Analysis of the Technological Innovation Effects of Government Big Data Policy: PSM-DID Based Estimation Post-print

Authors: Chen Ling, Duan Yaoqing

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] Technological innovation serves as a crucial driving force for economic development and transformation. Empirically analyzing the technological innovation effects of China's government big data policy implementation contributes to advancing the national innovation-driven development strategy, promoting the integrated development of data and technology elements, and achieving collaborative and open innovation. [Method/Process] Based on the release of the "Action Outline for Promoting Big Data Development" and the establishment of national-level big data comprehensive pilot zones, this study employs panel data from 31 Chinese provinces spanning 2000-2019 as the research sample. The PSM method is utilized to conduct propensity score matching for treatment and control group samples, upon which the DID method is applied for difference-in-differences analysis. Robustness checks are performed through variable replacement methods to investigate the causal relationship between government big data policy and technological innovation. [Results/Conclusion] By addressing endogeneity issues inherent in public policy and the unobservability of counterfactuals, the study reveals that government big data policy can effectively promote technological innovation.

Full Text

Preamble

Analysis of the Technological Innovation Effect of Government Big Data Policy: An Estimation Based on the PSM-DID Method

Chen Ling[1], Duan Yaoqing[1,2]

[1]School of Information Management, Central China Normal University, Wuhan 430079

[2]Hubei Data Governance and Intelligent Decision Research Center, Wuhan 430079

Abstract:
[Purpose/Significance] Technological innovation is a crucial driver of economic development and transformation. Empirically analyzing the technological innovation effects of China's government big data policy implementation helps advance the national innovation-driven development strategy, promotes the integrated development of data and technology elements, and achieves linkage and open innovation. [Method/Process] Based on the release of the *Action Outline for Promoting Big Data Development* and the establishment of national-level big data comprehensive pilot zones, this study employs panel data from 31 Chinese provinces from 2000-2019. Using Propensity Score Matching (PSM) to match experimental and control group samples, we subsequently apply the Difference-in-Differences (DID) method for double differencing, with robustness checks conducted through variable replacement to explore the causal relationship between government big data policy and technological innovation. [Result/Conclusion] By addressing endogeneity issues in public policy and the unobservability of counterfactuals, we find that government big data policy can promote technological innovation.

Keywords: Government big data; Technological innovation effect; Policy evaluation; Propensity Score Matching; Difference-in-Differences

Classification Number: G203
DOI: 10.13266/j.issn.0252-3116.2020.20.011

Currently, leveraging big data to drive economic development and enhance regional technological innovation capabilities is becoming a trend. To accelerate the construction of a data powerhouse and promote technological innovation, the State Council issued the *Action Outline for Promoting Big Data Development* (hereinafter referred to as the *Outline*) in 2015, China's first authoritative and systematic document for promoting big data development. Accordingly, China is committed to adapting its big data development and application processes to the national innovation-driven development strategy to achieve technological innovation. In 2020, the CPC Central Committee and State Council released the *Opinions on Improving the Market-oriented Allocation Mechanism of Factors* (hereinafter referred to as the *Opinions*), the central government's first document on factor market allocation, which formally introduced data as a new type of production factor in official documents for the first time. The *Opinions* propose accelerating the development of technology factor markets to promote scientific and technological innovation cooperation, while also calling for the accelerated cultivation of data factor markets to enhance the value of data resources. Adhering to innovation-driven development and deepening big data applications has become an inherent requirement and inevitable choice for promoting technological innovation.

Although the *Outline* emphasizes intensifying R&D on key big data technologies, deepening innovative applications of big data, releasing technology dividends and innovation dividends, and opening a new pattern of innovation-driven development, the technological innovation effects of the *Outline*'s implementation have not yet been empirically tested. Examining the impact of the *Outline*'s implementation on provincial technological innovation level growth helps correctly grasp the policy's effectiveness, understand the direction of the national innovation-driven development strategy, and provide reference for promoting technological, open, and linkage innovation. Therefore, this paper attempts to answer the following research question: Based on existing big data policies, by comparing technological innovation level differences between policy pilot zones and non-pilot zones before and after policy implementation, we reveal the actual technological innovation effects generated during China's government big data policy advancement process.

