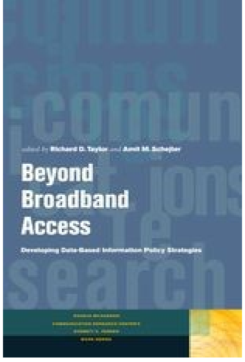


Beyond Broadband Access: Developing Data-Based Information Policy Strategies
Richard D. Taylor (ed.), Amit M. Schejter (ed.)
https://doi.org/10.5422/fordham/9780823251834.001.0001
Published: 2013 **Online ISBN:** 9780823268955 **Print ISBN:** 9780823251834

CHAPTER

2 Understanding Digital Gaps: A Quartet of Empirical Methodologies

Bin Zhang, Richard D. Taylor

https://doi.org/10.5422/fordham/9780823251834.003.0002 Pages 23–50
Published: June 2013

Abstract

This chapter presents a series of exercises in using statistical methods to analyze the digital divide, using China as an example. In particular, it describes four empirical methodologies and their underlying principles, along with their potential usefulness to planners and policy makers: static and dynamic analysis/analytic hierarchy process (AHP), hierarchical clustering analysis (HCA), time distance analysis (TDA), and data envelopment analysis (DEA). These four approaches have the same starting point, the Informatization Level index. The chapter suggests that there is no perfect way to measure either the digital divide or e-readiness, due to the limitations of dealing with proxies (indicators), and cites two general critiques: conceptual imprecision and issues of measurement. It calls for the development of a general theory of informatization, of which the analysis of the digital divide and e-readiness are just particular applications.

Keywords: statistical methods, digital divide, China, static and dynamic analysis, analytic hierarchy process, hierarchical clustering analysis, time distance analysis, data envelopment analysis, e-readiness, informatization

Subject: Museums, Libraries, and Information Sciences

Since the 1960s, the field of information studies has had a tradition of trying to understand the role of information in society by measurement, typically by counting things: media, words, bits, and so on. This reflects an intuitive sense that something important is happening. Because of the intangible nature of the subject "information," however, its role has been hard to grasp. This goal manifests itself in current times most often as the study of the so-called *digital divide* (or *e-readiness*). Approaches to this matter have grown more sophisticated over time, and now use complex statistical methodologies to parse huge databases for lessons from the past for potential gains from shaping the future.

This chapter provides a detailed exercise in four of those methods, as applied to one country, China, but with the view that their underlying principles may be of broad application and of substantial use to planners and policy makers.[1] These are not the only approaches—there are many—indeed, so many that the very idea of finding coherence in the field is challenging. While much progress has been made, it is argued that going forward there needs to be a new way of thinking about this field, and new emphasis on theory and testing as a way of developing analyses that are both explanatory and predictive.

This study is the product of an international collaboration (United States and China) to advance the thinking in this field, which in China is broadly referred to as *informatization*, a concept that resonates with industrialization—a sweeping industrial and social change affecting all aspects of life and society. China has

p. 24 highly prioritized informatization, and created a top-level state ↳ council Leading Group to coordinate and promote the concept. The United States has no such grand metaphor, focusing instead on *universal service* and *broadband access*, which are much more limited concepts. Continued advancement of theory and practice in this field will be of greatest benefit to whoever advances informatization the most. Our intention with this study is to encourage scholars and policy practitioners to give informatization greater consideration.

We chose China as a case study because it has pursued the empirical study of informatization for some time, and has collected extensive relevant data. This chapter provides examples of four methods for analyzing that data, with the hope that their underlying principles may be of general application and useful to policy makers. These methods are:

- Static and dynamic analysis/analytic hierarchy process (AHP)
- Hierarchical clustering analysis (HCA)
- Time distance analysis (TDA)
- Data envelopment analysis (DEA)

The examples are different in name, but similar in purpose. They have the same starting point, the Informatization Level index, described below. They use different assumptions, and the calculations are different. Their common purpose is to objectively describe and analyze the digital divides within China. Based on the particular situation and the data available, one or several methods may fit better than others. Rarely will one method describe all the relevant considerations. In the examples in this chapter, the results are in some ways different, providing different insights; but also similar, in that their respective conclusions converge. Which method to choose depends both on the researcher's goals and the philosophy.

