Article

Digital Inequalities in China in 2020: Spatial and Multivariate Analysis

James Pick *, Fang Ren and Avijit Sarkar

School of Business and Society, University of Redlands, Redlands, CA 92373, USA; fang_ren@redlands.edu (F.R.);
avijit_sarkar@redlands.edu (A.S.)
* Correspondence: james_pick@redlands.edu

Abstract: China's digital divide is explored through spatial and multivariate analysis. The dependent variables include general uses of information and communication technologies (ICTs) and mobile devices, measures of ICT infrastructure, purposeful uses for software services, and purposeful uses of e-commerce. Independent variables include a variety of demographic, economic, educational, ethnic, innovation, and knowledge production indicators. Data come from the China Yearbook. Theoretically, the study draws on the Spatially Aware Technology Utilization Model (SATUM). Digital disparities in Chinese provinces are analyzed using regression analysis, spatial autocorrelation, and k-means cluster analysis. The important correlates associated with digital inequality are expenditure for science and technology, income, R&D investment, full-time teachers, non-Han (minority) population, and proportion of urbanization. Longitudinal comparison reveals factors influencing ICT changes to be expenditure on science and technology, the unemployment rate, and college degree. Among the cluster findings are geographic concentrations of digital variables in Shanghai and Beijing and marked spatial pattern differences in central and central–east China between variable groups. Provincial and national policy implications, which are compared to China's five-year plans, include an emphasis on science and technology, technology education in all provinces, support for higher provincial incomes, and ICT training for the non-Han population. These steps are especially important in ICT-deprived provinces.

Keywords: information technology; digital divide; digital inequality; China; spatial; cluster; regression; SATUM model; spatial autocorrelation

1. Introduction

This paper studies the spatial patterns of the digital inequalities in China for an array of information and communication technology (ICT) and internet variables and examines demographic, socioeconomic, and innovation influences on digital inequalities in Chinese provinces. China's population was 1.45 billion in 2022, of whom 1.01 billion are internet users, representing a 70% internet penetration. Chinese internet users represent over one-third of Asia's internet user base. In 2017, China was the world's largest telecommunication market in terms of the number of mobile, fixed-telephone, fixed-broadband, and mobile broadband subscriptions. In addition, it was the world's leading exporter of ICT products in 2021 [1]. With this huge telecom market and expansive base of internet users, internet and e-commerce companies, such as Tencent, Alibaba, Baidu, and Bytedance, have become household names not only in China but all over the globe. Driven by the growth of internet and cybertechnologies, these companies have fueled a tremendous growth of e-commerce in China, which is consequently leading to social and economic transformations [2].

Despite this market size, the adoption and diffusion of ICT within China have consistently shown disparities. A study determined that significant provincial digital disparities exist between eastern and western Chinese provinces as well as between core cities and more peripheral ones [3]. Prior to 2010, university education, urban residential income,

and the urbanization rate were shown to be important educational and economic factors influencing the Chinese provincial digital divide. However, since 2010, adult literacy rate and rural residential income were identified to be key determinants of this divide [3]. Similar regional disparities were observed for e-commerce adoption, the key determinants of which were linked to local economic, political, and infrastructural conditions [2].

This study builds on prior work on the Chinese digital divide that has encompassed a broad array of ICT dependent variables, including ICT use, digital infrastructure, ICT services, and e-commerce. The present research has novel features that address gaps in prior research, including little inclusion of infrastructure factors; absence of analysis of longitudinal change in the correlates of digital divide factors; and lack of investigation of purposeful technology uses in ICT services and e-commerce. This study fills these gaps through more detailed spatial cluster analysis; inclusion of new, previously unreported variables, such as purposeful uses in ICT services and e-commerce; and inclusion of the influence of ethnicity.

The following research questions are formulated to address these research gaps:

1. What are the important demographic, economic, education, innovation, and knowledge determinants of digital infrastructure, ICT use, and purposeful uses of ICT in the services and e-commerce sectors?
2. What are the changes in the determinants of ICT and internet factors over the period of 2015 to 2020?
3. What are the spatial cluster patterns of digital inequalities segmented by ICT use, digital infrastructure, ICT services, and e-commerce?

After addressing the research questions, the study examines the implications of the research findings for national policymaking regarding the internet and ICT in China that could serve to reduce provincial digital inequalities.

The remainder of this paper is organized into the following sections: literature review, conceptual model of digital inequities, data and methodology, descriptive and regression findings, longitudinal analysis of change, cluster analysis, discussion and implications, and conclusion.

2. Literature Review

A digital inequality is defined as the unequal use of digital technologies, which extends to unequal impacts and outcomes. The related concept is the digital divide, which represents the gap in the utilization and benefits of digital technologies between people, organizations, and governments. The need to close the digital divide was recognized by the UN in its 2016 announcement that the internet is a human right and that efforts to remove internet access are a violation of human rights [4]. Its importance has been evident during the COVID-19 period in the disparities between and within communities in accessing the internet for family needs, including schooling, health information, social contacts, and commercial services.

Digital inequality is a complex phenomenon that has been formulated theoretically to include a range of types of access, cultural factors, socioeconomic and political aspects, and the importance of digital learning. A conceptual approach to access to digital technologies is to consider a sequence starting with motivation and extending to physical infrastructure, digital skills, and usage [5–8]. These laddered elements were encompassed in a complex model of digital media under the category of "access", along with personal characteristics, resources, positional categories, and participation outcomes [6].

The present study model includes the ladder elements of physical access, usage, and participation outcomes. In digital inequality/digital divide research, low-level physical access has become prevalent or even ubiquitous, while inequalities persist in purposeful uses and societal outcomes [9]. Our study examines both the low-level physical access digital factors and the purposeful use of ICT at the individual and organizational levels.

China's ICT growth can be classified into a succession of formative, developmental, and mature stages, with the formative stage ending around 2005, followed by a developmen-

tal stage until the mid-2010s, and finally a mature stage [10]. Throughout this progression, the central government adopted a top–down approach, seeking a balance between ICT for government control and improvement in economic prosperity. As China's exports grew significantly after 2000, businesses increasingly adopted ICT in order to participate in international supply chains and markets [11]. Yet, it is essential to recognize provincial unevenness and address the low levels of ICT usage among aged, undereducated, and rural, poor individuals [11,12].

Another strand of the literature has considered the interplay of the internet and ICT with aspects of the Chinese economy. As e-commerce expanded, Chinese consumers experienced unequal access and use of it depending on demographic characteristics, including age, gender, college education, and urban residency [13]. Geographically, there was greater online shopping in eastern coastal areas compared to western provinces [14]. A spatially driven, fixed-effects analysis found that ICT is directly related to socioeconomic development and that rapid prominence of ICT in one region can negatively influence ICT growth in neighboring provinces, termed a "spillover" effect [15]. Recently, a study found that broadband bolstered the nation's economic growth more during the pandemic compared to former "normal" times [16]. Our research contributes to the existing literature by examining the purposeful use of ICT on e-commerce and IT services and providing a detailed spatial pattern analysis of how ICT infrastructure, use of ICT, e-commerce, and IT services have been developed across the provinces.

China studies have also examined the spatial patterns in usage of the internet and ICT among the provinces through different approaches. In a study of China's rapid ICT growth from 2003 to 2008, location quotient (LQ) analysis was applied to map the ratios of provincial to national internet use, revealing provincial use differences as high as 7-fold, i.e., between Beijing and Tibet. In a spatial typology of regions, the internet LQ level varied from highest in the major cities to the southwest coast and southwest interior to lowest in the vast western and northern provinces [17]. Another study emphasizes wide urban–rural disparities in infrastructure and ICT use in the early 2000s [12]. A more detailed spatial study, based on the SATUM model, revealed the provincial determinants of ICT [18]. Another provincial study, utilizing an ICT index as the dependent variable, mapped ICT uses cross-sectionally in 2000 and 2010 in China [19], identifying the large provincial differences in the index, and assessed the relationship between ICT levels and economic growth. We update their investigations to 2020.

At a finer scale, research on 291 Chinese cities, drawing on 2010–2017 government and scraped website data, identified socioeconomic determinants of a "digital development index" (DDI) and classified China into five DDI regional clusters [3]. In sum, this body of non-spatial and spatial literature on China provides a baseline of prior findings for the present spatial study, which focuses on Chinese data in the 2020 pandemic year and in 2015 for a broader range of dependent digital factors.

This review has covered the broad context of study of digital inequalities and divides and the evolving research on digital inequalities in China. This study seeks to fill gaps in the literature of limited studies of China's provincial digital inequalities in the 2020s, the paucity of longitudinal studies, and the lack of research on purposeful uses of ICTs for China's provinces.

3. Conceptual Model

Approaches to modeling digital inequalities for China range from econometric models [15,16] to simple frameworks of ICT use [12,17] to spatial analytics [3,18]. Apart from China, more complex conceptual models of digital inequalities and divides include digital divide multi-level behavioral modeling [6,9], digital divide units of analysis and access–use progression [20], and spatial modeling [8,21].

Economic approaches tend to focus on digital inequalities among economic indicators, e.g., a study that analyzed and confirmed the effects of ICT on penetration levels of GDP,

technology cost, trade, and education [20]. Other studies have examined ICT impacts on economic growth based on time series, fixed effects, and econometric techniques [15,16].

Complex causal–sequential models incorporate into a multi-level model the digital inequality dimensions of digital access, capital and human resources, individual characteristics, and impacts/outcomes [6]. This holistic framework has the advantage of simulating sequential aspects over varied units of analysis (individual, organization, or society). However, obtaining sufficient data to test it is sometimes challenging. Consideration of technology development over time can employ the approach of adoption–diffusion modeling, which originated from Rogers [22], but this model fails to consider the broader social, economic, and cultural forces impacting events and for which longitudinal data often are not available or limited [21]. The same is true for technology acceptance theories, such as the Technology Acceptance Model (TAM), the Unified Theory of Acceptance and Use of Technology (UTAUT), and the Model of Adoption of Technology in Households (MATH), which are often deployed to model ICT use, including internet use [23], or to study transformational government in the United States [24].

The Spatially Aware Technology Utilization Model (SATUM) posits associations induced from the digital divide/inequality literature or justified by author reasoning. It proposes relationships of independent socioeconomic, infrastructural, innovation, and social capital variables with dependent factors that directly or indirectly include ICT and the internet [21,25–29]. SATUM includes the methods of descriptive mapping, cluster analysis, OLS regression, and spatial autocorrelation of dependent factors and model residuals. It is suitable for digital inequality investigations that focus on geographic disparities but does not provide for multi-level or bi-directional analysis.