1 Related Research

1.1 Public Policy Evaluation

Public policy evaluation is an essential component of the policy lifecycle and an important basis for deciding whether to continue, adjust, or terminate policies. The main research contents include theoretical system construction of evaluation indicators and empirical analysis of policy evaluation. In terms of theoretical system construction, perspectives include planning design, implementation, and evaluation stages; basic layer, data layer, and platform layer indicators; timeliness, completeness, and accuracy dimensions; and legal, regulatory, and policy levels. Empirical policy evaluation has undergone four stages: first, using multiple linear or nonlinear regression to address actual effect issues; second, employing simultaneous equations to address endogeneity; third, adopting Propensity Score Matching (PSM) to address sample selection bias; and fourth, combining PSM with Difference-in-Differences (DID) to control for unobservable factors.

1.2 Policy Technology Innovation Effect

Technological innovation, as a key driver of economic development and transformation, is a focal point in innovation research. Studies on policy technology innovation effects concentrate on three aspects: measurement of technological innovation, theoretical analysis of effects, and empirical testing. For measurement, three common indicator types are used: first, innovation input indicators such as R&D expenditure; second, financial performance indicators like market share and new product profit margins; and third, innovation output indicators such as patent applications. Theoretical analysis of government-guided innovation policies' effects on technological innovation has long lacked unified views. Some scholars argue these policies have incentive effects through fiscal support and reduced innovation costs. Others contend they have crowding-out effects due to information asymmetry between government and enterprises or mismatches between supported and actual R&D projects. Still others suggest non-linear relationships where effects vary with funding intensity and technology spillover strength. Empirical conclusions are similarly mixed, with some showing positive effects, others insignificant effects, and some neutral effects.

1.3 Research Review

Literature review reveals that: First, the direction and intensity of government-guided innovation policies' effects on technological innovation remain controversial. Second, existing literature mostly focuses on single aspects of innovation input or output, rarely combining both to examine joint effects on technological innovation input and output. Accordingly, this study employs the PSM-DID method based on relevant panel data, combining innovation input and output to answer three questions: (1) Does government big data policy affect technological innovation activities? (2) What is the direction of this effect—does it incentivize or inhibit technological innovation? (3) What is the magnitude of this effect—to what extent does the policy stimulate actual technological innovation input and output?

2 Research Design

2.1 Research Methods

Policy effect evaluation faces the challenge of endogeneity between policy and events. Unlike natural sciences, social sciences generally cannot use simulated experiments, instead relying on "quasi-experimental" opportunities and econometric tools like instrumental variables, regression discontinuity, PSM, and DID. PSM identifies control group samples most similar to experimental group samples to address selection bias from non-random data. DID, emulating natural science experiments, effectively combines "before-after differences" and "with-without differences" to isolate policy effects. PSM-DID eliminates unobservable factors' influence, yielding net policy effects. By matching experimental and control group samples through PSM and DID, we can compare "with-without" differences in technological innovation levels before and after *Outline* implementation, enabling scientific and accurate policy assessment.

2.2 Research Hypotheses

The *Outline* aims to guide and encourage innovation in key big data application technologies, forming a secure and reliable big data technology system. Empirical analysis of government big data policy's effect on urban technological innovation levels requires the following assumptions: (1) The common trend (parallel trend) assumption in DID analysis, where experimental and control groups would have the same effect trend without policy intervention. If violated, PSM should first match groups before DID. (2) The common support and balance assumptions in PSM. Common support requires experimental and control groups' propensity score ranges to be similar. Balance tests examine whether mean differences across matching variables become insignificant after matching.

