sustainability

MDPI

Article

Threshold Effect of the Internet on Regional Innovation in China

Congbo Chen * and Azhong Ye *

School of Economics & Management, Fuzhou University, Fuzhou 350108, China
* Correspondence: Chenconb@163.com (C.C.); ye2004@fzu.edu.cn (A.Y.)

Abstract: Internet business adoption is an essential determinant of regional innovation which has received little attention in the literature. This paper emphasizes the role and threshold effect of Internet business adoption in increasing regional innovation outputs. We construct a threshold spatial autoregressive model to illustrate the nonlinear positive impact of Internet business adoption on innovation, simultaneously estimating interregional knowledge spillovers. To test threshold effect and interregional knowledge spillover, we use province-level panel data set in China and calculate Moran's I and LR-like statistics to confirm the nonlinearity and spatial dependence. Within the empirical model, we find a positive relationship between the number of websites owned by local firms and the number of patents filed in that specific region. Our analysis suggests that Internet business adoption has a greater marginal benefit on the innovation of isolate regions. The results also indicate that ignoring interregional knowledge spillover may cause mistakes in the model on regional innovation systems. Policy implications for these results suggest that the government should not only pay attention to Internet development of the whole country but also encourage the reduction of digital divisions among regions

Keywords: information communication technologies; regional innovation; threshold spatial autoregressive model

Citation: Chen, C.; Ye, A. Threshold Effect of the Internet on Regional Innovation in China. *Sustainability* **2021**, *13*, 10797. https://doi.org/10.3390/su131910797

Academic Editors: Orlando Troisi, Anna Visvizi, Wadee Alhalabi, Shahira Assem Abdel Razek and Paolo Gerli

Received: 30 August 2021
Accepted: 26 September 2021
Published: 28 September 2021

1. Introduction

The Internet has been a general technology which enables an array of organizational changes in innovation activities. The Internet has induced significant increasement in innovation efficiency by changing the organization and structure of the innovation system. The new understanding of the role of the Internet on knowledge generation enables us to consider inscape and instress of regional innovation systems and to suggest an augmented role for regional telecommunications policy as a driving force for sustainable innovation in economic development. Chinese local government initiatives to foster the spread of Internet, such as the "Internet plus" strategy, are a testament to the expectation that the Internet has significant effect, and especially, that the Internet can provide new momentum to constructing innovation in the country [1,2]. Therefore, this paper provides evidence of nonlinear positive effects of adoption of the Internet on regional innovation using the case of China. It analyses whether the Internet has heterogeneous impacts on innovation in groups of provinces that differ with regard to their business adoption of the Internet.

Much of the literature on innovation management looks at the Internet's contributions to firms' innovation performance. Some study the contributions of the Internet from the perspective of gatekeeper. The communication function of the Internet largely replaces the role of gatekeeper and expands the scope of knowledge flow [3]; thus, gatekeepers' pioneering use of the Internet can drive the penetration of information technology in the firm [4], which will strengthen firm's whole ability to absorb external knowledge through the Internet. Some study the contributions of the Internet on open innovation. A widespread argument suggests that tacit knowledge can only be transmitted through face-to-face communication, thus tacit knowledge sharing is the key to the success of open innovation [5]. Hildrum [6] and Buunk et al. [7] provide cases of tacit sharing online, suggesting the Internet has positive effect on tacit sharing behavior and intention. Compared with the number of studies at the firm level, little attention has been paid to the driving mechanism of Internet use on regional innovation. The research closely related to ours has focused on the relationship between regional Internet and firms' innovation. Billion [8] identified two differentiated patterns of regional innovation and ICT use that combine strong disparities in ICT use and innovation, using data from the European Union. Zhang and Li [2] found that regional widespread ICT access would enable the positive impact of guanxi on entrepreneurial performance, while restricted ICT access would inhibit it. Zhang et al. [9] provided a case of official document exchange via microblogging (ODEM) in Haining of China and found that regional endogenous factors are key determinants of the emergence of open innovation in the public sector. Though the Internet has a positive effect on regional innovation, opportunities to support firms from disadvantaged backgrounds are not equal. Results from quantile regressions using data from 117 developing countries show that only the most productive firms reap productivity gains from Internet-enabled knowledge access [10].

The above literature confirms the positive effect of Internet use on innovation, but there are still more questions which can be studied: (1) Most of the literature studies the Internet's effect on innovation at the national level, while study at the provincial level is rare; (2) Nonlinearity of the Internet's effect on regional innovation has been neglected while research on the relationship between Internet and innovation focuses on whether there is a positive effect of the Internet on innovation, presupposing linearity. The effect of the Internet on regional innovation is likely to be limited and complex. The relationship between regional innovation and the Internet reflects how heterogeneous firms with Internet adoption choose their location. When Internet adoption has unbalanced spatial distribution, the opportunities to support regional innovation for the Internet will be not equal. Moreover, regional innovation can also be influenced by the technical distance of firms [11], norms of publication [12], and other unknown factors besides Internet adoption. (3) Most empirical literature on the Internet's effects presumes localized knowledge spillovers which might induce inaccurate estimation results, as it has been a consensus that there is spatial dependence of regional knowledge. Jaffe [13] firstly shows that local externalities in the innovation process may operate outside the firm but still within the local space of the region. Later, spatial econometric model proposed by Anselin et al. [14] were employed to discriminate empirically interregional knowledge spillovers [15,16].