The SATUM model is adopted for the present research due to its geographic approach, its feasibility for studying national states or provinces, its inclusion of spatial clustering for understanding digital inequalities, and its incorporation of testing for spatial bias and longitudinal analysis. The modified operational SATUM model for this study is presented in Figure 1. The figure on the left shows the exploratory analysis of spatial patterns for ICT variables as conducted using two methods [21]. One is k-means cluster analysis, which measures the groupings of geographical units with the least distance between provinces based on a given k number of clusters, while the second, Moran's Index, tests to what extent, for a single variable, like units aggregate together (positive value of the index) or are repelled from each other (negative value of the index). At the figure's lower right, the confirmatory analysis of ordinary least squares regression appears, which connotes the association of the independent variables with each dependent one. The Moran's I test, computed based on the regression residuals, evaluates if the model removes a spatial bias, leading to random residuals. If they are not random, regression findings must be treated cautiously, as the model has spatial bias [25]. The residuals also undergo standard OLS residual testing through the Joint Wald, Koenker, and Jarque Bera statistics.

SATUM's dependent variables are justified by their inclusion in the prior relevant literature. They are organized into four groups, which include the evolutionary stages of digitalization, which are access (infrastructure), use, and purposeful use (IT services and e-commerce) [6,9]. These two purposeful use categories were selected due to the scarcity of purposeful use variables in government data sources, while they represent important dimensions of the digital economy. The groups are as follows.

Infrastructure. Digital infrastructure is necessary for digital access, regardless of the type of device. It was included in the comprehensive digital divide model of van Dijk [6] and mentioned in economic models [. Nationwide information infrastructure was related to IT governance and socioeconomic development for a large sample of developing nations [30]. In the Chinese literature, infrastructure was shown to have grown significantly since 2000, a trend that partly reflects the support from China's central government [31]. Rapid growth in the nation's technology infrastructure during the years 2004 to 2014 can be linked to planning and investment by the government and major telecom enterprises [10].

In the same decade, electrical output was shown to have a modest influence on the use of mobile phone subscribers in Chinese provinces [18].
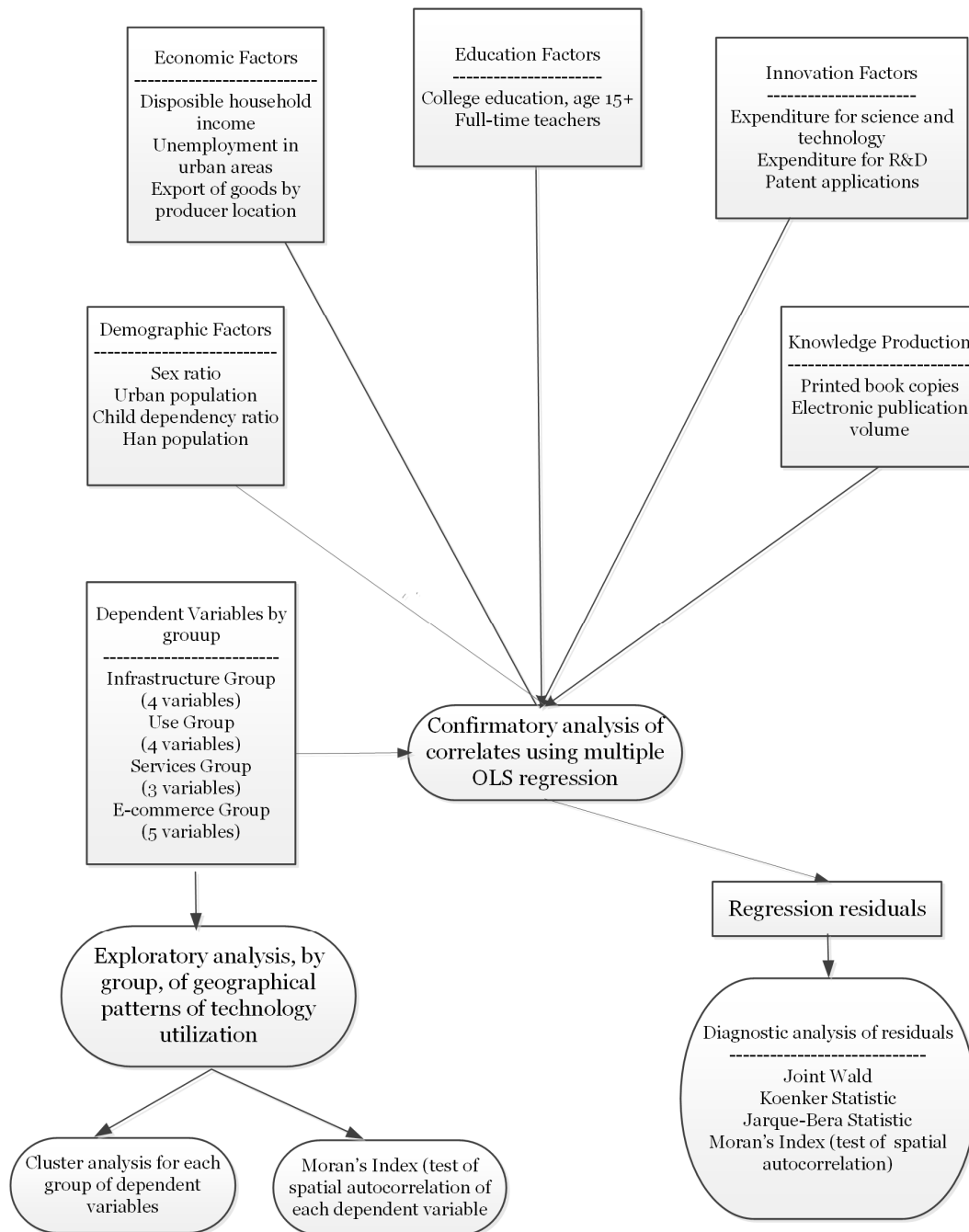
Figure 1. SATUM model for digital inequalities in China, 2020.

The following infrastructure-dependent variables are included in the present study: length of optical cable lines, base stations of mobile phones, capacity of mobile phone exchange, and number of domain names. The latter was included in a provincial study of China using government data from 2006 to 2009 [18].

Use. In prior studies of the Chinese provinces, use factors served as dependent factors measured based on mobile devices [3,12,18,32,33], broadband [3,19], the internet [12,17–19], and web pages [17,18]. Furthermore, these dependent variables for digital use have been frequently included in digital divide studies apart from China [6,32–34].

IT Services. IT services have rarely been included in digital divide research on China. They represent an important type of ICT purposeful use [35]. A study examined growth factors in domestic IT services in China and found that although technology infrastructure had grown substantially by 2012, IT services struggled to effectively support IT solutions [36]. However, the IT service revenue in China doubled in size from 2016 to 2023 to reach an estimated revenue of USD 25.6 billion in 2023 [37,38]. A study of early IT services' growth in developing nations suggests that IT services are likely to vary between technologically and economically advanced provinces and those that are less developed [39].

E-Commerce. E-commerce utilization in China has been examined at different levels, including at the individual level in 2011 [13] the provincial level for 2013–2016 [2], and for e-shopping usage in 2017–2019 [14]. At the individual level, a 2011 survey showed access to e-commerce was related to age, college education, female sex, and rural migration [13]. A study of individual e-shopping adoption based on Alibaba online shopping data at the county level for 2017 indicated that the characteristics associated most strongly with e-shopping were a local delivery system, income, the college enrollment ratio, and transportation accessibility [14]. Our study builds on this base by analyzing sales through e-commerce, purchases through e-commerce, and percent of enterprises with e-commerce transactions.

Next, we justify the independent variables, which are likewise divided into five groups of factors in the model. These groups are commonly seen in digital divide research, with the exception of knowledge production, which can be included through proxy variables.

Demographic. Urban population is commonly incorporated into digital inequality models for China [3,13]. It serves as a control for the large provincial differences in urban proportions. Sex ratio has also been included in prior studies [13,14] with significant results, which relates to China's relatively high sex ratios, i.e., more males than females on a national basis, potentially affecting the size of the technology workforce and ICT usage. Two other variables added to this group are unique for digital divide studies on China, namely the Han population and the young child dependency ratio. The Han population as a percent of the total population is included because US digital divide studies, which examined independent ethnicity factors, found them to be significant determinants of digital usage [34,40,41]. We posit that the Han proportion is linked to higher usage, as the Han majority in China is known to have higher educational and economic levels.

The young child dependency ratio is added because greater proportions of children in certain provinces can influence technology adoption, including, for instance, expansion in ICT use due to the presence of a relatively younger population. The child dependency ratio has been significant in studies of Japan [42] and the US [34,43].

Economics. A variety of income factors have been associated with ICT use across varied geographies of China. In particular, residential income was a major correlate of online shopping in Chinese counties [14], and urban and rural residential income were important correlates of a digital divide for prefectures in China [3]. Another economic factor, export of goods, was the leading determinant of ICT availability and usability in China's provinces [18]. We argue that it influences ICT use through the exporting process by being linked to advanced ICT use in the global supply chain.

Although unemployment has seldom been incorporated in digital inequality studies, it was included and shown to be inversely related to the proportion of internet users in a study of digital inequality in the UK [44]. Correspondingly, employment was a positive correlate for certain ICT variables in China in 2006 and 2009 [18]. We include unemployment, reasoning that unemployed persons tend to have lower work impetus to use information technology but may have a higher demand for personal communications. In China, the unemployment measure excludes unemployed workers without official state registration, and some people who are registered as unemployed may actually be working or out of the labor force [45].

Education. Educational attainment has long been established as an ICT determinant [26]. In a global study, secondary and tertiary education were significant determinants for a variety of ICT variables at the state and provincial levels in the US and India [25].

College education was shown to be significant in studies of ICT and internet variables at the county level in the US [32–34]. It has been influential in the Chinese provincial digital divide literature as well [3,13,14,18].

Innovation. Innovation has infrequently been incorporated in the prior digital divide literature and rarely for China. For Japanese prefectures, innovation was a leading determinant of ICT use [42]. For China, it was identified as an important influence on early internet adoption [46]. Likewise, R&D for enterprises was a moderate influence on certain ICT factors for provinces in 2006 and 2009 [18]. However, in a study of China's city data in 2016–2017, innovation as measured based on R&D and patents granted was found not to be significant [3] A further justification for including innovation in the present model is the increasing prominence of innovation in the two most recent 5-year plans of the central Chinese government [47].

Knowledge Production. Production of knowledge in digital content is an essential aspect of ICT use, yet it is hard to directly measure consistently by province. In an earlier study, the number of printed books stood as a proxy variable for digital knowledge and was determined to be significant [18]. For the present study, we again use the proxy of the number of printed book copies, reasoning that published knowledge in books in a province reflects a knowledge capacity that also informs the province's digital knowledge content. A second indicator is the number of electronic publications, which is growing in China and adds a direct measure of production of provincial digital knowledge content.

4. Materials and Methods

The data and methodology are chosen to enable testing of the SATUM model and correspond to research questions 1–4. For this study, nearly all of the provincial data were drawn from the Chinese Statistical Yearbooks in 2021 and 2016. The Chinese Statistical Yearbook is an annual publication compiled by the National Bureau of Statistics of China and serves as an authoritative source of official statistical information [37]. It has been widely used for research [46,48,49]. Although some weaknesses have been identified, namely problems with time series industrial classification definitions over time, coverage of the industrial sector [50], difficulties in searching for terms in English [48], and questions about data collection of certain variables [45], we feel the present dataset has limited exposure to these problems based on the detailed notes in the yearbooks. We do not include GDP or industrial production data, which have been criticized [45,48,50].