2.3 Research Framework

Our objective is to compare technological innovation levels between big data policy pilot zones and non-pilot zones before and after policy implementation, empirically analyzing the technological innovation effect of China's government big data policy. The research process includes: (1) Data processing and testing—conducting descriptive statistics and paired sample T-tests after data collection and cleaning. (2) PSM empirical analysis—using Logit models to calculate propensity scores, matching experimental and control groups via kernel matching, then testing common support and balance assumptions to calculate treatment effects. (3) PSM-DID empirical analysis—building DID regression models after PSM matching, conducting significance tests and analyzing results to assess net policy effects, with robustness checks through variable replacement. The framework is shown in Figure 1 [Figure 1: see original paper].

3 Data Processing and Testing

3.1 Sample Selection

The data covers 2001-2019, excluding Hong Kong, Macau, and Taiwan, covering 31 provincial-level administrative units for 589 panel observations. Data comes from the *China Statistical Yearbook*, with missing values interpolated. Provinces are coded 1-31. In 2015, the State Council issued the *Outline*, treated as a natural experiment. To implement the *Outline*, China launched two batches of national-level big data comprehensive pilot zones. Since direct measurement of provincial *Outline* implementation is impossible, we use the establishment of these pilot zones as the measure. Accordingly, Guizhou, Beijing, Tianjin, Hebei, Guangdong, Shanghai, Henan, Chongqing, Liaoning, and Inner Mongolia (10 provincial-level regions) serve as the experimental group, while Shanxi, Jilin, Heilongjiang, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Hubei, Hunan, Guangxi, Hainan, Sichuan, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang (21 regions) form the control group. The policy implementation year is set as 2016, with 2016 and earlier as pre-policy and post-2016 as post-policy.

3.2 Variable Setting

Variable definitions, calculations, and sources are shown in Table 1 . The dependent variable is technological innovation level under big data policy. Common measures include innovation input, output, and financial performance, each with merits and limitations. We use innovation output as the primary measure and innovation input for robustness checks, comprehensively capturing the technological innovation effect. Innovation output (*patent*) is measured as the logarithm of authorized patents, while innovation input (*innovation*) is measured as the ratio of R&D expenditure to GDP.

Key explanatory variables include policy dummy (*treated*), time dummy (*time*), and their interaction. *Treated* equals 1 for pilot zone provinces, 0 otherwise. *Time* equals 1 for 2016 and after, 0 otherwise. Their interaction term *time*treated* captures the net policy effect. Control variables include urbanization level (*city*), education level (*edu*), financial development level (*finance*), government size (*gov*), government support (*govsup*), industrialization level (*industry*), informatization level (*informatization*), labor cost (*labor*), economic development level (*lngdp*), industrial structure (*third*), and transportation convenience (*traffic*).

3.3 Descriptive Statistics

Descriptive statistics are shown in Table 2 . For all samples, innovation output ranges from 1.94591 to 13.07754, with a mean of 8.685775 and standard deviation of 1.862528, showing concentrated distribution. Other variables' statistics are omitted for brevity.

3.4 Paired Sample T-test

Paired sample T-test results between experimental and control groups across matching variables are shown in Table 3 . Significant differences exist across all control variables, indicating that pilot zone selection was non-random and featured specific characteristics, creating selection bias. Multivariate regression on all data would yield biased results. Therefore, we employ PSM-DID to address selection bias and better examine big data policy's innovation impact.

4 Propensity Score Matching Empirical Analysis

4.1 Logit Regression Results

Key matching variables selected through covariate screening include urbanization level, financial development, government size, government support, industrialization level, labor cost, economic development level, and transportation convenience. Logit regression results with pilot zone establishment as the dependent variable are shown in Table 4 . All matching variables except financial development are significant, indicating these factors importantly influenced pilot zone selection.

4.2 Common Support Test

To ensure matching quality, we examine the overlap of propensity score distributions between experimental and control groups (common support region), shown in Figure 2 [Figure 2: see original paper]. After kernel matching, approximately 81.23% of experimental and control group samples fall within the common support range (*onsupport*), with only a few outside (*offsupport*), satisfying the common support assumption.

4.3 Balance Test

Balance test results under kernel matching are shown in Table 5 . Before matching, significant differences exist across variables with t-values far exceeding 2 and P-values of 0.0000. After matching, differences decrease, t-values drop, and P-values become insignificant, indicating successful matching. All standardized bias absolute values are below 15%, showing good consistency between groups and satisfying PSM balance assumptions. Absolute bias changes before and after matching are shown in Figure 3 [Figure 3: see original paper], where post-matching biases shrink dramatically.