In the current paper, we examine in more detail nonlinearity and spatial dependence for regional innovation in China, and make several contributions to the literature. To the best of our knowledge, this paper provides evidence of Internet business adoption influencing regional innovation with a threshold mode. To test the threshold effect of Internet adoption on regional innovation, we built an empirical model in which we considered a basic knowledge production function based on the pioneering work of Griliches, and interregional knowledge spillovers based on the previous work of spatial econometric modeling. We find that greater Internet adoption does not mean more regional innovation. When Internet adoption reaches the threshold, its marginal effect on innovation decreases. Secondly, we use a novel empirical model, the panel data threshold spatial autoregressive model (TSAR). for our research. The TSAR model combines the spatial autoregressive model (SAR) with the threshold model so that it can estimate both the spatial spillovers and threshold effect of the Internet.

Our model focuses on the threshold effect of Internet business adoption on decreased discovery costs, thereby increasing regional patent innovation. On the basis of the spatial empirical model, we combine the SAR model with the threshold model proposed by Hansen [17] and construct TSAR in Section 2. In Section 3, we describe the data. Section 4 discusses a number of econometrics issues, Section 5 presents the empirical results and Section 6 concludes.

2. The Empirical Model

In this section we describe the econometric model used to investigate the determinants and mechanism of regional knowledge production. Regional knowledge production is modeled by using an econometric framework comprising four components: a proxy for regional innovative output I_{it}, region-specific inputs A_{it}, an explanatory variable capturing the contribution of interurban knowledge spillover KS_{it}, a set of additional explanatory variables that control for systematic differences between theory and real world $CONTROL_{it}$, and a random error term ε_{it}:

$$I_{it} = f\left(A_{it}, KS_{it}, CONTROL_{it}, \varepsilon_{it}\right) \quad (1)$$

where $i = 1, \ldots, n$ indexes 31 provinces in China and $t = 1, \ldots, T$ indexes years from 2014 to 2018. Some regions and provinces in China, such as Hong Kong and Taiwan are excluded from the analysis, mainly because some statistical data is less reliable than for the other regions.

2.1. Region-Specific Inputs on Knowledge Production

The basic knowledge production function (KPF) proposed by Griliches relates the innovative output in region i to its specific R&D inputs. Under Internet age, Internet is a key factor related to knowledge production but neglected by basic KPF. We depart slightly from this specification by introducing a further factor related to Internet business adoption. The general form of reformed KPF using a Cobb–Douglas production function is given as:

$$lnI_{it} = \alpha_0 + \alpha_1 lnL_{it} + \alpha_2 lnINT_{it} \quad (2)$$

where INT is approximated by the number of websites owned by regional firms and L is the resources of a regional innovation system, approximated by employment in research and development. Our main interest is the slope of INT, but in fact the slope may change due to differences in interregional economic backgrounds. We first introduce the resource inputs of regional innovation and then mainly argue the Internet's non-linear effect on regional innovation.

2.1.1. The Resource Inputs of Regional Innovation

The resources of the regional innovation system, L_{it}, are approximated by research and development personnel employed by local innovative actors (firms, universities, public research institutions). To appropriately capture the causality between the regional-specific inputs and knowledge production, a time lag of one year is allowed for. Research and development capital and regional knowledge stock may be further inputs. The present paper focuses on the Internet's effects of improving interaction and communication on the elements of the regional innovation system; therefore, we focused on research and development employment as the resources of regional innovation system using the log of such personnel employed by regional actors.

2.1.2. The Driving Mechanism of Internet for Regional Innovation

The concept of an "innovation system" is based on the idea that actors and elements constituting the innovation system interact in the production, diffusion and use of knowledge [18]. The main bodies of a regional innovation system include universities, industry, and local government [19]. Among them, local firms are the main actor of technology innovation and the core of the innovation system. The question arises, what would be induced by Internet business adoption by innovative firms. The inherent advantages of the Internet, such as interconnected sharing of knowledge and near-zero cost of information acquisition, effectively affect local firms' innovation activities in the regional innovation system. Next, we discuss two mechanisms by which Internet business adoption affects innovation. On the one hand, there exists a positive relationship between Internet business adoption and knowledge production. On the other hand, the effect of Internet business adoption on regional innovation may be characterized as non-linear.

Initially, increasing Internet business adoption promotes regional innovative activity by influencing related knowledge and information flow among firms. Each firm in the regional innovation system should not be isolated. The innovation process can be characterized as highly interactive, referring to internal collaboration between the departments of a company as well as to external links and co-operation with other firms. Local knowledge production increasingly depends on both agglomeration of firms and exchange of ideas. Theoretically, firms have been treated as special producers in the absence of idea exchange. Empirically, however, firms divide innovative time between pure production and exchange with others, including application firms and research and development firms, in order to raise their productivity. In modern knowledge economies driven by innovation, diversification and individuation, both types of firms contribute to regional innovative capacity due to the importance of idea exchange for innovation.