The variables are either indices or normalized by population, and all are from the years 2020 and 2015, except for unemployment, which was averaged for the year and the two preceding ones in order to even out year-to-year fluctuations. The data were checked for accuracy and entered into SPSS (v26.0) for statistical analysis and ArcGIS Pro (v3.2) for GIS mapping.

First, the Moran's I diagnostic for each of the dependent variables was calculated to specify the amount of spatial agglomeration or dis-agglomeration. Moran's I varies from 1.0 (nearby provinces are similar to each other) to −1.0 (nearby provinces are different from each other). A value of zero indicates that the digital divide variables are independently and randomly distributed in space. Moran's I was also calculated for the regression residuals to assess the independence and randomness of residuals. A significant Moran's I value for residuals indicates that the regression findings must be viewed with caution, discarded, or re-run with new variables that account for the bias.

Besides descriptive statistics, ordinary least squares (OLS) regression was employed to test proposed associations between the 16 dependent variables and 14 independent variables. OLS regression was chosen instead of alternative approaches, such as structural equation modeling [49,51–53] or path analysis [54] for the following reasons. For SEM, there are modeling assumptions that need to be met, including a cause-and-effect relationship such that cause precedes effect in time, linearity, and other assumptions [52]. The main reason is that the present sample size of 31 is too small to apply structural equation modeling. Studies have shown that SEM sample sizes commonly range from 200 to 400

but are certainly more than 50 [49,52,55]. Likewise, another type of multi-factor modeling, path analysis, requires sample sizes that are too large for the present sample size of 31. Kline [54] indicates a minimum sample size for path modeling of 10 times the number of parameters. The present study has 4 or more parameters, implying a minimum sample size of 40 or more.

OLS regression was conducted in a stepwise manner, allowing variables to enter with significance levels equal to or less than 0.10. The Variance Inflation Factor (VIF) was computed for each independent variable as a test of multicollinearity, with a cutoff of 5.0 [56]. Each of the 14 OLS regressions had a VIF lower than 5.0, indicating that multicollinearity problems were not present. All OLS regression models of this study were computed using IBM SPSS software, version 29, which produced diagnostics, such as the VIF values of each OLS model. OLS regression diagnostics of Joint Wald, Koenker, and Jarque–Bera were calculated to ensure that the assumptions of OLS regression are met (see discussion in [25], (pp. 9–11)). These diagnostics were obtained from the spatial statistics module of the ArcGIS Pro software, which is often used for mapping and geospatial analysis.

Assessing the change of a variable in two or multiple points in time and the factors that influence the change has been an interest of many social science studies [57]. Longitudinal changes have been particularly explored with demographic and socioeconomic variables. For example, Wilkinson and Pickett [58] discuss the longitudinal effects of income inequality on population health, highlighting how socioeconomic variables help understand changes in health outcomes over time. Furthermore, temporal changes are not always uniform across different spatial scales [59]. Using the Getis–Ord's Gi statistic and correlation analysis, Casali et al. [60] examine the co-evolution of urban characteristics, including population, income, housing, and urban infrastructure. In this study, we compare ICT use in China from 2015 to 2020 and investigate the factors influencing these changes. To this end, a longitudinal regression model was applied.

In a regression model, using the change score (Y_2-Y_1) as the dependent variable is an intuitive option. However, given documented substantial correlation between (Y_2-Y_1) and Y_1 and unreliability of change scores [61], regressing raw change scores on the explanatory variables is problematic. Alternative models have been proposed, including the regressor method, the change score method, and the residual score method [62]. Dalecki and Willits [63] further tested the three regression models with an empirical case study. The change score method, which is mathematically equivalent to the other two methods, is adopted in the present study, where the dependent variable is the difference in dependent variables from 2015 to 2020, $Y_{2020}-Y_{2015}$, while the independent variable set includes the dependent variable for 2015, Y_{2015}, as well as other independent variables.

To categorize the spatial distribution of ICT use in China, we grouped the Chinese provinces into a pre-specified number of nonoverlapping clusters so that each of the clusters is internally as homogenous as possible in terms of the four groups of dependent variables. For this reason, k-means cluster analysis was conducted.

K-means cluster analysis is an unsupervised machine learning technique used for grouping similar data points into clusters based on their similarities for a set of dependent variables. Machine-learning-based methods are being increasingly used in urban spatial planning [60]. In k-means clustering, the data points are grouped into a pre-specified number of clusters (K) by minimizing the sum of squared distances (variance) between each data point and its assigned cluster center. The choice of the k value can be subjective due to its exploratory nature. To determine the optimal k value, the elbow graph, which depicts the change of within-cluster variance with k, is often applied. In this study, ArcGIS Pro calculates the Euclidean distance with the standardized scores of the original variables and adopts the Calinski–Harabasz pseudo F-statistic, a ratio reflecting between-cluster variance to with-in cluster variance [64]. Because the F-statistic for this analysis yields an optimal k value that equals N (the number of data points), this is not a practical solution. Therefore, we considered a range of plausible k values from 3 to 6 and compared the

visual outputs. The final choice was made with the consideration of spatial contiguity and visual interpretation.

5. Descriptive and Regression Findings

Descriptive statistics, computed for the variables, include the mean, median, standard deviation, and coefficient of variation (CV). In the dependent-variable infrastructure group, the line length of optical cable, mobile phone base stations, and capacity of mobile phone exchange are low in CV, a nationwide evenness that we argue stems from governmental policies and planning for installing basic internet and mobile network structure for potential access across the nation (Table 1). The use group also has low variability, which can be explained by usage being largely individual, mobile, and inexpensive. In contrast, the CVs are high for e-commerce and very high for IT services, averaging, respectively, 101.6 and 206.1. Clearly, there are vast provincial differences in purposeful uses across the nation, which we argue represent substantial digital inequality arguably derived from their earlier stage of geographical diffusion. The percentage of enterprises with e-commerce transactions is an exception, with low CV, which suggests enterprises throughout the country have readily adopted and use e-commerce, evening out provincial differences. However, the intensity of e-commerce adoption still varies greatly.

Table 1. Descriptive statistics for dependent variables, 2020.

Variable Set	Definition	N	Minimum	Maximum	Mean	Std. Deviation	Coeff. of Variation
Infrastructure	Length of optical cable lines (km) per cap.	31	0.02	0.07	0.04	0.01	30.81
	Mobile phone base stations (10,000), per cap.	31	0.05	0.15	0.07	0.02	28.04
	Capacity of mobile phone exchange (10,000 subscribers), per cap.	31	1.28	7.73	2.25	1.19	52.97
	Number of domain names (10,000) per cap.	31	0.0044	0.17	0.03	0.03	101.14
Use	Mobile internet subscribers per cap.	31	0.77	1.49	0.97	0.15	15.36
	Broadband subscribers per cap.	31	0.26	0.46	0.34	0.05	14.70
	Flow of mobile internet (10,000 GB), per cap.	31	0.01	0.02	0.01	0.003	22.02
	Penetration rate of mobile phones per 100 persons	31	88.24	178.43	115.10	18.63	16.18
ICT Services	Income from IT services (100k yuan) per 100,000 pop.	30	0.0004	0.50	0.05	0.10	214.75
	Income of related software businesses (100k yuan) per cap.	31	0.0006	0.72	0.06	0.14	212.31
	Income from Information Safety (100k yuan), per cap.	30	0.01	0.06	0.02	0.01	74.39
E-commerce	Sales through E-commerce (1 million yuan) per cap.	31	0.0021	0.12	0.02	0.03	167.64
	Purchases through E-commerce (1 million yuan) per cap.	31	0.0034	0.70	0.09	0.15	175.41
	Percent of enterprises with E-commerce transactions	31	5.50	22.80	10.26	3.43	33.46
	Websites per 100 enterprises	31	24.00	59.00	45.19	8.54	18.90
	Computers used at the end of period per cap.	31	1.12	22.99	4.23	4.77	112.79

The descriptive findings for the independent variables (Table 2) reflect China's ascent as the second largest global economy but also reveal provincial variation, especially in export of goods, innovation, and knowledge production. China's advances are evident in its provincial averages, with an urban population percentage of 63.8 percent, disposable household income/capita of CNY 32,086 (USD 4650 dollars in 2020), and college education of 9.9 per 100 people, age 15+. The high CV for exports of goods stems partly from greater international and manufacturing potential for exports in the coastal or near-coastal provinces compared to those inland. The higher provincial CVs for innovation are partly

related to the uneven geographic distribution of government research centers and research universities. For knowledge production, there is a moderate CV for printed books but a very high CV for electronic publication volume. We argue that book production as a mature process is fairly evenly spread across the nation, but that e-publication, at a newer stage of use, is more varied provincially.

Table 2. Descriptive statistics for independent variables, 2020.

Variable Set	Definition	N	Minimum	Maximum	Mean	Std. Deviation	Coeff. of Variation
Demographic	Sex ratio	31	98.73	123.17	104.79	5.17	4.94
	Urban population	31	35.73	89.30	63.73	11.06	17.35
	Child dependency ratio	31	13.25	37.17	25.79	6.91	26.80
	Han population, per cap	31	0.12	1.00	0.85	0.21	24.47
Economic	Disposable household income, in yuan, per cap.	31	20,335.10	72,232.40	32,086.38	12,661.02	39.46
	Unemployment of registered persons in urban areas, per cap. (2018–2020 average)	31	2.30	3.20	3.03	0.16	5.21
	Export of goods by producer location, per cap.	31	0.0021	0.47	0.10	0.13	132.67
Education	College education, age 15+, per cap	31	0.07	0.15	0.10	0.02	19.29
	Full-time teachers, per cap	31	0.74	3.39	1.37	0.52	37.66
Innovation	Expenditure for science and technology, per cap.	31	0.01	0.19	0.04	0.04	99.68
	Expenditure for R&D, per cap.	31	0.02	2.81	0.89	0.74	83.38
	Patent applications, per cap.	31	0.03	2.43	0.69	0.67	97.57
Knowledge	Printed book copies, per cap.	31	0.17	2.01	0.59	0.36	60.73
Production	Electronic publication volume, per cap.	31	0.00	0.53	0.04	0.10	253.76

Overall, computation and subsequent analysis of the coefficients of variation for each dependent variable across the groupings of infrastructure, use, ICT services, and e-commerce provide descriptive insight into the extent of the digital divide among Chinese provinces. It is clear that the extent of disparity is higher for ICT services and e-commerce compared to the indicators of ICT infrastructure and use. This is further reinforced when the characteristics of the k-means clusters of these groupings of dependent variables are discussed later in Section 7.