Case 1: Using Static and Dynamic Analysis/Analytic Hierarchy Process

Based on a review and analysis of the structure (inductive/deductive), statistical methods, conclusions, and comparative strengths of twenty-eight existing digital divide index systems, a comprehensive index system was developed for the measurement of regional digital divides in thirty-one regions of China. Using the AHP model, rankings of indicators were obtained from the work of experts using factor analysis, and indicators measuring more effective access were given higher weights. Index weights were then determined for twenty-nine factors by computing a comparison matrix. Using regional data in China from 2002 to 2007,
p. 25 normalized by means of equalization, the ↳ index values and ranks of informatization levels for each region for each year were obtained; then mean deviation was used to analyze the changing trends in the digital divides during those years. Attention was focused on five core factors affecting digital divides in China: technology, economy, government, education, and society.

Case 1 focuses on the measurement of the *digital divide*: this refers to the *effective access* gap between regions in information and communication technology (ICT). *Effective access* targets the uses of ICTs through which people, organizations, or society can obtain economic and cultural advantages. If people who have access to ICT do nothing but upload or download music online, then their access to ICT does not make sense (is not *effective*) because they lack the ability or the opportunities or the will to use ICT to broaden their cultural knowledge or better their economic situation. So the importance of effective access is emphasized.

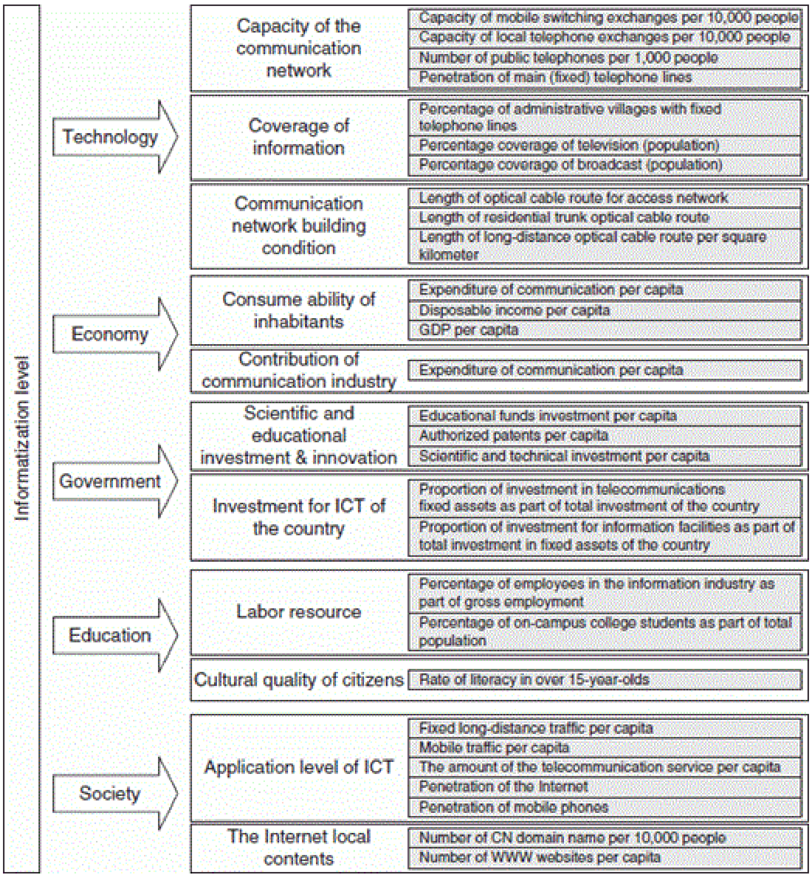
There are twenty-nine indicators in this index system, which reflect the main factors affecting the digital divide. Indicators measuring effective access are given higher weights. From this point of view, therefore, AHP has greater flexibility, and under the guidance of experienced experts, becomes a more accurate measurement model for the selected target.

In recent years China has increased its level of informatization rapidly. All regions have shown a narrowing trend in their digital divides, yet the real levels of the regions and the gaps between them are not precisely known. Due to the fact that the indicators have different measurement units and orders of magnitude, equalization is used to make indexes dimensionless during the process of static analysis. Thus the *mean deviation* is used as a reference to describe the digital divide in a dynamic analysis—to observe changes over time, which is the more intuitive way to present the regions' real informatization levels and the disparities between them and the average level.