Through Internet business adoption, the expenditure of each firm on idea exchange is next to zero; thus, the regional innovation system creates a favorable atmosphere. It is embodied that: (1) For research and development firms, only if such firms have a high level of Internet business adoption can they discover effective information related to their research patents, such as the prior discovery of patents and the number of others pursuing them. Not knowing this information due to high discovery cost can cause firms to choose not to pursue innovation. The Internet is conducive to innovative activity and will, to a greater extent, place high demands on the such firms so that they are forced to continuously develop new technologies, create new ideas, and constantly improve their Internet business application level and innovation efficiency level. (2) For application firms, the continuous proliferation of Internet business adoption in this area has made it easier to turn knowledge production into economic benefits by promoting the ability to commercialize new technologies. (3) For interaction between research and development firms and application firms, the former bring new knowledge and technologies to the latter more effectively through the Internet, while in the use of new technologies by application firms, more demand and increasing experience for innovation will feed back to research and development, forcing the continuous pursuit of innovation.

Furthermore, the effect of Internet business adoption on regional knowledge production might be nonlinear. While Internet business adoption reduces discovery costs, there may be considerable barriers to Internet-driven knowledge flow. Prior literature has suggested that individuals use the Internet disproportionately to reach out to existing contacts and those who share similar tastes or knowledge bases, [20] and Internet connectivity is able to facilitate knowledge flows between locations only when they share a common knowledge base [11]. There is potential for Internet business adoption to reduce barriers due to distance while at the same time increasing barriers from technical and recognitive distance. Consequently, in the initial stage of Internet development firms with common or similar knowledge bases are connected, while in the mature stage firms are grouped based on common knowledge, and different groups are not able to connect together, expanding the technical and cognitive difference.

Based on the mechanisms above, we assume that there is a threshold effect of the Internet on regional knowledge production. The definition of threshold in this paper may differ from those in prior literature, as it estimates the turning point of the Internet's decreasing marginal effect. The Internet's decreasing marginal effect can be observed when it is beyond this given threshold level. We believe that the Internet, as proposed by Van Alstyne and Brynjolfsson [21], reduces the importance of geographic distance but simultaneously increases distance arising from technical difference. We therefore focus on the 'threshold', which is the turning point of the Internet's promoting effect, and aim to provide empirical evidence of interregional Internet business adoption differences in the impact of innovation gaps among regions.

According to Hansen's [18] research framework for the panel data threshold model, the extended KPF is given as:

$$lnI_{it} = \alpha_i + \alpha_1 lnL_{it} + \theta_1 lnINT_{it},\ \ lnINT_{it} < \gamma \tag{3}$$

$$lnI_{it} = \alpha_i + \alpha_1 lnL_{it} + \theta_2 lnINT_{it},\ lnINT_{it} \geq \gamma \tag{4}$$

where regional Internet business INT_{it} is approximated by number of websites owned by local firms. For convenience of expression we write single threshold model, but the number of the threshold needs be estimated through testing of the threshold effect.

2.2. Spatial Knowledge Spillovers

The indicator of $\boldsymbol{W}ln\boldsymbol{I}$ for KS_i has been used in many studies [15,16]. The pros are based on omitted variables arguments [22]. It is difficult to find sample data that adequately reflects the regional innovative environment, entrepreneurial spirit, research institution, government and a host of other influences that may be important for knowledge spillover modeling problems. Using $\boldsymbol{W}ln\boldsymbol{I}$, we can model the spatial dependence of omitted variables.

Formally, the explanatory variable representing spatial knowledge spillovers to region i can be defined as the weighted sum of innovative output from region $i\prime s$ neighboring regions, i.e., $KS_{it} = \sum_{j=1}^{n} w_{ij} lnI_{jt}$. $\boldsymbol{W}$ is a known spatial weight matrix of dimension n with zero diagonal. $\boldsymbol{W}$ is also assumed to be row normalized, i.e., $\sum_{j=1}^{n} w_{ij} = 1$. The special weight w_{ij} is the $(i, j)th$ element of the spatial weight matrix $\boldsymbol{W}$, which generates the spatial dependence between cross-section i and j. Ceteris paribus, the higher I_{jt} and the higher spatial share w_{ij}, the more easily knowledge flows from neighboring region j to i, the higher contribution of external knowledge to knowledge production in region i. $\sum_{j=1}^{n} w_{ij}$ is the spatial weighted average of the value innovative output in the neighboring cross-sections of region i, which is referred to as the spatial lag of regional innovative output. In other words, $\boldsymbol{W}ln\boldsymbol{I}$ is a weighted measure of R&D output in the regions with which region i has contracts. We interpret an influence of $\boldsymbol{W}ln\boldsymbol{I}$ on lnI_{it} as evidence of spatial spillovers of the knowledge located outside i, whereas the coefficient of $\boldsymbol{W}ln\boldsymbol{I}$ would indicate the degree of knowledge spillover from neighboring regions for i.

As an important component of KS, the spatial weights w_{ij}, determines (1) from which region j knowledge spillovers to region i may occur, and (2) which share of the knowledge in region j actually found its way to i. The present paper chooses inverse distances to reflect spatial weights. In the inverse-distance contiguity specification, the type of dependence between regions can be modeled by assigning each urban $j(j \neq i)$ a weight proportional to the inverse distance between i and j. At the same time, we also tested binary order spatial weight matrix and Economic distance spatial weight matrix, and used inverse distance spatial weight matrix to estimate the model in the present paper. The test results can be seen in Section 4.1. Using the inverse distance spatial weight matrix, KS is defined as the weighted mean of knowledge available in neighboring regions after $\boldsymbol{W}$ is normalized as convention. An economically plausible reason for inverse-distance contiguity is information stickiness. Entrepreneurs rely heavily on local knowledge though research and development employment search information across distance in the Internet age. Being different from researchers, entrepreneurs have limited ability to absorb professional knowledge and therefore tend to understand technological progress in a geographic area, which is assumed to be local and neighboring regions. The longer distance between regions, the more difficult it thus becomes for entrepreneurs to absorb new knowledge and transform it into economic benefits.