An example map in Figure 2 presents the geographic variation for broadband subscriptions per capita. Broadband subscription rates are very high in the eastern coastal provinces and in Beijing and Tianjin, moderately high in the central and far northwestern provinces, and low in the southwestern and northeastern provinces. The spatial pattern somewhat resembles the spatial distribution of the ICT development index in a provincial 2010 study of China [19], which suggests that some spatial patterns remain consistent longitudinally, even with the large growth over the last decade in ICT capacities. The global Moran's I for this variable is also shown to be significant, indicating the spatial agglomeration of provinces that have similar broadband subscription rates.

The OLS regression findings reveal for the full set of 14 independent variables that expenditure for science and technology is the dominant determinant, which is significant and positive for eight of the dependent variables (Table 3). An implication is that IT use might be an outcome of this science/technology government expenditure across the country. As examined in the discussion section, China's 13th five-year plans for the period of 2016–2020 emphasized the development of scientific innovation, ICT, and the internet [65–67]. The present significant innovation findings are consistent with results for Japanese prefectures of associations of patents with broadband subscribers, mobile phone subscribers, and social media use in 2010 [42].
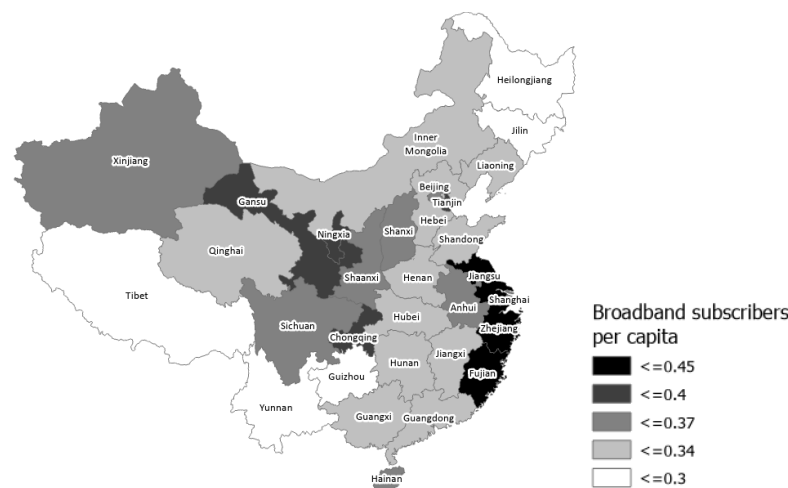
Figure 2. Broadband subscribers per capita.

The second leading factor is disposable household income, which is supported by antecedent China studies [1,3] and by the worldwide digital divide literature outside of China [34,42,68].

The third leading inverse relationship is an association between the percentage of the Han population and infrastructure and ICT use, implying that a higher percentage of non-Han minorities is associated with an increase in ICT, including, in particular, optical cable lines, mobile phone base stations, the capacity of mobile phone exchange, mobile phone internet flow, and the number of computers. Because most of the high-minority provinces are in the west and north of the country, the strong associations between the non-Han percentage of the population and infrastructure can be explained by the need for more extensive infrastructure to cover the larger land areas and lower population densities. The inverse association of the Han population with the use factors relates to leapfrogging catch-up in use by minority populations once income and scientific–technical expenditures are controlled for.

Trends towards minority catch-up are evident in studies in US states and counties; in particular, usage by ethnic minorities and White people converged during the first two decades of the 21st century [41]. Furthermore, the percentage of use of internet apps by Black and Hispanic people has converged to become comparable to use by White people in 2019 [69]. For Chinese research, a deeper investigation is called for to understand the spatial disparities of digital usage by non-Han minorities compared to the Han population.

To better investigate determinants other than the dominant one, a second set of regressions was performed, removing science and technology expenditures. In the second-round findings, per capita income and expenditure in R&D are the two dominant independent variables, which is supported by antecedent Chinese studies [3,14] and by the worldwide digital divide literature outside of China [34,42,69] (Table 4). R&D, which differs from science and technology expenditures due to its presence in private- and state-owned enterprises, has the arguable explanation that it influences higher ICT usage by attracting a digitally talented workforce that exerts multiplicative effects by sharing their technology-based social capital with citizenry in the province. Furthermore, we speculate that provincial R&D may have positively assisted in improving software and infrastructure to help individuals' experiences with technologies during the COVID-19 period. Full-time teachers are important for IT services variables, which suggests that better education relates to higher income from IT services. The percentage of the Han population is again apparent as a secondary inverse influence.

The third set of regressions removed the dominant variables in the first and second rounds. Findings for the third set, shown in Table 5, indicate the dominant variable of full-time teachers, present for 10 of the 16 variables. For 9 out of the 10 variables, the association

with full-time teachers is positive at the 0.05 level or lower. The influence of teachers corresponds to an earlier Chinese study in which the secondary school enrollment ratio was an important provincial factor for ICT use [3]. Education's impact on digital divides is supported by studies in US counties [33,34,43] and in other nations, such as Japan [70] and the UK [44]. Only slightly behind education is the Han inverse effect, present for eight variables, including three in the infrastructure group and all of the variables in the IT service group, reinforcing the explanations of non-Han minority catch-up in digital usage. Finally, urban population is associated with mobile internet subscribers, the penetration rate of mobile phones, and the flow of mobile internet. The percentage of urbanization has been important in a variety of prior studies [25]. In China, the urbanization rate was a key determinant of the provincial digital divide between 2001 and 2014 [3]. Although it was somewhat significant in stages 1 and 2, in stage 3, it becomes more significant as an overall study finding.

It is important to note that despite standardized regression coefficients changing from the first OLS set to the second and then from the second OLS set to the third set, the Variance Inflation Factor (VIF) never exceeded the 5.0 threshold, indicating that multicollinearity was not a problem in any of the OLS models. In addition, there were very few instances of important predictor variables not being statistically significant in their association with the dependent variables. This reinforces that multicollinearity was not a problem.

Table 3. Regression findings for 16 dependent factors and 14 independent variables, Chinese provinces, 2020.

Independent Variable Definition	All Length of Optical Cable Lines (km), Per capita	Base Stations of Mobile Phones (10,000), Per capita	Capacity of Mobile Phone Exchange (100,000) Subscribers	Number of Domain Names Per capita	Mobile Internet Subscribers Per capita	All Broadband Subscribers of Internet, Per 1000 Persons	Flow Accessed to Mobile Internet (10,000 GB) Per capita	Penetration Rate of Mobile Phones (Set/100 Persons)	Income from IT Services (100,000 yuan), Per capita	Gross Income from Related Software Business (100,000 yuan), Per capita	Income from Information Safety (100,000 yuan) LN Per 100,000 Pop	Sales through E-Commerce (1 million yuan), Per capita	Purchases through E-Commerce (1 million yuan), Per capita	With E-Commerce Transactions Enterprises, Percent of Enterprises	Websites Per 100 Enterprises (Unit)	Computers Used Per 100 Persons (Unit)
Sex Ratio							0.28 *		−0.11 *					−0.35 *	−0.29	
Urban Pop 2020 (%)					0.42 **			0.40 **								
Children Dependency Ratio	−0.43 **															
Han Percent of Population		−0.60 ***	−0.69 ***		−0.18 **			−0.68 ***								−0.78 ***
Per Capita Disposible Income of Household (yuan)		0.60 ***	0.44 **		0.92 ***			0.99 ***								
Exports of goods by location of producer (in 100k RMB), per capita																
College Subtotal, per pop. Age 15+ Full-Time Teachers, per capita		0.29							0.48 ***	0.34 ***	0.61 ***			−0.36		
Expenditure for Science and Technology, per capita	−0.61 **		0.90 ***				0.41 **		0.79 ***	0.76 ***		1.12 ***	1.12 ***	1.27 ***	0.35	0.96 ***
Expenditure on R&D (10,000 yuan), per 1000 persons					−0.54 ***	0.63 ***		−0.57 ***							0.44 *	
Number of Patent Applications, per 1000 persons	0.35											−0.27 ***	−0.29 ***			
Unemployment Avg. Rate of Persons in Urban Areas, 2018–2020																
Printed Book Copies (100 million copies), per capita		−0.36 *							−0.22 **							
Volume of Electronic Publications, per capita				−0.37 **							−0.20 *			−0.57 **		
Adjusted R^2	0.541 ***	0.658 ***	0.609 ***	0.653 ***	0.839 ***	0.380 ***	0.534 ***	0.898 ***	0.933 ***	0.878 ***	0.391 ***	0.918 ***	0.904 ***	0.555 ***	0.408 ***	0.913 ***
Spatial Autocorrelation (Moran's I)																
of Dependent variable	0.32 **	0.16			0.20 *	0.35 ***	0.18 *	0.22 *	0.20 **	0.17 **	0.20 **	0.14	0.01	0.10	0.16	0.22 **
of Regression Residual	0.01	−0.140			−0.11	−0.002	−0.14	0.12	−0.050	−0.06	−0.19	−0.04	−0.04	0.16	−0.13	0.09
Diagnostics																
Joint Wald Statistic	138.94 ***	139.83 ***			442.49 ***	15.90 ***	37.50 ***	607.24 ***	568.03 ***	167.55 ***	108.12 ***	407.53 ***	333.90 ***	198.00 ***	50.27 **	183.77 ***
Koenker Statistic	2.69	11.418 *			3.31	0.96	4.72	2.57	4.49	19.88 ***	11.88 **	0.53	0.95	9.01	0.82	2.50
Jarque-Bera Statistic	1.40	1.60			0.33	1.03	2.17	3.66	1.97	1.38	8.492 *	1.16	1.66	1.01	2.22	1.70
Sample Size	31	31	31	31	31	31	31	31	30	31	30	31	31	31	31	31

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4. Regression findings for 16 dependent factors and 13 independent variables, Chinese provinces, 2020.