Index System

The *informatization level* (IL) developed in this case is a composite index. Specifically, there are twenty-nine indicators composing the entire framework—seven indicators for Society, ten for Technology, five for Government, four for Economy, and three for Education. Figure 2-1 shows the structure and indicators of the index.

Figure 2-1.



Hierarchical structure of the informatization level.

Determination of Index Weights

Using weighting scores from experienced experts, AHP was used to derive the local and global priorities; the consistencies of hierarchy (overall and local) were also checked. As a result, the weights of every level in the informatization level framework were determined. In accordance with the results ↳ p. 26 of the local priorities, global priorities were computed as well. The complete results will be provided on request. Some of the more noteworthy results are:

- The application of ICTs is shown as the key aspect of this index system, and the weight of Penetration of the Internet hit 0.147, which is the highest among all indicators. It suggests that indicators related to the application of the Internet have been stressed, which demonstrates the concept of the digital divide emphasizing *effective access*.
- Besides emphasizing application, this index system indicates that the Economy and Education components have significant influence upon the regions' informatization levels. This is evidenced by the expenditure of communication per capita and rate of literacy over ↳ p. 27 fifteen years old, both reaching 0.103. It also shows that the digital divide is not just equality of access; to truly and effectively narrow the digital divide we have to improve the citizens' cultural quality, literacy, awareness of access to information and extent of their ICT application level.
- In terms of technology, more attention has to be paid to the coverage of information; from an economic perspective, the study focuses on the consuming capacity of inhabitants; and in the government component, investment for ICT of the country is quite vital.

Static Analysis of the Digital Divide in China

Data normalization. Due to the fact that indicators have different dimensions, the data was normalized within the same indicator by using the following formula (equalization):

$$x'_{ij} = x_{ij}/\bar{x}_j$$
$$(\text{if } \bar{x}_j = 0, \text{ let } \dot{x}'_{ij} = 1 + x_{ij})$$

Using the original data divided by the mean of data of the same variable, the mean of every indicator turns out to be 1, and the standard deviations are the coefficients of variation of the original variables. This method effectively eliminates the impact of dimension and magnitude, and at the same time it retains the variation information of the original data. The greater the variation, the greater will be the influence on the comprehensive analysis. This kind of undimensionalization tries to preserve the variation information through the coefficients of variation of the original variables, not the standard deviation of the original variables, which can save both the comparability and the variation information of the original data.

Analysis of index values and ranks. Applying the index system and normalization method discussed above, the index values were developed. The analysis shows that:

- Beijing, Shanghai, Guangdong, Zhejiang, and Tianjin, which ranked at the top of the list in recent years, maintained their leading positions during the six years.
- Regions that made gradual progress, indicating that the importance of informatization has been recognized step by step, and that scored some achievements in these areas include Jiangsu, Hainan, Shanxi, Jiangxi, Inner Mongolia, and Hena.
- Liaoning, Jilin, Hebei, and Yunnan apparently showed deterioration, which reflects that those regions had lower awareness p. 28 ↳ of informatization building compared to other areas, or lack of sustainable development, causing the backward rankings. On the other hand, it implies those regions have more room for improvement in the future.
- If we take a closer look at the results for 2007, we find that Beijing and Shanghai had much higher index values than all the other regions, with figures arriving at over 2.7 and 2.5, respectively. They were the top group of all the regions, presenting advanced informatization levels and there was very large disparity between them and other areas. Guangdong, Zhejiang, Tianjin, and Fujian were the next highest group, with index values all above 1.1, which indicates that those regions were in a relatively good position; and there were smaller gaps between the remaining regions, which all pointed at an intermediate or low level.
- *Group analysis of index values:* Table 2-1 shows the results when applying a descriptive statistical analysis to these numbers.