2.3. Control Variables

A few control variables have to be added to the empirical model to control for systematic differences between theory and the real world due to systematic effects present in the data but neglected by model. The first, $RICH$, is approximated by GDP per capita. It is defined as the effect of economic environment on regional innovation [23]. The second, INS, measures industrial structure by the ratio of service to manufacturing workers in the regional economy. The propensity to patent is usually different between service and manufacturing industries, so INS is added to prevent the output elasticity of R&D employment from impacting the heterogeneous manufacturing and service workers ratio in urban regions. The control variables are also logarithmic in our model.

2.4. Summary

Assuming a non-linear functional form for urban Internet development, INT_{it}, reflects regional innovation outputs and a linear form for the control variables, respectively; the full empirical models to explain regional innovation read:

Regime 1:

$$\begin{aligned} lnI_{it} = \alpha_i + \alpha_1 lnL_{it} \quad & + \theta_1 lnINT_{it} + \rho ln\left(\sum_{j \neq i} w_{ij} I_{jt}\right) + \beta_1 lnRICH_{it} + \beta_2 lnINS_{it} \\ & + \varepsilon_{it}, \ \ lnINT_{it} < \gamma \end{aligned} \tag{5}$$

Regime 2:

$$\begin{aligned} lnI_{it} = \alpha_i + \alpha_1 lnL_{it} \quad & + \theta_2 lnINT_{it} + \rho ln\left(\sum_{j \neq i} w_{ij} I_{jt}\right) + \beta_1 lnRICH_{it} + \beta_2 lnINS_{it} \\ & + \varepsilon_{it}, lnINT_{it} \geq \gamma \end{aligned} \tag{6}$$

where I is the number of applied invention patents; L is the physical R&D inputs, approximated by R&D personnel equivalent to full time equivalent; $RICH$ is approximated by GDP per capita; INS is ratio of service and manufacturing outputs.

Equations (5) and (6) can be written compactly as

$$\begin{aligned} lnI_{it} = \alpha_i + \alpha_1 lnL_{it} \quad & + \theta_1 lnINT_{it} \cdot 1(lnINT_{it} < \gamma) + \theta_2 lnINT_{it} \cdot 1(lnINT_{it} \geq \gamma) \\ & + \rho ln\left(\sum_{j \neq i} w_{ij} I_{jt}\right) + \beta_1 lnRICH_{it} + \beta_2 lnINS_{it} + \varepsilon_{it} \end{aligned} \tag{7}$$

where INT_{it} is core explanatory variable and threshold variable, $1(\cdot)$ is the indicator function, i.e., $1(lnINT_{it} < \gamma)$ takes value one, if $lnINT_{it} < \gamma$ holds; otherwise, it is zero. The threshold parameter $\gamma \in \Gamma$ where Γ is the strict subset of the support of $lnFIRM_{it}$. The major focus of STAR model is the threshold parameter γ, which is unknown in this case.

Since the coefficient of the endogenous variable is not segmented, we can firstly suggest the instruments for spatial lag [24] and then use a least squares estimator as proposed by [17] for the slope parameters and the sum of square errors. Following [24], the instruments for model (7) are $Z = \{X, WX\}$. After replacing the spatial lag term with instruments, run S2SLS estimation on model (7) for each unique threshold value and calculate the sum of square errors $S_n(\gamma)$, following Chan (1993) to estimate threshold value by minimizing the squared errors $S_n(\gamma)$. Find the estimated value of γ that

$$\hat{\gamma} = \underset{\gamma}{\mathrm{argmin}} S_n(\gamma) \tag{8}$$

Then substitute $\hat{\gamma}$ for γ and re-run S2SLS to obtain the same estimate of slope coefficients in different regimes.

Although we write the STAR model with two regimes for simplicity, we do not know the real number of regimes. To determine the number of regimes, we need to test the threshold effect with test statistics for hypotheses $H_0 : \theta_1 = \theta_2$. Following [17], we consider the LR-like statistic

$$F_1 = \frac{S_0 - S_1(\hat{\gamma})}{S_1(\hat{\gamma})} \tag{9}$$

where $S_0 = \hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*$, $\hat{\varepsilon}^*$ is residual vector for the model without threshold effect ($\theta_1 = \theta_2$) and $S_1 = \hat{\varepsilon}'\hat{\varepsilon}$, $\hat{\varepsilon}$ is residual vector for the model (7). Similarly, we consider the LR-like statistic for hypotheses $H_0 : \theta_1 = \theta_2 = \ldots = \theta_{n+1}$

$$F_n = n\frac{S_0 - S_n(\hat{\gamma})}{S_n(\hat{\gamma})} \tag{10}$$

For the classical threshold model, the sum of the square error S_0 can be estimated with OLS, and $S_n(\hat{\gamma})$ can be estimated with threshold estimator proposed by [25], but the estimators may not hold in the context due to the spatial lag term in STAR. We derived both the threshold estimator and instrument variable estimator and ran an S2SLS (spatial two-stage least squares) estimator for our model to obtain S_0 and $S_n(\hat{\gamma})$ in Equations (9) and (10).