Independent Variable Definition	All Length of Optical Cable Lines (km), Per capita	Base Stations of Mobile Phones (10,000), Per capita	Capacity of Mobile Phone Exchange (100,000) Sub-scribers	Number of Domain Names Per capita	Mobile Internet Sub-scribers Per capita	All Broad-band Sub-scribers of Internet (10,000 Sub-scribers), Per capita	Flow Accessed to Mobile Internet (10,000 GB) Per capita	Penetration Rate of Mobile Phones (Set/100 Persons)	Income from IT Services (100,000 yuan), Per capita	Gross Income from Related Software Business (100,000 yuan), Per capita	Income from Information Safety (100,000 yuan) LN Per 100,000 Pop	Sales through E-Commerce (1 million yuan), Per capita	Purchases through E-Commerce (1 million yuan), Per capita	With E-Commerce Transactions Enterprises, percent of Enterprises	Websites Per 100 Enterprises (Unit)	Computers Used Per 100 Persons (Unit)
Sex Ratio							0.27								−0.31	
Urban Pop 2020 (%)					0.42 **			0.37 **								
Children Dependency Ratio													0.15			
Percent of Populaton HAN	−0.60 ***	−0.94 ***	−0.83 ***				−0.71 ***				−0.60 ***					
Per Capita Disposible Income of Household (yuan)		0.60 ***	0.44 ***	0.29	0.92 ***		.41 **	0.98 ***	0.51 ***	1.40 ***		1.27 ***	1.18 ***	0.86 ***		1.17 ***
Exports of goods by location of producer (in 100k RMB), per capita										−0.43 ***		−0.47 ***				−0.29 ***
College Subtotal, per pop. Age 15+																
Full-Time Teachers, per capita		0.29		0.46 **					0.52 ***		0.61 ***		0.22 *			
Expenditure on R&D (10,000 yuan), per 1000 persons					−0.54 ***	0.63 ***		−0.52 ***					−0.43 ***		0.63 ***	
Number of Patent Applications, per 1000 persons																
Unemployment Avg. Rate of Persons in Urban Areas, 2018–2020			−0.34 **												0.26	
Printed Book Copies (100 million copies), per capita		−0.36 *														
Volume of Electronic Publications, per capita											−0.33 **		−0.52 *			
Adjusted R^2	0.335 ***	0.658 ***	0.609 ***	0.652 ***	0.839 ***	0.380 ***	0.516 ***	0.906 ***	0.866 ***	0.878 ***	0.391 ***	0.912 ***	0.903 ***	0.333 ***	0.404 ***	0.928 ***
Spatial Autocorrelation (Moran's I)																
of Dependent variable	0.32 **	0.16			0.20 *	0.35 ***	0.18 *		0.20 **	0.17 **	0.20 **	0.14	0.01	0.10	0.16	0.22 **
of Regression Residual	0.16	−0.14			−0.11	0.00	−0.13		0.14	−0.01	−0.19	−0.05	−0.16	0.296 **	−0.11	−0.03
Diagnostics																
Joint Wald Statistic	31.61 ***	139.83 ***			442.49 ***	15.91 ***	39.94 ***		49.06 ***	140.34 ***	108.12 ***	221.12 ***	362.17 ***	16.80 ***	31.34 ***	242.21 ***
Koenker Statistic	0.35	11.42 *			3.31	0.96	8.34 *		18.50 ***	7.95 *	11.88 **	14.89 ***	12.28 *	4.44	1.11	6.30 *
Jarque-Bera Statistic	0.67	1.60			0.33	1.03	1.96		7.847 *	0.13	8.492 *	3.71	17.623 ***	1.91	0.35	0.79
Sample Size	31	31	31	31	31	31	31	31	30	31	30	31	31	31	31	31

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5. Regression findings for 16 dependent factors and 12 independent variables, Chinese provinces, 2020.

Independent Variable Definition	All Length of Optical Cable Lines (km), Per capita	Base Stations of Mobile Phones (10,000), Per capita	Capacity of Mobile Phone Exchange (100,000) Sub-scribers	Number of Domain Names Per capita	Mobile Internet Sub-scribers Per capita	All Broad-band Sub-scribers of Internet (10,000 Sub-scribers), Per capita	Flow Ac-cessed to Mobile Internet (10,000 GB) Per capita	Penetration Rate of Mobile Phones (Set/100 Per-sons)	Income from IT Ser-vices (100,000 yuan), Per capita	Gross Income from Related Software Business (100,000 yuan), Per capita	Income from Infor-mation Safety (100,000 yuan) LN Per 100,000 Pop	Sales through E-Commerce (1 million yuan), Per capita	Purchases through E-Commerce (1 million yuan), Per capita	With E-Commerce Transac-tions Enter-prises, Percent of Enter-prises	Websites Per 100 Enter-prises (Unit)	Computers Used Per 100 Persons (Unit)
Sex Ratio											−0.16				−0.31	
Urban Pop 2020 (%)					0.81 ***		0.55 *	0.87 ***								0.43 **
Children Dependency Ratio																
Percent of Populaton Han	−0.60 ***	−0.95 ***	−0.69 ***		−0.32 *		−0.85 ***		−0.25 *	−034 **	−0.60 ***		−0.20			
Exports of goods by location of producer (in 100k RMB), per capita									0.34 ***							
College Subtotal, per pop. Age 15+																
Full-Time Teachers, per capita		0.45 **	0.43 **	0.64 ***	0.25				0.89 ***	0.88 ***	0.61 ***	0.67 ***	0.74 ***	0.38 *		0.32 *
Expenditure on R&D (10,000 yuan), per 1000 persons						0.66 ***		−0.27							0.63 ***	
Number of Patent Applications, per 1000 persons		0.31 *								0.44 ***						
Unemployment Avg. Rate of Persons in Urban Areas, 2018–2020				−0.38 **											0.26	
Printed Book Copies (100 million copies), per capita								0.319								0.28 *
Volume of Electronic Publications, per capita												0.44 ***	0.46 ***			
Adjusted R^2	0.335 ***	0.580 ***	0.583 ***	0.612 ***	0.710 ***	0.380 ***	0.454 ***	0.741 ***	0.809 ***	0.805 ***	0.391 ***	0.729 ***	0.731 ***	0.116 *	0.404 ***	0.723 ***
Spatial Autocorrelation (Moran's I)																
of Dependent variable	0.32 **	0.16			0.20 *	0.35 ***	0.18 *	0.22 *	0.20 **	0.17 **	0.20 **	0.14	0.01	0.10	0.16	0.22 **
of Residual	0.16	−0.12			−0.20	0.00	−0.09	−0.11	−0.19	−0.10	−0.19	−0.08	−0.17	0.21	−0.11	0.04
Diagnostics																
Joint Wald Statistic	31.61 ***	36.28 ***			46.31 ***	15.90 ***	38.89 ***	59.85 ***	68.74 ***	41.27 ***	108.12 ***	70.56 ***	87.98 ***	1.44	31.34 ***	49.23 ***
Koenker Statistic	0.35	9.96 *			14.62 **	0.96	1.32	7.54	23.49 **	24.11 ***	11.88 **	18.69 ***	12.53 **	16.49 ***	1.11	14.96 **
Jarque-Bera Statistic	0.67	0.41			1.10	1.03	0.61	0.27	7.165 *	0.25	8.492 *	2.89	3.73	0.20	0.35	2.18
Sample Size	31	31	31	31	31	31	31	31	30	31	30	31	31	31	31	31

$* p < 0.05$, $** p < 0.01$, $*** p < 0.001$.

6. Longitudinal Comparison of ICT Factors between 2015 and 2020

A comparison was performed between 2015 ICT/internet levels and the same variable set in 2020, both drawing from the China Yearbook series. A paired sample *t* test was conducted to compare the overall ICT change of Chinese provinces during the designated five years. As seen in Table 6, all of the dependent variables had a significant increase, except for the number of websites per 100 enterprises, indicating that most dimensions of ICT had kept developing in China from 2015 to 2020. The growth, however, may not be even across the Chinese provinces. Accordingly, a regression analysis of the change in ICT factors was carried out, and the results are summarized in Table 7.

Table 6. Paired *t* test of changes in ICT and internet determinants, 2015–2020.

	Paired Differences		t	df
	Mean	**Std Deviation**		
All lengths of optical cable lines (km), per capita	0.02	0.01	12.40	30
Base stations of mobile phones (10,000), per capita	0.04	0.01	14.44	30
Mobile internet subscribers per capita	0.26	0.10	13.90	30
All broadband subscribers to internet (10,000 subscribers), per capita	0.16	0.04	22.97	30
Flow accessed to mobile internet (10,000 GB) per capita	0.01	0.003	25.30	30
Penetration rate of mobile phones (set/100 persons)	21.41	11.54	10.32	30
Income from related software business (CNY 100,000), per capita	0.03	0.09	2.09	29 [1]
Income from IT services (CNY 100,000), per capita	0.03	0.07	2.19	29 [1]
Sales through e-commerce (CNY 1 million), per capita	0.01	0.01	2.85	30
Purchases through e-commerce (CNY 1 million), per capita	0.004	0.01	3.27	30
Enterprise with e-commerce transactions (%)	1.17	2.15	3.03	30
Computers used per 100 persons in enterprise	10.55	4.10	14.33	30
Websites per 100 enterprises (unit)	−9.16	6.42	−7.95	30

[1] Tibet had missing values for ICT service variables. Only these two tests are significant at $p < 0.05$, while all others are significant at $p < 0.01$.

As expected, most of the 2015 ICT dependent variables significantly influenced the changes during the five-year period. Provinces with higher penetration rates for mobile internet, broadband subscribers, and mobile phones experienced smaller growth in the five years subsequent to 2015. This suggests a saturation of the market for these ICT uses. Similarly, the inequality of ICT infrastructure lessened, as evidenced by the negative coefficient of capacity of mobile phone exchange, domain names, and websites of enterprises. On the other hand, gaps in income from ICT services, intensity of mobile internet usage, and e-commerce continued to increase among Chinese provinces.

In addition to the 2015 ICT dependent variables, the most dominant independent variable is college degree or above, which positively affects the growth in digital infrastructure, IT services, and ICT use, suggesting the sustained development in ICT is highly dependent on the education attainment of the population.

The second influential indicator is the unemployment rate. The higher the unemployment rate, the slower the development in digital infrastructure, IT services, and e-commerce, which is in line with our expectation. However, the unemployment rate had a positive relationship with the mobile phone adoption rate. This might be because although fewer unemployed people would obtain company-subsidized mobile phones for daily use, personal uses, such as job searching, are still in demand.

Table 7. Regression findings for changes in ICT and internet determinants, 2015–2020.

	All Length of Optical Cable Lines (km), Per capital	Base Stations of Mobile Phones (100k), Per capita	Capacity of Mobile Phone Ex-changes, Per capital	Number of Domain Names Per capita	Income from Related Software Business (100k yuan), Per capita	Income from IT Services (100k yuan), Per capita	Mobile Internet Sub-scribers Per capita	Flow Accessed to Mobile Internet (10k GB) Per capita	All Broadband Sub-scribers of Internet, Per capita	Penetration Rate of Mobile Phones (Set/100 Persons)	Sales through E-Commerce (1 million yuan), Per capita	Purchases through E-Commerce (1 million yuan), Per capita	Percent of Enterprises with E-Commerce Percent of Enterprises	Computers Used Per 100 Persons (Unit) in the Enterprise	Websites Per 100 Enterprises (Unit)
Sex Ratio													−0.50 *		
Urban Pop (%)															
Children Dependency Ratio						0.23 *								−0.70 ***	
Disposable Income of Household (yuan), per capita															
Unemployment Average Rate in Urban Areas			−0.42 **		−0.26 **					0.41 ***	−0.17 *				
Exports of goods by location of producer, per capita															
College Degree or above, per pop. Age 15+		0.70 ***			0.80 ***	0.57 **				0.74 ***	0.36 **				
Full-Time Teachers, per 1000 pop.															
Expenditure for Sci & Tech, per capita (10K Yuan)				−0.058 **			0.69 ***								
Expenditure on R&D (1K yuan), per capita	−0.49 **														
Patent Applications (piece), per 1K persons															
Printed Book Copies (10 copies), per capita			0.47 **												
Volume of Electronic Publications, per capita												0.43 ***			
Y_15			−0.503 **	−0.360 *		0.547 **	−1.292 ***	0.509 **	−0.497 **	−0.979 **	0.566 ***	0.713 ***			−0.593 **
Adjusted R-square	0.209	0.473	0.535	0.799	0.864	0.915	0.717	0.369	0.213	0.747	0.888	0.944	0.129	0.467	0.322
F statistic	7.094 **	21.637 ***	9.813 ***	35.505 ***	61.365 ***	88.722 ***	18.992 ***	10.132 **	9.339 **	30.581 ***	80.230 ***	195.08 ***	4.305 *	21.318 ***	17.776 ***
Sample Size	31	31	31	31	30	30	31	31	31	31	31	31	31	31	31

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Compared to the prior discussion, the impact of innovation and knowledge production had mixed impacts on ICT changes. Expenditure on science and technology boosted mobile internet penetration and knowledge production in books; electronic publications facilitated the growth in ICT infrastructure and e-commerce. On the other hand, however, provinces that have more investment in R&D and science and technology seem to have a smaller increase in the ICT infrastructure variables of length of optical cables and capacity of mobile phone exchanges. The latter findings might represent saturation of ICT infrastructure, leading to lessening of ICT infrastructure's relationship with R&D and science and technology investment. This shows that innovation did not play a vital role in ICT infrastructure development compared to IT services and ICT use. Lastly, the inverse effect of sex ratio on e-commerce suggests that the digital divide of involvement in e-commerce has been reduced between women and men.

