in Section 4.2.3. Robust standard errors, adjusted using the OIM method, are reported in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

	(1)	(2)	(3)	(4)
VARIABLES	*Dig Inno:2008-2013*		*Dig Inno:2008-2022*	
*Treatment*Distance*			0.013*	0.021***
			(0.008)	(0.007)
Distance	-0.057**	-0.049**	0.032***	0.036***
	(0.023)	(0.024)	(0.009)	(0.009)
Treatment			0.126***	0.039
			(0.041)	(0.039)
Constant	-0.019	-0.542	-1.289***	-0.767**
	(0.102)	(1.197)	(0.080)	(0.391)
Observations	2,961	2,961	20,465	20,465
Year fixed-effect	YES	YES	YES	YES
Firm fixed-effect	YES	YES	YES	YES
Control variables	NO	YES	NO	YES
Log-Likelihood Full Model	-4818	-4684	-38126	-36916
Wald Chi2	362.1	854.8	2954	7504
Prob >Chi2	0	0	0	0

Table 8: Data Trading Platforms, Digital Innovation and TFP

Table 8A presents the relationship between the establishment of data trading platforms and the total factor productivity (TFP) of companies. Table 8B presents the relationship between firms' digital innovation and their TFP. Odd-numbered columns report the estimation results without control variables, while even-numbered columns include them. The variables are defined in Table A1, and the control variables are described in Section 4.2.3. Standard errors are clustered at the firm level and are reported in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

(A)

VARIABLES	(1)	(2)	(3)	(4)
	TFP_OP		*TFP_LP*	
Treatment	0.035**	0.025*	0.067***	0.053***
	(0.015)	(0.014)	(0.017)	(0.016)
Constant	6.636***	6.300***	8.258***	7.818***
	(0.008)	(0.095)	(0.009)	(0.105)
Observations	24,418	24,418	24,418	24,418
R-squared	0.857	0.874	0.881	0.895
Year fixed-effect	YES	YES	YES	YES
Firm fixed-effect	YES	YES	YES	YES
Control variables	NO	YES	NO	YES

(B)

VARIABLES	(1)	(2)	(3)	(4)
	TFP_OP		*TFP_LP*	
L. Digital Innovation	0.001***	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	6.691***	5.094***	8.338***	6.267***
	(0.004)	(0.266)	(0.004)	(0.294)
Observations	21,139	21,139	21,139	21,139
R-squared	0.865	0.891	0.890	0.912
Year fixed-effect	YES	YES	YES	YES
Firm fixed-effect	YES	YES	YES	YES
Control variables	NO	YES	NO	YES

Appendix A. Additional tables and figures

Table A1: Variable Definitions

Variable	Definition
Treatment	A dummy variable that equals one in a year if a data trading platform has been established in the province where a firm is located by that year, and zero otherwise
Digital Innovation	The number of digital patent applications a firm files
DPM (Digital Product Manufacturing)	The number of digital patents a firm applies that fall under the "Digital Product Manufacturing" category
DPS (Digital Product Service)	The number of digital patents a firm applies that fall under the "Digital Product Service" category
DTA (Digital Technology Application)	The number of digital patents a firm applies that fall under the "Digital Technology Application" category
DFD (Digital Factor Drive)	The number of digital patents a firm applies that fall under the "Digital Factor Drive" category
Green Innovation	The number of green patent applications a firm files
Digital Others	The number of all patent applications minus digital patents a firm files
Green Others	The number of all patent applications minus green patents a firm files
Gov Attention	Natural logarithm of one plus the number of digital keywords in the annual work reports of provincial governments
Digital Policy(14^{th} 5-year)	A dummy variable that equals one in a year if there is a government policy related to the development plan for the digital economy, especially the 14th Five-Year Plan, in the province where a firm is located by that year, and zero otherwise
Digital Policy(1^{st})	A dummy variable that equals one in a year if there is a government's first policy related to the digital economy in the province where a firm is located by that year, and zero otherwise
Distance	Natural logarithm of one plus the distance of the city to the data trading platform location within the same province

Table A1: Variable Definitions (continued)