3. Patents and Regional Internet Data

For consistency, availability and objectivity of data, panel data from 31 provinces of China from 2014 to 2018 were selected as study samples. The main data on regional innovation inputs and number of pending patents come from Chinese Science and Technology Statistical Yearbook and Chinese Statistical Yearbook.

The number of filed patents, the number of scientific research papers and the sales revenue of new products can be categorized as innovation. Similar to previous work [26], we used the number of filed invention patents as an indicator of regional innovation. Although this has some shortcomings as an indicator, as noted by Griliches (1979), it is still considered one of the most reliable measures of regional innovation due to its availability and integrity. There are three types of patents in China including the invention patent, utility model patent and design patent. To eliminate the interference of patent quality diversity we only used invention patents, which usually have the highest technical content and the most rigorous examination procedures.

China does not have many indicators for regional Internet development. Some studies use the level of informatization, but the concept of informatization covers a larger area than that of the Internet. In addition, the Internet penetration rate has also been used as an Internet index in some studies [27]. However, Internet penetration is likely to correlate with infrastructure construction and has obvious space-time convergence at the regional level, so it cannot objectively reflect the actual level of regional Internet business adoption. Considering the availability of data, we selected the number of websites owned by local firms as an indicator of the level of Internet business adoption in a region. Currently, Chinese firms' offline innovative activity is often synchronized with online development. As we focused on mechanisms by which the Internet promotes regional innovation by stimulating local firms' innovative activity, the number of websites should be a better indicator for business adoption than Internet penetration. The descriptive statistics of independent and dependent variables can be seen in Table 1.

Table 1. Statistics.

Variable	Obs	Mean	SD	Min.	Max.
lnI_{it}	155	9.622	1.557	4.522	12.285
lnL_{it}	155	11.497	1.292	7.655	13.688
$lnRICH_{it}$	155	10.871	0.400	10.172	11.851
$lnINS_{it}$	155	−0.147	0.366	−1.471	0.405
$lnINT_{it}$	155	9.192	1.182	5.814	11.224
$lnFIRM_{it}$	155	9.843	1.136	6.392	11.733

4. Spatial Threshold Econometrics Issues

The Spatial threshold autoregressive (STAR) model, a combined panel data spatial autoregressive (SAR) with threshold model, was faced with a series of econometric issues such as identification of the threshold effect and spatial dependence. Although both issues have been solved in the research on the threshold model and SAR model, respectively, we needed to test whether there was spatial dependence or a threshold effect for the STAR model.

4.1. Tests of Spatial Dependence

We studied the global spatial dependence of regional innovation by Global Moran's I statistic, which measures spatial autocorrelation based on locations and feature values and characterizes the clustering as a whole. The Moran's I is defined as follow:

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y_i - \overline{y}) (y_j - \overline{y})}{S^2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \tag{11}$$

where $S^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2$; $\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$; y_i refers to the logarithmic amount of filed patents in region i and w_{ij} is the element in spatial weighted matrix W. Here we emphasize that we respectively use the binary order spatial weight matrix W_1, inverse distance spatial weight matrix W_2 and economic distance spatial weight matrix W_3 introduced above to calculate w_{ij} in the Moran's I statistic. Moreover, $I \in (-1, 1)$ and when I is close to zero, it indicates weak spatial dependence or no spatial dependence. When I is near one or negative one, it indicates strong positive or negative dependence among regions. We also use standardized statistics z to test the significance level of global spatial dependence; z is defined as follow:

$$z = \frac{[I - E(I)]}{\sqrt{Var(I)}} \tag{12}$$

where $E(I) = -1/(n-1)$ refers to theoretical expectations of Moran's I and $Var(I)$ refers to theoretical variance of Moran's I.

Table 2 reports the Moran's I statistic of innovation using three types of spatial weight matrix for 31 regions in China between 2014 and 2018. The Moran's I statistic of innovation is positive and all p values for the Moran's I statistic are less than 10%, which suggests that there is a significant positive spatial dependence of regional innovation in China. Comparing the values of Moran's I statistics using different weight matrices the value with W_3 is the smallest, which suggests that economic differences offset part of the geographical correlation, making the spatial dependence of regional innovation smaller. Furthermore, the value with W_1 and W_2 rises steadily from 2014 to 2018, while the value with W_3 shows an upward trend of fluctuation. We also test the local spatial dependence and plot scatter diagrams from 2014 to 2018 (They are not shown in this paper due to limited space). In the scatter diagrams, innovation in most provinces is distributed in the first and third quadrants, which further suggests the characteristic of high spatial agglomeration of regional innovation. The scatter diagrams also show obvious characteristics of spatial dependence that is not a random distribution, both of which further reveal the existence of spatial knowledge spillovers. In summation, it is necessary to pay attention to the spatial dependence of innovation when exploring the relationship between regional Internet business adoption and innovation, otherwise the estimation results will be biased.

Table 2. The Moran's *I* of regional innovation.

Year	Binary Order Spatial Weight Matrix			Inverse Distance Spatial Weight Matrix			Economic Distance Spatial Weight Matrix		
	I	*z*	*p*	*I*	*z*	*p*	*I*	*z*	*p*
2014	0.290	2.843	0.002	0.186	2.440	0.007	0.130	2.002	0.023
2015	0.292	2.845	0.002	0.192	2.496	0.006	0.124	1.926	0.027
2016	0.293	2.847	0.002	0.205	2.629	0.004	0.114	1.792	0.037
2017	0.313	2.987	0.001	0.216	2.725	0.003	0.081	1.384	0.083
2018	0.349	3.287	0.001	0.260	3.190	0.001	0.139	2.074	0.019

Note: *I* is global Moran's statistic; *z* is standardized global Moran's statistic obeying asymptotic standard normal distribution.