In summary, China experienced overall growth in ICT from 2015 to 2020, but the growth was uneven among provinces. More educated areas tended to have higher growth in base stations of mobile phones, mobile phone penetration rates, IT services, and sales through e-commerce, indicating an increased digital divide gap for these applications in the country, while science and technology expenditure and unemployment rate had mixed effects.

7. Cluster Analysis

To determine spatial patterns of China's digital divide, k-means cluster analysis was conducted separately for each of the four groups of dependent variables: digital infrastructure, ICT use, IT service, and e-commerce. This allows for recognition of each of the four stages of digital divide, with access to technology represented by the infrastructure group, use of technology represented by the use group, and purposeful use represented by two areas of purposeful use, the area of ICT services and e-commerce. Clusters were mapped, and cluster centers are summarized in tables.

7.1. Infrastructure

The ICT infrastructure cluster map (Figure 3) contains five distinct clusters, with the underlying cluster centers for the variables in Table 8. Beijing, the sole member of cluster 1, ranks highest in domain names, mobile phone base stations, and capacity of mobile phone exchange, as expected given its status as the political and cultural hub of the country. Shanghai, the sole member of cluster 2, is ranked highest for its optical cable coverage, reflecting its leadership in economic development and technological investment. This is evident from the cluster center values of clusters 1 and 2 in Table 8.

Figure 3. Clusters of infrastructure, China, 2020.

Table 8. K-means clusters of digital infrastructure, China 2020.

	Cluster Number				
	1	**2**	**3**	**4**	**5**
Number of domain names per capita	0.17	0.06	0.02	0.03	0.004
Mobile phone base stations (100,000) per capita	0.11	0.07	0.08	0.06	0.15
Capacity of mobile phone exchange per capita	4.27	2.60	2.06	1.85	7.73
Optical cable lines (km) per area (km^2)	28.83	113.42	12.43	11.16	0.21
Number of provinces	1	1	16	12	1

Compared to the first two clusters, clusters 3 and 4 (in blue and purple, respectively, in Figure 3) have much lower mean centers across all four infrastructure variables. Between these two clusters, the 12 provinces in cluster 4, located in the northeast and north–south flank from Beijing to Guangdong, has even lower infrastructure per capita, which could just be explained by their high population density. Cluster 3 (in blue) includes the eastern coastal states, the north, Yunnan Guizhou, and Chongqing in the south. Most of these states in the north and south have lower overall ICT use and therefore less need for infrastructure, while the low levels of per capita infrastructure in the eastern coastal states are unexplained. Lastly, Tibet (the sole member of cluster 5) has its unique pattern. It has a strikingly low level of domain names per capita, showing the slow penetration of internet use in that province. However, it has the largest capacity for mobile phone exchange per capita, which indicates that the physical infrastructure for communications has been much improved in the remote area.

7.2. ICT and Internet Use

For the cluster analysis of the use group of variables, cluster 1 stands out, with Beijing and Shanghai (shown in red in Figure 4) showing the highest ICT use in terms of mobile phone subscribers and phone penetration rate, as evident from the cluster center values in Table 9. This is expected given their status as the technological centers of the country. Following these two megacities, in cluster 2 (in yellow) are eastern coastal provinces and the central provinces of Chongqing and Ningxia. This cluster is characterized by the highest ranking of all broadband subscribers to the internet, but it is lower in mobile internet flow and phone penetration compared to the first cluster. The inclusion of Ningxia in this cluster is unexpected given its relatively low levels of urbanization and economic development.

Figure 4. Clusters of ICT and internet use, China, 2020.

Table 9. K-means clusters of ICT use, China 2020.

	Cluster Number				
	1	2	3	4	5
Broadband subscribers to internet per capita	0.36	0.42	0.33	0.33	0.28
Mobile internet flow (10,000 GB) per capita	0.02	0.01	0.01	0.01	0.02
Broadband subscribers to internet per capita	1.42	1.01	0.99	0.88	0.88
Phone penetration rate (set/100 persons)	198.98	136.47	128.45	117.02	116.32
Number of provinces	2	6	8	11	4

Clusters 3 and 4 (in blue and purple, respectively, in Figure 4), located in the nation's center and north and comprising two-thirds of the provinces, have relatively moderate–low ICT use compared to the first two clusters, reinforced by the middling cluster center values shown in Table 9. Cluster 5 (in green), consisting of Tibet, Qinghai, Yunnan, and Guangxi, has levels of mobile factors similar to clusters 3 and 4, which is not surprising given its average ICT infrastructure, shown in Figure 3. But, the lower adoption of broadband internet indicates that individuals in these regions may rely more on their own mobile data plan to access the internet while on the go, as there is a shortage of urban infrastructure and opportunities like restaurants and shops that provide free Wi-Fi hotspots.

7.3. IT Services

The four clusters for IT services are straightforward in explanation (Figure 5 and Table 10). Beijing dominates, in this case, on all three underlying variables: IT services, software-related services, and safety services. Tianjin and Shanghai in cluster 2 are moderate in income from software business and IT services but very low in income from information safety, which is unexpected.

Figure 5. Clusters of ICT services, China, 2020.

Although clusters 3 and 4 are both low in IT-related services compared to the first two clusters, cluster 3, including eastern coast provinces, Guangxi, and the provinces extending from Hubei to Chongqing and Sichuan, has higher income for IT services, which largely corresponds to the e-commerce pattern discussed below. This is reasonable, as IT-related services are an important factor for e-commerce.

Table 10. K-means clusters of IT services, China 2020.

	Cluster Number			
	1	**2**	**3**	**4**
Income from software business (CNY 100,000) per capita	0.72	0.22	0.07	0.01
Income from IT services (CNY 100,000) per capita	0.50	0.18	0.04	0.01
Income from information safety (CNY 100,000)	11.61	0.12	1.79	0.12
Number of provinces	1	2	9	18

7.4. E-Commerce

In terms of e-commerce, Beijing and Shanghai are again the leaders by far, and they are especially high in e-commerce sales and purchases per capita (Figure 6 and Table 11). This is explained by the megacities in those provinces being the predominant headquarter locations for e-commerce firms, as well as the location of high concentrations of computers per capita.

Figure 6. E-commerce clusters, China, 2020.

Table 11. K-means clusters of e-commerce, China 2020.

	Cluster Number			
	1	**2**	**3**	**4**
Computers used in enterprise per 100 persons	20.56	4.67	2.27	2.83
Number of websites of enterprises	57.50	52.75	44.67	36.44
Enterprises with e-commerce transactions	17.00	12.15	10.50	6.76
Sales through e-commerce (CNY 1 million)	0.12	0.02	0.01	0.01
Purchases through e-commerce (CNY 1 million)	0.63	0.08	0.03	0.05
Number of provinces	2	8	12	9

The other three clusters are similar in their levels of sales/purchases of e-commerce and computers. What differentiates these three clusters are the low levels of percentage of enterprises using e-commerce, reflecting that the northern tier plus Guangxi (cluster 4) has relatively little e-business (blue on the map), whereas the central east to west flank of the nation has higher levels of these two variables (clusters 2 and 3) yet somewhat less than the leaders in cluster 1. The subtle geographical differences between the two intertwined geographies (blue and yellow) cannot be explained and call for further research.

Overall, the analysis of the clusters for the four groups of variables indicates that digital divides in China persist geographically but with distinctive group-related patterns of geographies, especially in the nation's center and east–center. The constant across all four variable groups is the dominance of Beijing and Shanghai, which may be considered high outliers.

8. Discussion and Implications

The provincial variation across the infrastructure and use groups is relatively small, whereas provincial inequality for IT services and e-commerce is large, as reflected by their elevated coefficients of variation. This finding relates to prior studies identifying the succession of stages of ICT access, use, and purposeful use [6–8]. In China, access and use have mostly achieved evenness among the provinces. The low variation in infrastructure implies that access is available throughout the provinces. Likewise, there is a tendency towards evenness of use variables across provinces. The high variation among provinces in purposeful uses (IT services and e-commerce) implies that these specialized uses are in earlier adoption and diffusion stage in China and hence represent a large digital divide that persists in the country.

Regarding the spatial patterning of these purposeful uses, we argue, based on the mapping results, that the enterprises providing IT services and e-commerce are concentrated in Beijing, Shanghai, and the eastern coastal provinces, with reduced presence in the western, northern, and south–central provinces. On the other hand, the more even provincial spatial distribution of infrastructure per capita is the outcome of the Chinese government's policy that has encouraged digital infrastructure throughout the country as well as the business strategies of provider enterprises that have realized market opportunities in previously "poorly wired" provinces. The Chinese government's 13th five-year plan also sought to even out the infrastructure throughout the nation. The use group of variables is mostly mobile, inexpensive, and supported by the widening of available internet infrastructure.

The regression findings regarding determinants of the internet and ICT point to which factors dominate but also which ones show little or no relationships with ICT. The dominant importance of science and technology expenditures is in concert with China's 13th 5-year plan in 2016, which emphasized innovation-driven development as a major pillar [47]. The latest Chinese national 5-year plan approved in 2021 put even more emphasis on scientific innovation, with seven technology targets, which include semiconductors, quantum computers, and artificial intelligence. We argue that the dominance of science and technology expenditures is partly the result of the national government's multi-year planning effort to build up domestic innovation and technology [61]. The second most important factor of household income is unsurprising, as income has been reported in prior studies and also because given that the year 2020 data come after the start of the COVID-19 pandemic, household income would encourage increased use of the internet and technology at home to offset work-related exposure to the virus. Teachers' secondary importance is consistent with education's prominence in multiple digital divide studies, both in China and external to it.