4.4 Matching Results

Kernel density distributions of propensity scores before and after matching are shown in Figure 4 [Figure 4: see original paper]. Pre-matching distributions differ substantially between groups, but post-kernel-matching distributions align closely, indicating satisfactory matching results. Matching results are detailed in Table 6 . First, pre-matching differences in innovation performance between groups are positive and significant, suggesting an amplified pre-policy effect that PSM corrects. Second, the Average Treatment Effect on the Treated (ATT) is positive and significant post-matching, indicating a positive policy effect on innovation. Third, the Average Treatment Effect on the Untreated (ATU) is also positive and significant, confirming the positive impact. Finally, the Average Treatment Effect (ATE) is positive and significant, demonstrating that government big data policy positively affects innovation performance.

5 PSM-DID Empirical Analysis

5.1 Model Construction

After PSM matching, experimental and control groups show similar trends across matching variables, satisfying the parallel trend assumption. The model is specified as:

$$Patent_{it} = \beta_0 + \beta_1 time_{it} + \beta_2 treated_{it} + \beta_3 time_{it} * treated_{it} + \alpha X_{it} + \varepsilon_{it}$$

where $Patent_{it}$ reflects technological innovation level of province i in year t; $time_{it}$ is the time dummy; $treated_{it}$ is the policy dummy; X_{it} represents control variables; and ε_{it} is the error term. The interaction term coefficient β_3 captures the net effect of government big data policy.

5.2 Difference-in-Differences

With validated PSM results, we estimate the model using matched data. PSM-DID significance tests and regression results are shown in Tables 7 and 8 . Table 7 shows significant positive differences between experimental and control groups both before (0.670*) and after (3.393*)* **policy implementation, with a significant positive net policy effect (2.724***). Table 8 confirms the positive and significant interaction term coefficient, demonstrating that government big data policy significantly promotes technological innovation. Among control variables, urbanization and financial development show positive but insignificant effects; education, government support, industrialization, informatization, and industrial structure have significantly positive effects; while government size, labor cost, economic development level, and transportation convenience have significantly negative effects.

5.3 Robustness Test

To verify baseline results and avoid 偶然性 from single indicator selection, we conduct robustness checks using innovation input level (*innovation*) as an alternative to innovation output (*patent*). Results are shown in Tables 9 and 10 . The policy-time interaction coefficient remains significantly positive (2.441***), confirming that baseline results are not sensitive to innovation indicator selection and demonstrating robust conclusions.

6 Summary and Outlook

6.1 Research Conclusions

Through PSM, DID, and robustness tests, consistent results show that government big data policy promotes technological innovation. The reasoning process is: (1) PSM results satisfy common support and balance assumptions, with ideal matching quality and significantly positive treatment effects (ATE, ATT, ATU), supporting the hypothesis after controlling for observable selection bias. (2) PSM-DID regression shows significantly positive net policy effects and interaction terms, confirming the hypothesis after controlling for both observable and unobservable factors. (3) Robustness tests with alternative indicators yield consistent, significant results, confirming conclusion stability. Thus, the hypothesis passes multiple validations: government big data policy significantly promotes technological innovation development.

6.2 Policy Recommendations

Based on empirical findings, we propose recommendations for provinces to achieve and enhance technology and innovation dividends: First, vigorously promote national big data comprehensive pilot zone construction and encourage deep policy advancement. Second, governments should establish complete professional research institutions or scientific laboratories to support innovation infrastructure, providing human capital support for pilot zone construction under innovation capital and equipment investment to enhance innovation levels. Finally, different pilot cities can achieve faster technological innovation progress from government big data policies, so the nation should grasp overall development patterns, layout rationally, and accelerate big data technological innovation progress while promoting big data development and application.