Table 2-1. Descriptive Statistics from 2002 to 2007

	2002	2003	2004	2005	2006	2007
Maximum	3.39502	3.24164	2.99855	2.88652	2.92703	2.71674
Minimum	0.57630	0.59443	0.60965	0.61940	0.66275	0.67054
Mean	1	1	1	1	1	1
Case Number	31	31	31	31	31	31

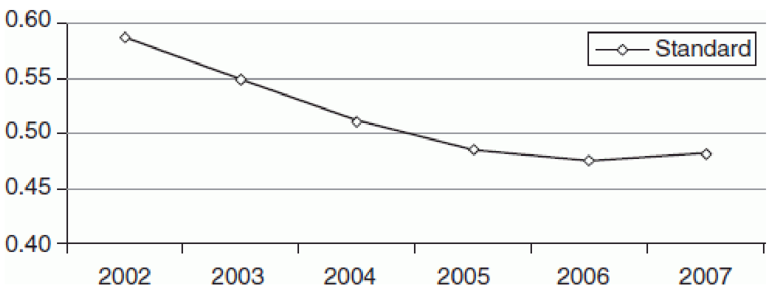
Table 2-1 shows that in these six years, the minimum index values increased moderately from 0.57630 in 2002 to 0.67054 in 2007, approaching the mean 1. Meanwhile, the maximum decreased from 3.39502 in 2002 to 2.71674 in 2007, approaching the mean 1 as well. So we can see that the digital divide in China has been narrowing in recent years.

But these conclusions only provide rough information about the entire body of data. For a closer look at the index values, we divided the data into ten groups. They are: below 0.5 (the first group), 0.5–0.625 (the second group), 0.625–0.75 (the third group), 0.75–0.875 (the fourth group), 0.875–1 (the fifth group), 1–1.125 (the sixth group), 1.125–1.25 (the seventh group), 1.25–1.375 (the eighth group), 1.375–1.5 (the ninth group), over 1.5 (the tenth group). We classified and applied frequency statistics, so that the following additional conclusions could be reached.

p. 29 The number of regions that can be included in the third group (0.625–0.75), the fourth group (0.76–0.875), and the fifth group (0.875–1) were the majority of the total thirty-one. Moreover, the numbers grew slightly in the six-year period. All of this shows that the index values of the thirty-one regions tend to be concentrated in the center.

There is a trend towards concentration seen in the grouping process, and in order to better support the conclusion by means of data, we computed the standard deviation of the index values for the six years and the result is presented in Figure 2-2. It is manifest that the standard deviation generally declined from 2002 to 2007, especially during the period between 2002 and 2006, when it showed a clear drop, which suggests that the digital divide was controlled during those years. But there was a small increase in the figure from 2006 to 2007, which showed the gap widened slightly. In fact, the Internet and the number of Internet users in China saw rapid growth in 2006, which was mainly due to the dramatic expansion of the Internet and the drastically increasing number of websites, web pages and Internet users in the eastern part of China. This unparalleled situation led to a greater gap between the eastern and western parts of China. So here we need to analyze the digital divide using both horizontal and vertical comparisons.

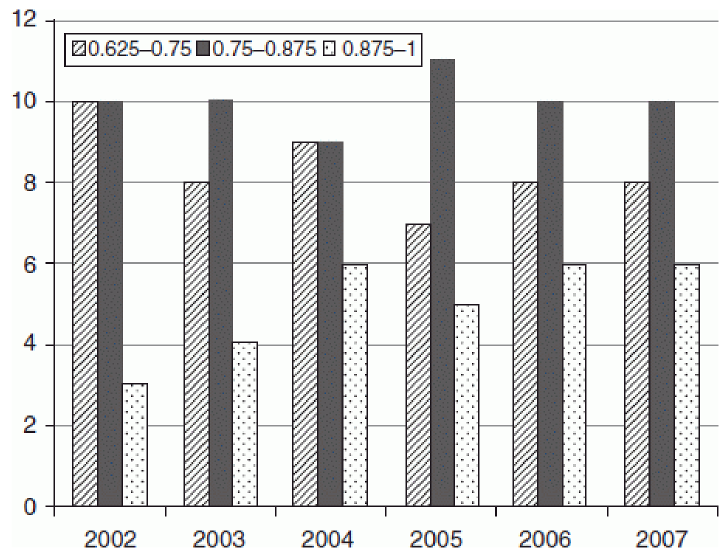
Figure 2-2.