Variable	Definition
Age	Natural logarithm of one plus the number of years since a firm's establishment
Growth	Annual growth rate of a firm's operating revenue
Leverage	The ratio of a firm's total liabilities to total assets
Size	Natural logarithm of a firm's total assets for the year
Employee	Natural logarithm of a firm's number of employees
Market Value	The ratio of a firm's total market value to total assets
Profitability	The ratio of a firm's net profit to operating revenue
Liquidity	The ratio of a firm's current assets to total assets
Cash Flow	The ratio of a firm's net cash flow from operating activities to total assets
Board Size	Natural logarithm of a firm's number of board members
Ownership	1 for state-owned firms, 0 otherwise
Firm R&D	The ratio of a firm's research and development expenditure to operating revenue
TFP_OP	The total factor productivity (TFP) of firms, constructed based on Olley and Pakes (1996)
TFP_LP	The total factor productivity (TFP) of firms, constructed based on Levinsohn and Petrin (2003)
Real GDPpc	Per capita real GDP indexed to 2000 (one hundred thousand RMB)
Social Consumption	The ratio of total retail sales of consumer goods to regional GDP
Fiscal Expenditure	Province-level fiscal expenditures
Industrialization Level	The ratio of industrial value-added to regional GDP
Province R&D	The ratio of province-level expenditures on research and development to regional GDP
Province Innovation	Natural logarithm of the number of patent applications in each province
Informationization Level	The ratio of telecommunication business volume to regional GDP
Technological Development	The ratio of technical market trading volume to regional GDP

Population	The ratio of resident population to administrative area of the region

Figure A1: An Example of Data Trading Platform: Guiyang Data Exchange

This figure illustrates the Guiyang Data Exchange, highlighting the essential features of a data trading platform. The Guiyang Data Exchange offers a diverse range of data products and services for trading, spanning industries such as agriculture, forestry, mining, manufacturing, and public administration. Companies like Guizhou Zhengyu Technology Co., Ltd. provide specialized data resources that are tailored to specific market needs. The platform facilitates price negotiations and provides a variety of data products and services, enhancing data-driven economic activities.



Appendix B. Proofs

Proof of Equation (31):

Define the following Hamiltonian,

$$\mathcal{H}\left(c_{it}, a_t, \lambda_t\right) = lnc_t + \lambda_t \left(r_t a_t + w_t - \int_0^{N_t} p_{it} c_{it} di\right) \tag{35}$$

We then have following FOCs,

$$\frac{\partial \mathcal{H}}{\partial c_{it}} = \frac{1}{c_t}\left(\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{1}{\sigma-1}} c_{it}^{-\frac{1}{\sigma}} - \lambda_t p_{it} = 0 \tag{36}$$

$$\frac{\partial \mathcal{H}}{\partial a_t} = -\dot{\lambda}_t + \rho\lambda_t = \lambda_t r_t \tag{37}$$

FOCs of incumbent firm's problem are,

$$\left(1 - \frac{1}{\sigma}\right)\left(\frac{Y_t}{Y_{it}}\right)^{\frac{1}{\sigma}} \frac{Y_{it}}{L_{it}} = w_t \tag{38}$$

then the labor market clearing becomes,

$$\int_0^{N_t} l_{it} di + l_{et} = l_{pt} + l_{et} = 1$$

Using symmetry, the production structure of the economy can be simplified considerably. Consumption per person is,

$$c_t = \left(\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{\sigma}{\sigma-1}} = N_t^{\frac{\sigma}{\sigma-1}} c_{it} = N_t^{\frac{\sigma}{\sigma-1}} \frac{Y_{it}}{L_t}$$

The labor hired in the incumbent firms becomes,

$$L_{pt} = \int_0^{N_t} L_{it} di = N_t L_{it}$$

The production of a variety is,

$$Y_{it} = AL_{it} = A\frac{L_{pt}}{N_t}$$

Aggregate output in the symmetric economy is,

$$Y_t = \left(\int_0^{N_t} Y_{it}^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{\sigma}{\sigma-1}} = N_t^{\frac{\sigma}{\sigma-1}} Y_{it} = N_t^{\frac{\sigma}{\sigma-1}} A\frac{L_{pt}}{N_t} = N_t^{\frac{1}{\sigma-1}} AL_{pt}$$