4.2. Tests of Threshold Effect

For the threshold model, we must determine the number of the threshold value. If we test threshold effect directly using the classical testing measure proposed by Hansen (1999), we have to ignore the spatial lag term $\sum_{j\neq i} w_{ij} lnI_j$ or regard it as exogenous, which may lead to a false test result. Since the spatial lag dependent variable is correlated with the error term, we suggest the instrument proposed by [24] to replace $\sum_{j\neq i} w_{ij} lnI_j$ in model (7); the accuracy of the test derived in Hansen (1999) for the threshold model can hold in this context. After replacing the spatial lag dependent variable with instrument variable $Z = \{WX\}$, we use a bootstrap method to repeat the sampling 300 times to the estimated p-value. The result is shown in Table 3, which shows a single threshold value for the STAR model. We use a single-threshold spatial autoregressive model since the p-value of the single-threshold model is significant at the 5% level, whereas the double-threshold and triple-threshold models do not pass the significance test. We also obtain the threshold estimator (8.1719) and its upper and lower estimators (8.1719 and 7.7415) at the 95% level. The results suggest that when the logarithm of the number of firms' websites in the region is greater or lesser than 8.1719, there will be corresponding effects of the Internet on the regional innovation system.

Table 3. Test of Internet's threshold effect.

Model	*F*-Value	*p*-Value	Critical Value		
			10%	5%	1%
Single-threshold	26.15	0.0433	21.1121	23.8455	36.5259
Double-threshold	14.92	0.2667	23.4736	27.4578	33.5923
Triple-threshold	12.38	0.3367	21.1642	24.1024	31.4249

5. Empirical Results and Discussion

5.1. Empirical Results

In this section, we first report our baseline specification. For our baseline specification, we ran the panel data SAR model without the threshold effect of the Internet on regional innovation. Further, we used the estimated threshold value $\hat{\gamma} = 8.1719$ to divide the sample into two segments of which the threshold variable is higher or lower than the estimated threshold value. We also report the TSAR model results. Table 4 reports the estimation results of the threshold and spatial autoregressive models for the effect of Internet and interregional knowledge spillover on patent output.

Firstly, we see that both region-specific resources of R&D employment lnL_{it} and interregional knowledge spillovers $WlnI$ have a positive effect on regional innovation outputs (the number of new filed patents) and coefficients of lnL_{it} and $WlnI$ in all models passing the significance test for p-value less than 1%. This fully shows that interregional knowledge spillovers actually contribute to regional innovation. Compared with interregional knowledge spillovers, research and development employment plays a more important role in regional knowledge production, which corresponds with the development stage in China now that high concentration of such employment in the eastern area contributes to a gap in innovation between regions.

Table 4. Regression results for SAR model and TSAR model.

Variable	SAR without Threshold Effect	TSAR
lnL_{it}	0.984 *** (0.171)	0.802 *** (0.180)
$WlnI$	0.254 ** (0.100)	0.265 *** (0.100)
$lnINT_{it}$	0.007 (0.164)	—
$lnINT_{it} \cdot 1(lnINT_{it} < \gamma)$	—	0.473 ** (0.206)
$lnINT_{it} \cdot 1(lnINT_{it} \geq \gamma)$		0.378 ** (0.192)
$lnRICH_{it}$	0.447 ** (0.223)	0.451 ** (0.220)
$lnINS_{it}$	−0.193 (0.152)	−0.284 * (0.149)
cons	−9.156 *** (1.787)	−10.764 *** (1.806)

Note: Standard errors in parentheses. * Significance at 10%level. ** Significance at 5%level. *** Significance at 1% level.

Secondly, to account for possible heterogeneity, we used the threshold estimation approach. Tests of the threshold effect above have shown that there is a single threshold value for the Internet's effect on regional innovation, and we have obtained a threshold value of 8.1719. We differentiate regions with high levels of Internet business adoption from other regions and investigate the Internet's respective effects. More specifically, $\alpha_i, \alpha_1, \rho, \beta_1, \beta_2$ should be same for both regimes. Hence, we perform spatial estimation on model (7) using the estimated threshold value to replace unknown γ so that all the coefficients can be jointly estimated. Looking at Column 3 in Table 4, when $lnINT_{it}$ is higher than 8.1719, the coefficient of $lnINT_{it}$ on explained variable is 0.378, passing the significance test at 5%, which suggests Internet business adoption has a significant promotional effect on innovation. When $lnINT_{it}$ is lower than 8.1719, the coefficient of $lnINT_{it}$ on explained variables is 0.473, passing the significance test at 5%, which shows that the promotional effect of the Internet is enhanced. Consequently, there is an inverted U-shaped positive relationship between the level of Internet business adoption and regional innovation outputs. With the development of Internet business adoption, its positive effects on regional innovation are reduced.

Thirdly, we simply discuss the coefficients of the control variables on the explained variable. The coefficients of $lnRICH_{it}$ in SAR and TSAR models are similar, and both are significantly positive at the 5% level. For every 1% increase in regional GDP per capita, the number of new filed patents increases about 0.45%. Innovation usually requires a large amount of capital investment, and a good economic environment can widen the financing channels of firms. The coefficients of $lnINS_{it}$ in the SAR and TSAR models is negative because the service sector has now become the main driver of innovation in China.