Standard deviation of the index values, 2002–2007.

Further analysis shows that the frequency of the 0.625–0.75 group fluctuates drastically, but it stopped at 8 in 2007, which followed a drop compared to 10 in 2002; the figure of the 0.76–0.875 group climbed with less dramatic fluctuation, ending up with the same level in 2007 as it did five years ago (10); however, the number of the 0.875–1 group generally rose in those years, and reached 6 in 2007, which doubled the magnitude in 2002 (see Figure 2-3).

Figure 2-3.



Frequency statistics for three groups, 2002–2007.

Overall, whether through charts or data, we can see the current levels and developing trends of informatization in China, that is, the digital divide in China is narrowing, and the informatization level
p. 30 index is centralizing to 1. ↳

Dynamic Analysis of the Digital Divides in China

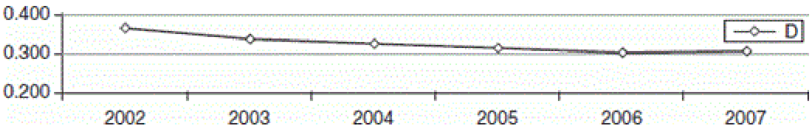
In the prior analysis, the index value for the digital divide and the ranks of thirty-one regions were obtained, and equalization methods for detailed analysis were applied. A number of other methods can also be applied to calculate the digital divide. In order to see changes in the digital divide over time, the concept *mean deviation* (a more intuitive way to show the disparity between the index value of a specific region and the mean value) is referred to, to determine the shape of the digital divide over time as presented below (assume the year has been fixed):

$$Digital\ Divide\ (DD) = \frac{\sum_{i=1}^{n} |Index\ value(i) - 1|}{n}$$

DD stands for Digital Divide; *Index value* means the index value of informatization level in the No. *i* region; *n* equals the total number of regions in our research, which is 31 including provinces, municipalities, autonomous regions and municipalities; 1 stands for the mean of index values.

Figure 2-4 shows the changing trend of the digital divide from 2002 to 2007 according to the computed
p. 31 results and the digital divide measurement model. ↳

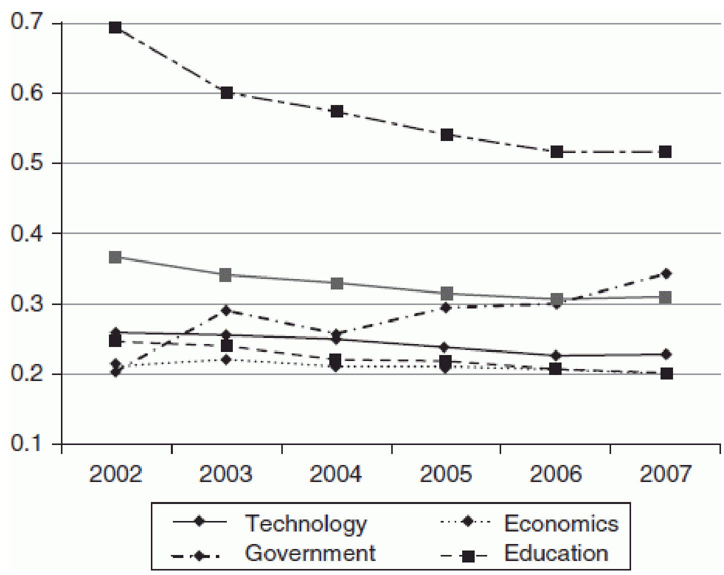
Figure 2-4.



Changing trend of the digital divide in China, 2002–2007.

The result indicates the different conditions of informatization in China's thirty-one regions, and the reduction in the digital divide over recent years accordingly; in other words, the gaps between regions are being bridged. Starting at 0.367, the number plunged rapidly from 2002 to 2005, followed by a steady period during the next two years, then leveling off at approximately 0.309. Overall, the result shows a decreasing trend, as seen in Figure 2-5.

Figure 2-5.



Digital divide trends comparing five indexed components, 2002–2007.

To better understand the cause of the narrowing trend, it is necessary to analyze the changes from the five key components: Technology, Economy, Government, Education, and Society.