A surprising secondary determinant is the influence of the non-Han (minority) population. We reason that minorities are leapfrogging from their prior very low ICT and internet usage levels to approach or even exceed national averages, although the data do not include the quality and sophistication of ICT use. There would presumably be more catching up to do in specialized, highly skilled uses. An example is e-commerce variables, with which the non-Han proportion presently has almost no association.

The cluster analyses for the four groups show that the spatial distribution of ICT varies considerably in China. Comparing across the four groups of cluster analysis, a common feature is the leadership of the megacities of Beijing and Shanghai. It is well-known that those giant cities have the most presence of science and technology R&D, tech company headquarters, income, and education. For the other 29 provinces, the spatial patterns for IT service and e-commerce are similar in that they are concentrated in the eastern coastal

provinces and in an east–west flank of provinces extending from Anhui and Zhejiang westward across Hubei and Chongqing to Sichuan and into Tibet. This vast area of over 300 million people has heightened e-commerce and IT services compared to provinces in the north, northwest, and south, a pattern calling for further investigation. The pattern for ICT infrastructure emphasizes, besides the megacity provinces of Beijing and Shanghai, levels in the north, mid-south, and eastern coast of a third to a half of the megacities, which supports the earlier point about infrastructure being spread widely. Tibet, however, has very low infrastructure with the exception of having the nation's highest mobile phone exchange capacity per capita, due to, we argue, Tibet's mobile phone dependence in its vast and rugged landscape.

Besides the two megacities and Tianjin, which adjoins Beijing, the pattern for ICT use emphasizes Ningxia, Chongqing, and the east coast. Other than Ningxia, these areas are modern and encouraging for use, such that high per capita use in Ningxia, a province that is economically weak and sparsely populated, is unexplained. Overall, the cluster analysis patterns for four groups of variables portray considerable geographic differences among groups, which can inform the government on where the government could bring forward policy to even out or build specific technological capacities.

Based on the findings, this research supports that the national government should consider the following policies.

Continue to emphasize science and technology investment, especially in underserved provinces to reduce the gaps in IT services and e-commerce.

View household income levels as essential to stimulating provincial ICT and emphasize steps to increase household income in less prosperous provinces.

In educational planning for provinces, consider giving emphasis to the metric of number of teachers, as well as the training and quality of teachers. As ICT use gaps narrow, there remain specialized digital divides that will require new education and training with an abundance of appropriately trained teachers as a key factor.

Plan to go farther in elevating the technological training and capacities of the non-Han population in provinces. Although this study shows positive associations of minorities with ICT infrastructure, the metrics might be changed from subscriptions, uses, and e-commerce prevalence to raising the quality of ICT participation, ICT teaching/learning, and specialized ICT services and e-services.

Central government planners should take into account distinctive spatial "hot spots" and "cold spots" appropriate for each step in the progression ladder of ICT from access to use, specialized use, and outcomes. Knowing those hot and cold spots along the progression ladder might improve the geographic areas in which to focus specific planning projects.

Many of these recommendations for this investigation are in concert with the 13th and 14th 5-year plans of the Chinese government [47,65–67]. In the 13th 5-year plan, innovation in science and technology was a leading emphasis, which ranged from strengthening ICT and internet infrastructure to enhanced and widespread training and education, encouraging businesses to conduct R&D, emphasizing resources, and establishing organizations for cutting-edge ICT innovation [66]. The emphasis is national, while specific regions are occasionally stressed. Our findings do support that the dependent variables largely increased significantly during the period of 2015–2020, which nearly matches the plan's period of 2016–2020. One sub-sector that was mostly not statistically significant in growth was IT services (Table 7), a type of purposeful use not specifically mentioned in the plan but clearly of importance in succeeding with large-scale ICT growth. Our study also indicates that although ICT and internet uses did spread fairly evenly across the nation's provinces, the spread of ICT services and sales/purchases of e-commerce were uneven (Table 2). These purposeful use types are driven by government and businesses and may depend somewhat on broad Chinese economic shifts from export to internal consumption [71] in order to reach evenness throughout the provinces. Interestingly, the need for nationwide improvements in these two areas is addressed in the 14th 5-year plan for 2021–2025 [67], which emphasizes a

productive, flourishing service industry (Article X) and mentions promotion of e-commerce and related smart logistics under the topic of digital industrialization (Article XV, Section 2).

The theoretical implications of the study are three-fold. First, the SATUM model recognizes the geographical relationships between the dependent variables. For the OLS regression analysis, each dependent variable is examined for its positive or negative spatial agglomeration, i.e., for each dependent variable, are there distinctive spatial patterings of similarly valued provinces. The OLS regression residuals are examined in the same way [21]. Second, the SATUM model has been expanded in the present research to recognize four distinct groups of dependent ICT variables, rather than only for the full set of variables. Accordingly, in the present study, the focus is on the patterns of spatial agglomeration of each of four groups of dependent variables. The earlier approach of clustering a full set of dependent variables may obscure key underlying geographic patterns, while the present four separate groupings of purposeful uses reveal the distinctive patterns, and, for each group, clusters can be characterized based on the averages of the independent variables. This approach can yield deeper insights into pattern differences of variable groups, including, for example, for e-commerce variables, in a digital divide analysis. As far as we know, this is a novel theoretical approach for ICT research.

Third, adding to SATUM a new feature for longitudinal analysis of change provides an understanding of how the relative importance of correlates evolves over a prior time interval. This model's enlargement can be operationalized if there is a comparable set of variables at a prior point, which is increasingly available for nations and for sub-national geographic units, such as states and provinces. Regression change methodologies, such as the technique of regression change scores in dependent variables in this study, can help to explain the dynamics of change in regression correlates over time.

Limitations

There are limitations to this exploratory approach to studying digital inequality in China in 2020. First of all, because most provinces in China are large and populous geographic areas, uneven intra-provincial digital development is common. The analysis at the provincial level cannot reveal the hidden patterns at a granular scale. Another limitation is that the variable choice depends on the China Yearbooks, which do not collect some well-known variables of importance in the digital divide literature, such as social media factors, social capital, detailed ethnic groupings, telecom regulations, the rule of law, and freedom.

9. Conclusions

This study examines the state of provincial digital inequalities in China after the start of the COVID-19 pandemic. Novel aspects of this research include an expanded theoretical model, heretofore unreported COVID-19-period findings, longitudinal change analysis and findings, emphasis on purposeful uses, and comparative findings from cluster patterns, which are segmented into four groups of ICT variables.

K-means cluster analysis spatially reveals the dominance of the megacities of Beijing and Shanghai, followed by coastal regions and southwestern population centers that are marked by prominent levels of e-commerce activities. At the other end of the spectrum are provinces in western and central China that are relatively low in terms of ICT use, with the exception of modern inexpensive uses, such as mobile internet. Comparisons between the patterns of four ICT/internet purposeful use categories provide new insights.

OLS regression findings reveal that expenditure for science and technology is the key determinant of the digital inequities within China in 2020. Other important variables are per capita household income and expenditure in R&D, full-time teachers and minority (non-Han) population, and proportion of urbanization. Some of these correlates have been found to be associated with sub-national digital disparities in other nations, such as the United States, Japan, and India [25]. A longitudinal comparison of 2015–2020 reveals specific shifts in the dominant determinants, with the significant variables being science and

technology expenditure, unemployment rate, and college education, which partly reflect the advent of the COVID-19 pandemic in early 2020. Methodologically, the combination of supervised (OLS regressions) and unsupervised (k-means clustering) machine learning methods to study China's provincial digital divide is a research approach that could be helpful for other researchers in the future.

Overall, associations posited by the research questions and supported by an enlarged SATUM conceptual model are validated by the study's empirical findings. Despite the enormous progress made by China to bridge its provincial digital divide, this study reveals that gaps remain between Chinese provinces, pointing to the need for information systems and telecommunications policy development that focuses on closing these gaps and creating a mature, equitable digital society.

References

1. CEIC. *Exports: ICT Goods*; CEIC Data Ltd.: Singapore, 2023. Available online: https://www.ceicdata.com/en/indicator/exports-ict-goods (accessed on 1 June 2024).
2. Zhang, X. Investigation of e-commerce in China in a geographical perspective. *Growth Chang.* **2019**, *50*, 1062–1084. [CrossRef]
3. Song, Z.; Wang, C.; Bergmann, L. China's prefectural digital divide: Spatial analysis and multivariate determinants of ICT diffusion. *Int. J. Inf. Manag.* **2020**, *52*, 1–12. [CrossRef]
4. UN Human Rights Council. *The Promotion, Protection, and Enjoyment of Human Rights on the Internet*; Oral Revisions of 30 June, 32nd Session, Agenda Item 3; United Nations: New York, NY, USA, 2016.
5. van Dijk, J.A.G.M. A theory of the digital divide. In *The Digital Divide: The Internet and Social Inequality in International Perspective*; Ragnedda, M., Muschert, G.W., Eds.; Routledge: London, UK, 2013; pp. 29–51.
6. van Dijk, J. *The Digital Divide*; Polity Press: Cambridge, UK, 2020.
7. Warschauer, M.; Matuchniak, T. New Technology and Digital Worlds: Analyzing Evidence of Equity in Access, Use, and Outcomes. *Rev. Res. Educ.* **2010**, *34*, 179–225. [CrossRef]
8. Skaletsky, M.; Pick, J.B.; Sarkar, A.; Yates, D.J. Digital divides: Past, present, and future. In *The Routledge Companion to Management Information Systems*; Galliers, R.D., Stein, M.K., Eds.; Routledge: London, UK, 2017; pp. 416–443.
9. Scheerder, A.; van Duersen, A.; van Dijk, J. Determinants of internet skills, uses, and outcomes. A Systematic review of the second and third-level digital divide. *Telemat. Inform.* **2017**, *34*, 1607–1624. [CrossRef]
10. Loo, B.P.; Wang, B. Progress of e-development in China since 1998. *Telecommun. Policy* **2017**, *41*, 731–742. [CrossRef]
11. Zhang, M.; Sarker, S.; Sarker, S. Unpacking the Effect of IT Capability on the Performance of Export-focused SMEs: A Report from China. *Inf. Syst. J.* **2008**, *18*, 357–380. [CrossRef]
12. Fong, M.W.L. Digital divide between urban and rural regions in China. Electron. *J. Inf. Syst. Dev. Ctries.* **2009**, *36*, 1–31. [CrossRef]
13. Zhu, S.; Chen, J. The Digital Divide in Individual E-Commerce Utilization in China: Results from a National Survey. *Inf. Dev.* **2013**, *29*, 69–80. [CrossRef]
14. Song, Z. The geography of online shopping in China and its key drivers. *Environ. Plan. B Urban Anal. City Sci.* **2022**, *49*, 259–274. [CrossRef]
15. Wang, D.; Zhou, T.; Lan, F.; Wang, M. ECT and Socio-economic Development: Evidence from a Spatial Panel Data Analysis in China. *Telecommun. Policy* **2021**, *45*, 1–13. [CrossRef]
16. Zhang, X. Broadband and Economic Growth in China: An Empirical Study during the COVID-19 pandemic period. *Telemat. Inform.* **2021**, *58*, 2–9. [CrossRef] [PubMed]