6.3 Research Outlook

This study has limitations: (1) Innovation level measurement indicators are imperfect. Scholars have not unified innovation level indicators. While we measured from input and output perspectives, future research could expand innovation level connotations more comprehensively. (2) Methodological requirements for data selection are fixed. While PSM-DID overcomes selection bias and unobservable factors, it requires panel data with fixed effects and cannot be applied to time series data. Future research could adopt more scientific methods like regression discontinuity or treatment effect models.

References

[1] Zhang Ke. Institutional setup and functional allocation of provincial big data bureaus: Based on a new round of institutional reform empirical analysis [J]. E-Government, 2019(6): 113-120.
[2] Huang Huang, Sun Xuezhi. Preliminary study on local government data governance institutions in China: Current situation and patterns [J]. Chinese Public Administration, 2018(12): 31-36.
[3] Yang Daifu, Yun Zhan. Research on public policy evaluation innovation in the big data era: Based on a process perspective [J]. E-Government, 2020(2): 92-99.
[4] Song Yi, An Xiaomi, Ma Guanghui. Research on big data governance capacity of the US, UK, and Australian governments—Based on big data policy content analysis [J]. Information and Documentation Services, 2018(1): 12-17.
[5] Zhang Huiping, Guo Ning, Tang Xikai. Promotion logic and future path: Textual analysis of China's government big data policy [J]. Journal of Intelligence, 2018, 37(3): 152-157, 77-84.
[6] Liu Yaya, Qu Wan, Feng Haihong. Research on the evolution of China's big data policy system [J]. Science Research Management, 2019, 40(5): 15-25.
[7] McNeely CL, Hahm JO. The big (data) bang: Policy, prospects, and challenges [J]. Review of policy research, 2014, 31(4): 304-310.
[8] Hoechtl J, Parycek P, Schoellhammer R. Big data in the policy cycle: Policy decision making in the digital era [J]. Journal of organizational computing and electronic commerce, 2016, 26(1-2): 147-169.
[9] Zhang Yongjin, Wang Jingxuan. Comparative study on big data policies of major developed countries [J]. Chinese Public Administration, 2014(12): 113-117.
[10] Heitmueller A, Henderson S, Warburton W, et al. Developing public policy to advance the use of big data in health care [J]. Health affairs, 2014, 33(9): 1523-1530.
[11] Zhou Jingyan, Zhang Huina, Huang Yurong, et al. Textual quantitative analysis of China's big data policy from a policy tool perspective [J]. Informa-

tion Research, 2016(12): 7-10.
[12] Ding Wenyao, Zhang Zili, Yu Guoxian, et al. Research on diffusion patterns and transfer characteristics of local big data policies in China [J]. Big Data Research, 2019, 5(3): 76-95.
[13] Xu Lei, Li Qing, Xiao Xiangze. Research on co-word network of big data policy based on grounded theory [J]. Modern Information, 2018, 38(6): 157-164.
[14] Huang Cui, Ren Xing, Zhang Jian. Quantitative research on policy documents: A new direction for public policy research [J]. Journal of Public Management, 2015, 12(2): 129-137.
[15] Shi Dai. Research on evaluation of China's government open data policy [D]. Harbin: Heilongjiang University, 2018.
[16] Jiang Yongjian. Research on maturity evaluation of China's government data openness [D]. Hangzhou: Zhejiang University, 2019.
[17] Sun Jing. Analysis of Shanghai's data opening policy—Based on SPDQM data quality evaluation [D]. Wuhan: Huazhong University of Science and Technology, 2018.
[18] Cai Chengcheng, Liu Xinping, Zheng Lei. Open government data readiness assessment: Laws, regulations and policies [J]. E-Government, 2017(9): 49-55.
[19] Duan Xiufang, Kou Minglong. Impact of the "Belt and Road" initiative on economic growth of five eastern provinces and cities—Based on the PSM-DID model [J]. Economic Forum, 2019(10): 25-34.
[20] Sun Wenjun. Empirical study on the relationship between the "Belt and Road" initiative and China's economic development—Based on panel data from 31 provinces [J]. Economic Research Guide, 2019(29): 91-93.
[21] Cao Jing, Guo Zhe. Policy effect of China's sulfur dioxide emission rights trading pilot—Policy effect evaluation based on PSM-DID method [J]. Chongqing Social Sciences, 2019(7): 24-37.
[22] Li Linhong, Wang Juan, Xu Yanfeng. Impact of low-carbon pilot city policy on enterprise technological innovation—Empirical research based on DID model [J]. Ecological Economy, 2019, 35(11): 48-54.
[23] Liu Xiaodan, Zhang Bing. Informal institutions and learning effects of emerging economy enterprises' OFDI—Based on PSM-DID analysis method [J]. World Economy Studies, 2019(11): 80-93, 135-136.
[24] Wang Peng, Wu Silin, Li Yan. Can the establishment of national high-tech zones promote urban industrial structure upgrading?—Empirical analysis based on PSM-DID method [J]. Comparative Economic & Social Systems, 2019(4): 17-29.
[25] Xu Yuan. Research on the impact of equity incentives on innovation performance [D]. Jinzhou: Bohai University, 2019.
[26] Gao Qing. Impact of intellectual property barriers on export innovation based on PSM-DID method [J]. Social Sciences in Xinjiang, 2019(3): 41-48.
[27] Li Zhuowen. Does the construction of national independent innovation demonstration zones promote urban technological innovation?—Empirical research based on DID [J]. Financial Supervision, 2019(6): 101-106.
[28] Tian Renhe, Zhang Zhiqiang, Wang Fei, et al. Empirical study on funding