- The Society index mainly describes the application of ICT. Its values are much higher than the others over recent years. But at the same time, its decline was very fast until 2006 and 2007 when it began to stabilize. That suggests that instead of the huge differences between regions in the past, now almost all
p. 32 regions have an increasing awareness of the importance of the application of ICT, and have ↳ started using ICT products and services simultaneously. This change, therefore, brought about a progressively reduced gap in ICT application and produced a rising informatization level in China.
- The Technology index primarily embodies the ICT infrastructure. The index values show that there was a small divide in this factor, with a slow decrease during those years which followed a similar trend with the average divide. This indicates that the whole country has been paying relatively greater attention to the ICT infrastructure.
- The Government index chiefly represents ICT and educational investments from the government. Its values fluctuated more strongly than the other ones, and interestingly its values have grown over the years, meaning that the investment divide of Government seemed to widen between regions as opposed to narrowing. This surprising situation shows that the government formulated national ICT development policies with some strategic considerations. Yet, there are gaps between regions in terms of support and investment for ICT, which is the foundation for further improvement.
- The values of Economy and Education were both small and showed a gradual decrease, following the same trend as the average. This suggests that education and the economy were generally less emphasized.

All in all, the digital divide in China has dropped from year to year.

Case 2: Using Hierarchical Clustering Analysis

case 2 adopts the indicators system in case 1. Then an HCA for China's Digital Divide Index in thirty-one regions was carried out from 2002 to 2007. The results of this approach provide a deeper and more useful understanding of the digital divide in China. case 2 also takes into consideration the five core factors, that is, Technical, Economic, Governmental, Educational, and Social factors, and ranks each indicator for the thirty-one regions from 2002 to 2007. Based on this, through relevance measure clustering, case 2 identifies twelve different types of factors which influence the digital divide. It analyzes the reasons for rank changes in the digital divide index in some provinces, and suggests policy directions.

An advantage of cluster analysis is that it can simplify a situation that might otherwise be too complicated to analyze. There are so many provinces in China that it is necessary to cluster them by their similarities relative to the digital divide. In case 2, hierarchical clustering is used to cluster the thirty-one regions based on their digital divide indexes. Hierarchical clustering cannot only be used to conduct horizontal
p. 33 comparisons among regions, but ↳ also vertical time-distance comparisons of classifications and rankings of regions, which can lead to a clearer understanding and awareness of the changes of the digital divide status of regions during the six years studied.

The similarity method of hierarchical clustering. The method used to determine the degree of similarity between samples is called the *similarity measure*. When different types of similarity measures are used, the same sample can be divided into different classes. Therefore, it is important to be clear about the kinds of similarity measures used to classify when we do hierarchical clustering analysis. There are many kinds of similarity measures, which can be divided into two types: relevance measures and distance measures. A distance measure focuses on the comprehensive distance between samples, while a relevance measure focuses on the structural similarity between samples.

Explanation of the index system and weight distribution. The same index system is used as in case 1. There are two reasons for reusing that index system. First, it was systematically derived from twenty-eight prior index systems, so it is comprehensive. Second, using different methods on the same index renders the results comparable. The same applies to the weights. So here we also adopt the weight distribution of case 1.

The Hierarchical Clustering Process

The data of thirty-one regions from the year 2002 to 2007 is used as the sample data, using SPSS to apply the hierarchical clustering method for analysis. Those data are collected from the *China Statistical Yearbook* of the National Bureau of Statistics of China, the *Statistical Yearbook of Communication*, *Information Statistical Yearbook*, *Communication Statistics Annual Report*, and the *Internet Statistics Report*" of CNNIC. The procedure of hierarchical clustering is as follows:

1. Data standardization: As the method "mean of 1" can eliminate the effects of the dimension and order of magnitude, while still retaining the information about the degree of variability of the original data, we use this method to standardize data, that is, to make the mean value of data in the scope of 1. The formula is:

$$x'_{ij} = x_{ij}/\bar{x}_j, \text{ if } \bar{x}_j = 0, \text{ then } x'_{ij} = 1 + x_{ij}.$$

2. Calculate the values of the criterion levels' indicators—which are Technology, Economy, Government, Education, and Society—and the digital divide index (the higher the value the higher the level of informatization, which means the digital divide between this subject province and the most advanced level is smaller).
3. Cluster the variables according to the digital divide index and get the result of clustering. We use p. 34 distance measurement to measure the ↳ comprehensive distance among samples, which means to select the Squared Euclidean distance in SPSS. We use the *between-group linkage* (also known as the *category average*) method to measure the distance between the different categories.
4. We can cluster the variables in accordance with the ranking of five indicators for criterion level, that is, Technology, Economy, Government, Education, and Society. We use *Cosine* in SPSS, a Relevance Measure method, to measure the degree of structural similarity for samples. We also use the between-group linkage method to measure the distance between the different categories.