5.2. Discussion

Building on the TSAR model, this study examines the nonlinearity and spatial dependence of regional innovation. The study here includes the findings that both local Internet business adoption and interregional knowledge spillovers have a positive effect on regional innovation outputs. However, only considering the threshold effect of Internet can we obtain the statistically significant coefficient. The study also confirms the existence of a single threshold effect of the Internet simultaneously considering spatial dependence, which is also confirmed.

The findings contribute to the existing literature in several ways. First, the study helps to unpack the black box of the Internet's impact on regional innovation by empirically delimitating local Internet business adoption scales that determine the Internet's promotional effect. The findings provide support for the view that the Internet has a heterogeneous effect on regional innovation. Using different kinds of threshold variables such as economic development, government funding, industry structure and trade openness, the Internet has significant positive and non-linear effect on regional innovation. Based on the previous literature, the study uses the Internet as a threshold variable and finds that Internet business adoption has an inverted-U positive effect on regional innovation. This finding suggests that regions with low levels of Internet business adoption have more opportunity to catch up with developed regions in innovation. Consequently, Internet business adoption may contribute to bridging interregional innovation gaps.

Second, the study confirms that ignoring interregional knowledge spillover may cause misjudgment of the effect of factors in the regional innovation system. Although interregional knowledge spillover has been confirmed as a determinate factor on innovation [16,17,28], it is still difficult to use empirical research to estimate the threshold and spatial spillover effects. When we only estimate the threshold effect of the Internet on regional innovation, we may obtain an incorrect threshold number, incorrect threshold value or incorrect coefficients for regimes due to ignorance of spatial lag terms.

6. Conclusions and Policy Implications

The use of the Internet has increased tremendously since the middle of the 1990s. Although some literature provides policy suggestions for regional Internet and innovation, empirical evidence that the Internet has positive effects on regional innovation is still scarce. This paper studies whether the impact of Internet business adoption on regional innovation has had threshold effects when the interregional knowledge spillovers are considered in one framework. We also postulate an explanation based on the perspective of knowledge and information flow among local research and development and application firms. To test our results empirically, we use a panel data TSAR model, which may be similar to other studies. We draw three main conclusions: (1) There are significant and positive interregional knowledge spillovers influencing regional innovation. This means that innovative activity in a certain regional innovation system is correlated to the innovation outputs of its contiguous areas. Even in the Internet age, interregional knowledge spillovers are still limited by geographical distance and economic distance. (2) Regional increases in innovation are dependent on Internet business adoption. This is due to the lowering of costs related to information and knowledge flow, which can decrease uncertainty associated with innovative activity and increase innovative efficiency. (3) Internet business adoption has had a threshold effect on regional innovation in China and when Internet business adoption reaches threshold value, its promotional effect decreases. Generally, results suggest that the nonlinear effects of Internet business adoption on regional innovation are positive and significant for lower-Internet-adoption regions. It also suggests that the Internet's reinforcement of knowledge and information flow is limited due to technical distance between firms. The Internet helps knowledge and information to flow among firms within regions, overcoming geographical distance, but barriers from cognitive and technical differences cannot be overcome by the Internet, which determines the upper limit of the Internet's effect on regional innovation.

Policy implications for these conclusions suggest that if there were a minimum level of Internet business adoption, there could potentially be a significant increase in regional innovation outputs in China. Therefore, the government should not only pay attention to Internet infrastructure development but also encourage Internet business adoption by research and development and application firms. At the same time, we have also noted that Internet business adoption might help decrease interregional innovative divides due to its threshold effect on regional innovation. We use the 31 Chinese provinces as a case study in this paper; however, our methodology and analysis could be extended to other countries.

In addition to the problem addressed in this paper, the TSAR model could be applicable to other research fields where there are spatial spillovers and threshold effects. For example, the influence of environmental regulation on technological progress can also be modeled by TSAR. Our analysis is subject to limitations. For one, we studied a specific time set of five recent years. The impact of Internet business adoption on regional innovation may be different over a later time period due to ICT progress; innovations such as 5G, AR, and AI may change the mode of knowledge and information flows. Second, we selected only firms to represent the actors of regional innovation system. More work may examine how Internet adoption by local governments, universities, and research institutions influences regional innovation systems if researchers can find good indicators to describe more kinds of Internet adoption. Another limitation is that our research focused primarily on the threshold effect on regional knowledge production. Future work should test whether there is a threshold effect on interregional knowledge spillovers. Research on this topic would provide new insights on whether the Internet can help overcome localization of knowledge and narrow the innovative divide at the regional level.