effect of science and technology policy innovation capability based on DID model—Taking the Earth Science Project of the National Science Fund for Distinguished Young Scholars as an example [J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(8): 782-795.

[29] Tian Renhe, Zhang Zhiqiang, Yu Jie, et al. Empirical study on funding effect of science and technology policy based on DID model—Taking the 20-year implementation of the Earth Science Project of the National Science Fund for Distinguished Young Scholars as an example [J]. Library and Information Service, 2018, 62(18): 110-121.

[30] Liu Yuhong. Resource allocation effect of R&D subsidies on enterprise technological innovation—Review of foreign empirical research [J]. Economic Forum, 2020(1): 24-36.

[31] Wang Chunhui, Li Ping. Analysis of policy effects of government support for enterprise technological innovation [J]. Science & Technology Progress and Policy, 2012, 29(2): 106-109.

[32] Wang Banban. Review of research on environmental policy and technological innovation [J]. Economic Review, 2017(4): 133-150.

[33] Chen Xiao, Li Meiling, Zhang Zhuangzhuang. Environmental regulation, government subsidies and green technological innovation—Empirical research based on mediation effect model [J]. Industrial Technology & Economy, 2019(9): 18-25.

[34] Li Dongqin. Environmental policy tool combination, environmental technological innovation and performance [J]. Science Research Management, 2018, 36(12): 160-169.

[35] Guo Xu, Sun Xiaohua, Xu Ran. On the innovation effect of industrial technology policy—Demand pull or supply push? [J]. Science Research Management, 2017, 35(10): 31-44.

[36] Wu Xianming, Zhang Yumei. Do overseas M&As by state-owned enterprises create value: Empirical test based on PSM and DID methods [J]. World Economy Studies, 2019(5): 80-92, 136.

[37] Liu Hewang, Liu Botao, Zheng Shilin. Environmental regulation and industrial transformation and upgrading: DID test based on the "Eleventh Five-Year" emission reduction policy [J]. China Soft Science, 2019, 341(5): 45-57.

[38] Li Weibing, Wang Peng. Will increasing pollution fees inhibit FDI inflow?—Estimation based on PSM-DID method [J]. Journal of Xi'an Jiaotong University (Social Sciences), 2020, 40(3): 91-100.

Author Contributions:

Chen Ling: Responsible for outline formulation, data collection, data analysis, and initial draft writing.

Duan Yaoqing: Responsible for topic selection and comprehensive revision.

Chen Ling[1], Duan Yaoqing[1,2]

Technological Innovation Effect Analysis of Government Big Data Policy—Estimation Based on PSM-DID Method