17. Song, W. Development of the Internet and digital divide in China: A spatial analysis. *Intercult. Commun. Stud.* **2008**, *17*, 20–43.
18. Pick, J.; Nishida, T.; Zhang, X. Determinants of China's Technology Availability and Utilization 2006–2009: A Spatial Analysis. *Inf. Soc.* **2013**, *29*, 26–48. [CrossRef]
19. Song, Z.; Liu, W.; Ma, L.; Dunford, M. Measuring spatial differences of informatization in China. *Chin. Geogr. Sci.* **2014**, *24*, 717–731. [CrossRef]
20. Dewan, S.; Riggins, F.J. The digital divide: Current and future research directions. *J. Assoc. Inf. Syst.* **2005**, *6*, 298–337.
21. Pick, J.; Sarkar, A. Theories of the digital divide: Critical comparison. In Proceedings of the IEEE 49th Hawaii International Conference on System Sciences, Koloa, HI, USA, 5–8 January 2016; IEEE: New York, NY, USA, 2016; pp. 3888–3897.
22. Rogers, E. *Diffusion of Innovations*, 5th ed.; Free Press: New York, NY, USA, 2003.
23. Niehaves, B.; Plattfaut, R. Internet adoption by the elderly: Employing IS technology acceptance theories for understanding the age-related digital divide. *Eur. J. Inf. Syst.* **2014**, *23*, 708–726. [CrossRef]
24. Sipior, J.C.; Ward, B.T.; Connolly, R. The digital divide and t-government in the United States: Using the technology acceptance model to understand usage. *Eur. J. Inf. Syst.* **2011**, *20*, 308–328. [CrossRef]
25. Pick, J.; Sarkar, A. *The Global Digital Divides: Explaining Change*; Springer: Berlin/Heidelberg, Germany, 2015.
26. Roztocki, N.; Soja, P.; Weistroffer, H.R. The role of information and communication technologies in socioeconomic development: Towards a multi-dimensional framework. *Inf. Technol. Dev.* **2019**, *25*, 171–183. [CrossRef]
27. Ramadhanti, H.D.; Astuti, E.T. Digital Divide and A Spatial Investigation of Convergence in ICT Development Across Provinces in Indonesia. *J. Stat. Appl. Comput.* **2020**, *12*, 69–83. [CrossRef]
28. Myovella, G.; Karacuka, M.; Haucap, J. Determinants of digitalization and digital divide in Sub-Saharan African economies: A spatial Durbin analysis. *Telecommun. Policy* **2021**, *45*, 102224. [CrossRef]
29. Ntim, S.K. Covid-19 Pandemic and Disparity in Household Adaptations to School Lockdown: Redressing the Myth of Educational Equality. *Int. J. Educ.* **2022**, *14*, 37–55. [CrossRef]
30. Meso, P.; Musa, P.; Straub, D.; Mbarika, V. Information infrastructure, governance, and socio-economic development in developing countries. *Eur. J. Inf. Syst.* **2009**, *18*, 52–65. [CrossRef]
31. Ben, S.; Bosc, R.; Jinpu, J.; Li, W.; Simonelli, F.; Zhang, R. *Digital Infrastructure: Overcoming the Digital Divide in China and the European Union*; The Centre for European Policy Studies: Brussels, Belgium, 2017. Available online: https://www.ceps.eu/ceps-publications/digital-infrastructure-overcoming-digital-divide-china-and-european-union/ (accessed on 5 May 2024).
32. Azari, R.; Pick, J.B. Technology and society: Socioeconomic influences on technological sectors for United States counties. *Int. J. Inf. Manag.* **2005**, *25*, 21–37. [CrossRef]
33. Sarkar, A.; Pick, J.; Moss, G. Geographic patterns and socio-economic influences on mobile internet access and use in United States counties. In Proceedings of the IEEE Hawaii International Conference on System Sciences, Village, HI, USA, 4–7 January 2017; IEEE: New York, NY, 2017; pp. 4148–4158.
34. Sarkar, A.; Pick, J.B.; Rosales, J. Multivariate and geospatial analysis of technology utilization in US counties. *Telecommun. Policy* **2023**, *47*, 1–9. [CrossRef]
35. National Bureau of Statistics of China. *China Statistical Yearbook*; China Statistics Press: Beijing, China, 2021.
36. McFarlan, W.; Jia, N.; Wong, J. China's growing it services and software industry: Challenges and implications. *MISQ Exec.* **2012**, *11*, 1–9.
37. Statista. *Penetration Rate of Internet Users in China from 2008 to June 2022*; Statista: Hamburg, Germany, 2023. Available online: www.statista.com (accessed on 14 May 2024).
38. Statista. *IT Services—China*; Statista: Hamburg, Germany, 2023. Available online: www.statista.com (accessed on 27 April 2024).
39. Bon, A.; Akkermans, H.; Gordijn, J. Developing ICT Services in a Low-Resource Development Context. *Complex Syst. Inform. Model. Q.* **2016**, *9*, 84–109. [CrossRef]
40. Perrin, A.; Duggan, M. *Americans' Internet Access: 2000–2015*; Report; Pew Research Center: Washington, DC, USA, 2015.
41. National Telecommunications and Information Administration. *New NTIA Data Show Enduring Barriers to Closing the Digital Divide, Achieving Digital Equity*; May 11 Report; U.S. Department of Commerce: Washington, DC, USA, 2022.
42. Nishida, T.; Pick, J.B.; Sarkar, A. Japan's prefectural digital divide: A multivariate analysis. *Telecommun. Policy* **2014**, *38*, 992–1010. [CrossRef]
43. Pick, J.; Sarkar, A.; Rosales, J. Social Media Use in American Counties: Geography and Determinants. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 424. [CrossRef]
44. Blank, G.; Graham, M.; Calvino, C. Local Geographies of Digital Inequality. *Soc. Sci. Comput. Rev.* **2018**, *36*, 82–102. [CrossRef]
45. Plekhanov, D. Quality of China's Official Statistics: A Brief Review of Academic Perspectives. *Cph. J. Asian Stud.* **2017**, *35*, 76–101. [CrossRef]
46. Liu, S.; Chen, C. Regional innovation system: Theoretical approach and empirical study of China. *Chin. Geo-Graph. Sci* **2003**, *13*, 193–198. [CrossRef]
47. Koleski, K. *The 13th Five Year Plan*; Staff Research Report; U.S.-China Economic and Security Review Commission: Washington, DC, USA, 2017.
48. Lim, E.J.; Stubbs, J.; Xu, Q. Working with Chinese Government Data Sets: Potential Issues and Solutions. 2019. Available online: https://library.ifla.org/id/eprint/2511/1/185-lim-en.pdf (accessed on 14 April 2024).
49. Hair, J.F.; Ringle, C.M.; Sarstedt, M. PLS-SEM: Indeed a silver bullet. *J. Mark. Theory Pract.* **2011**, *19*, 139–152. [CrossRef]

50. Holz, C.A. Chinese statistics: Classification systems and data sources. *Eurasian Geogr. Econ.* **2014**, *54*, 532–571. [CrossRef]

51. Hair, J.; Hult, G.T.M.; Ringle, C.M.; Sarstedt, J. *A Primer on Partial Least Squares Structural Equation Modeling*, 3rd ed.; SAGE Publications: Los Angeles, CA, USA, 2022.

52. Kline, R.B. *Principles and Practice of Structural Equation Modeling*, 3rd ed.; Guilford Press: New York, NY, USA, 2010.

53. Kline, R.B. *Principles and Practice of Structural Equation Modeling*, 5th ed.; Guilford Press: New York, NY, USA, 2023.

54. Kline, R.B. Software Review: Software Programs for Structural Equation Modeling: Amos, EQS, and LISREL. *J. Psychoeduc. Assess.* **1998**, *16*, 343–364. [CrossRef]

55. Boomsma, A. Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika* **1985**, *50*, 229–242. [CrossRef]

56. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2021.

57. Smith, T.W. Developing comparable questions in cross-national surveys. *Cross-Cult. Surv. Methods* **2003**, *325*, 69–92.

58. Wilkinson, R.G.; Pickett, K.E. Income inequality and population health: A review and explanation of the evidence. *Soc. Sci. Med.* **2006**, *62*, 1768–1784. [CrossRef] [PubMed]

59. Petrović, A.; Manley, D.; van Ham, M. Multiscale contextual poverty in the Netherlands: Within and between-municipality inequality. *Appl. Spat. Anal. Policy* **2022**, *15*, 95–116. [CrossRef] [PubMed]

60. Casali, Y.; Aydin, N.Y.; Comes, T. A data-driven approach to analyse the co-evolution of urban systems through a resilience lens: A Helsinki case study. *Environ. Plan. B Urban Anal. City Sci.* **2024**. [CrossRef]

61. Campbell, D.T.; Stanley, J.C. *Experimental and Quasi-Experimental Designs for Research*; Rand McNally: Chicago, IL, USA, 1963.

62. Allison, P.D. Change Scores as Dependent Variables in Regression Analysis. *Sociol. Methodol.* **1990**, *20*, 93. [CrossRef]

63. Dalecki, M.; Willits, F.K. Examining change using regression analysis: Three approaches compared. *Sociol. Spectr.* **1991**, *11*, 127–145. [CrossRef]

64. Calinski, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **1974**, *3*, 1–27.

65. Central Compilation and Translation Press. *The 13th Five-Year Plan for Economic and Social Development of the People's Republic of China*; Central Compilation and Translation Press: Beijing, China, 2016.

66. Center for Security and Emerging Technology. *13th National Five-Year Plan for the Development of Strategic Emerging Industries*; Central People's Government of the People's Republic of China; Center for Security and Emerging Technology, Georgetown University: Washington, DC, USA, 2019.

67. Center for Security and Emerging Technology. *Outline of the People's Republic of China 14th Five-Year Plan for National Economic and Social Development and Long-Range Objectives for 2035*; Center for Security and Emerging Technology, Georgetown University: Washington, DC, USA, 2021.

68. Chinn, M.D.; Fairlie, R.W. The determinants of the global digital divide: A cross-country analysis of computer and internet penetration. *Oxf. Econ. Pap.* **2007**, *59*, 16–44. [CrossRef]

69. Auxier, B.; Anderson, M. *Social Media Use in 2021*; Pew Research Center: Washington, DC, USA, 2021.

70. Akiyoshi, M.; Ono, H. The Diffusion of Mobile Internet in Japan. *Inf. Soc.* **2008**, *24*, 292–303. [CrossRef]

71. Wang, T. *Making Sense of China's Economy*; Taylor and Francis: New York, NY, USA, 2023.