Author Contributions: C.C. conceived and designed the study, and performed the analytical model. A.Y. reviewed and edited the paper. All authors discussed the results and implications and commented on the paper at all stages. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [the National Natural Science Foundation of China] grant numbers [72073030] and [71571046].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: https://data.cnki.net/yearbook/Single/N2020100004 (accessed on 1 July 2020).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, C.; Wang, X. The influence of ICT-driven innovation: A comparative study on national innovation efficiency between developed and emerging countries. *Behav. Inform. Technol.* **2019**, *38*, 876–886. [CrossRef]
2. Zhang, F.; Li, D. Regional ICT access and entrepreneurship: Evidence from China. *Inf. Manag.* **2018**, *55*, 188–198. [CrossRef]
3. Whelan, E.; Teigland, R.; Donnellan, B.; Golden, W. How Internet technologies impact information flows in R&D: Reconsidering the technological gatekeeper. *R D Manag.* **2010**, *40*, 400–413.
4. Huang, M.; Bhattacherjee, A.; Wong, C. Gatekeepers' innovative use of IT: An absorptive capacity model at the unit level. *Inform. Manag.* **2018**, *55*, 235–244. [CrossRef]
5. Terhorst, A.; Lusher, D.; Bolton, D.; Elsum, I.; Wang, P. Tacit Knowledge Sharing in Open Innovation Projects. *Proj. Manag. J.* **2018**, *49*, 5–19. [CrossRef]
6. Hildrum, J.M. Sharing Tacit Knowledge Online: A Case Study of e-Learning in Cisco's Network of System Integrator Partner Firms. *Ind. Innov.* **2009**, *16*, 197–218. [CrossRef]
7. Buunk, I.; Smith, C.F.; Hall, H. Tacit knowledge sharing in online environments: Locating 'Ba' within a platform for public sector professionals. *J. Libr. Inf. Sci.* **2019**, *51*, 1134–1145. [CrossRef]
8. Billon, M.; Lera-Lopez, F.; Marco, R. Patterns of Combined ICT Use and Innovation in the European Regions. *J. Glob. Inf. Technol. Manag.* **2017**, *20*, 28–42. [CrossRef]
9. Zhang, N.; Zhao, X.; Zhang, Z.; Meng, Q.; Tan, H. What factors drive open innovation in China's public sector? A case study of official document exchange via microblogging (ODEM) in Haining. *Gov. Inf. Q.* **2017**, *34*, 126–133. [CrossRef]
10. Paunov, C.; Rollo, V. Has the Internet Fostered Inclusive Innovation in the Developing World? *World Dev.* **2016**, *78*, 587–609. [CrossRef]
11. Forman, C.; van Zeebroeck, N. Digital technology adoption and knowledge flows within firms: Can the Internet overcome geographic and technological distance? *Res. Policy* **2019**, *48*, 103697. [CrossRef]
12. Sorenson, O.; Fleming, L. Science and the diffusion of knowledge. *Res. Policy* **2004**, *33*, 1615–1634. [CrossRef]
13. Jaffe, A.B. Real effects of academic research. *Am. Econ. Rev.* **1989**, *79*, 957–970.
14. Anselin, L.; Varga, A.; Acs, Z. Local geographic spillovers between university research and high technology innovations. *J. Urban Econ.* **1997**, *42*, 422–448. [CrossRef]
15. Bode, E. The spatial pattern of localized R&D spillovers: An empirical investigation for Germany. *J. Econ. Geogr.* **2004**, *4*, 43–64.
16. Moreno, R.; Paci, R.; Usai, S. Spatial spillovers and innovation activity in European regions. *Environ. Plan. A Econ. Space* **2005**, *37*, 1793–1812. [CrossRef]
17. Hansen, B.E. Threshold effects in non-dynamic panels: Estimation, testing, and inference. *J. Econ.* **1999**, *93*, 345–368. [CrossRef]
18. Lundvall, B.A. Why study national systems and national styles of innovation? *Technol. Anal. Strateg.* **1998**, *10*, 407–421. [CrossRef]

19. Etzkowitz, H.; Leydesdorff, L. The dynamics of innovation: From National Systems and "Mode 2" to a Triple Helix of university-industry-government relations. *Res. Policy* **2000**, *29*, 109–123. [CrossRef]
20. Rosenblat, T.S.; Mobius, M.M. Getting closer or drifting apart? *Q. J. Econ.* **2004**, *119*, 971–1009. [CrossRef]
21. Van Alstyne, M.; Brynjolfsson, E. Global village or cyber-balkans? Modeling and measuring the integration of electronic communities. *Manag. Sci.* **2005**, *51*, 851–868. [CrossRef]
22. LeSage, J.P.; Pace, R.K. Spatial econometric modeling of origin-destination flows. *J. Reg. Sci.* **2008**, *48*, 941–967. [CrossRef]
23. Gao, X.; Guo, X.; Sylvan, K.J.; Guan, J. The Chinese innovation system during economic transition: A scale-independent view. *J. Inf.* **2010**, *4*, 618–628. [CrossRef]
24. Kelejian, H.H.; Prucha, I.R. Generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J. Real Estate Finance Econ.* **1998**, *17*, 99–121. [CrossRef]
25. Chan, K.S. Consistency and limiting distribution of the least-squares estimator of a threshold autoregressive model. *Ann. Stat.* **1993**, *21*, 520–533. [CrossRef]
26. Xu, X.; Watts, A.; Reed, M. Does access to internet promote innovation? A look at the US broadband industry. *Growth Chang.* **2019**, *50*, 1423–1440. [CrossRef]
27. Kumar, R.R.; Stauvermann, P.J.; Samitas, A. The effects of ICT* on output per worker: A study of the Chinese economy. *Telecommun. Policy* **2016**, *40*, 102–115. [CrossRef]
28. Caragliu, A.; Nijkamp, P. Space and knowledge spillovers in European regions: The impact of different forms of proximity on spatial knowledge diffusion. *J. Econ. Geogr.* **2016**, *16*, 749–774. [CrossRef]