Define the following Hamiltonian,

$$\mathcal{H}\left(c_{it}, d_t, a_t, \lambda_t\right) = lnc_t - d_t^{\kappa} + \lambda_t \left[r_t a_t + w_t + p_{dt} d_t - \int_0^{N_t} p_{it} c_{it} di \right] \tag{39}$$

We then have following FOCs,

$$\frac{\partial \mathcal{H}}{\partial c_{it}} = \frac{1}{c_t} \left(\int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{1}{\sigma-1}} c_{it}^{-\frac{1}{\sigma}} - \lambda_t p_{it} = 0 \tag{40}$$

$$\frac{\partial \mathcal{H}}{\partial d_t} = -\kappa d_t^{\kappa-1} + \lambda_t p_{dt} = 0 \tag{41}$$

$$\frac{\partial \mathcal{H}}{\partial a_t} = -\dot{\lambda}_t + \rho \lambda_t = \lambda_t r_t \tag{42}$$

The Hamiltonian for the optimal allocation is,

$$\mathcal{H}\left(l_{pt}, d_t, N_t, \lambda_t\right) = ln \left(N_t^{\frac{1}{\sigma-1}} l_{pt} L^{\alpha} d_t^{\alpha} \right) - \frac{\kappa}{2} d_t^2 \ + \lambda_t \eta N_t l_{et}^{1-\beta} L d_t^{\beta} \tag{43}$$

We then have following FOCs,

$$\frac{\partial \mathcal{H}}{\partial l_{pt}} = \frac{1}{l_{pt}} - \lambda_t \left(1-\beta\right) \eta N_t l_{et}^{-\beta} L d_t^{\beta} = 0 \tag{44}$$

$$\frac{\partial \mathcal{H}}{\partial d_t} = \frac{\alpha}{d_t} - \kappa d_t^{\kappa-1} + \lambda_t \beta \eta N_t l_{et}^{1-\beta} L d_t^{\beta-1} = 0 \tag{45}$$

$$\frac{\partial \mathcal{H}}{\partial N_t} = -\dot{\lambda}_t + \rho \lambda_t = \frac{1}{\sigma-1} \frac{1}{N_t} + \lambda_t \eta l_{et}^{1-\beta} L d_t^{\beta} \tag{46}$$

From (42),

$$\kappa d_t^{\kappa} = \alpha + \beta \lambda_t \dot{N}_t = \alpha + \frac{\beta}{1-\beta} \frac{l_{et}}{l_{pt}} \tag{47}$$

Then, we have,

$$l_{et} = \frac{\frac{1-\beta}{\beta}\left(\kappa d_t^{\kappa} - \alpha\right)}{1 + \frac{1-\beta}{\beta}\left(\kappa d_t^{\kappa} - \alpha\right)}$$

The balanced growth rate of varieties,

$$\frac{\dot{N}_t}{N_t} = \eta l_{et}^{1-\beta} L d_t^{\beta} = \eta L \left[\frac{\frac{1-\beta}{\beta}\left(\kappa d_t^{\kappa} - \alpha\right)}{1 + \frac{1-\beta}{\beta}\left(\kappa d_t^{\kappa} - \alpha\right)} \right]^{1-\beta} d_t^{\beta} \tag{48}$$

From (43),

$$-\dot{\lambda}_t N_t + \rho \lambda_t N_t = \frac{1}{\sigma-1} + \lambda_t \dot{N}_t$$

Obtain

$$\lambda_t N_t = \frac{1}{\sigma-1} \frac{1}{\rho}$$

Since (44) gives,

$$\lambda_t \dot{N}_t = \frac{\kappa d_t^{\kappa} - \alpha}{\beta}$$

Therefore,

$$\frac{\dot{N}_t}{N_t} = \frac{(\sigma - 1)\,\rho}{\beta}\left(\kappa d_t^{\kappa} - \alpha\right) \tag{49}$$