Clustering Results and Analysis

Clustering the data for six years by distance measurement. One hundred and eighty-six digital divide indexes—which were samples calculated by using the data of thirty-one provinces from 2002 to 2006—divided into twenty-four categories, which were then clustered. The provinces were then ranked in accordance with the distance between the samples.

case 2 shows the differences in the level of informatization for all regions in the years from 2002 to 2007. The samples are divided into twenty-four categories and there are digital divides between categories. Beijing and Shanghai are clearly in the leading position in their level of informatization. Guangdong, Tianjin, Zhejiang, and Fujian follow closely. The level of informatization of Gansu, Guizhou, and Anhui are lower, and the rest of the provinces are in the mid-range. The time distance of the digital divide between regions by the category rankings of all 186 samples can be analyzed. For example, if category 6 included "Beijing 3, Shanghai 6," we could say that the level of ICT development of Shanghai in 2006 had almost reached the level of Beijing in 2003, which means the time-distance between Beijing and Shanghai would be three years.

Clustering for each year's data by distance measure. Based on the distance matrix obtained from cluster analysis by distance measure, we can get the ranking of the informatization levels among regions in each year. The better the ranking the narrower the digital divide; and the worse the ranking the larger the digital divide.

According to digital divide indexes, we cluster thirty-one regions into twelve categories, and in accordance with the ranking of regions, three types of the categories can be merged into one class, and eventually thirty-one regions are divided into the following four classes:

- First-class regions: smallest digital divide, highest ICT application level.
- Second-class regions: smaller digital divide, higher ICT application level.

p. 35 ↳ • Third-class regions: medium digital divide, medium ICT application level.

- Fourth-class regions: large digital divide, low ICT application level.

At the same time, in each class, the regions are divided into three categories, representing three ICT application levels in each region: high, medium, and low.

According to the classification table for the digital divide for 2002–2007, the total classification table of those six years was produced. Beijing and Shanghai, as the leading regions in ICT application, have been in the first class, and their digital divide index rankings have stayed first and second for six years.

Clustering the data for six years by relevance measure. From 2002 to 2007, ranking for Technology, Economy, Government, Education, and Society indicators for each year among the thirty-one regions were made into a set of data. By relevance measure clustering for this set of data, the thirty-one regions can be divided into a variety of types. In clustering the results using SPSS, twelve types of regions were selected and fine-tuned based on the actual situation. Then the characteristics of each type were summarized.

The clustering results can be used as the reference for regional classification. When the characteristics of each type are described, the use of “outstanding” for a certain indicator means the ranking of that indicator is relatively higher than others, and “performs evenly” means the ranking of that indicator is a little lower than the “outstanding” indicator. When a certain indicator’s ranking is “relatively weak,” it means that indicator’s ranking is much lower than the “outstanding” indicator; the specific difference between them depends on the outcome of the clustering.

Reasons for changes in digital divide ranking for regions in 2002–2007. We consolidated all of the rankings data to do the cluster analysis. There was interaction between different years’ data, so it was not useful for making policy based on the performance of each year. Because of this, relevance measure clustering was done separately for each criterion level indicator of thirty-one regions. Since showing all the results would require too much space, only one region, Hainan, is analyzed here as an example to show the rationale of ranking changes and to suggest policy implications.

Hainan. There were fluctuations in Hainan’s digital divide index rankings, but not large. As to the classification by distance measure clustering, Hainan showed a steady increase in the third-class regions. Hainan stayed in the low end of third-class regions from 2002 to 2004, in the middle in 2005 and 2007, and in the high end in 2006.

The data show that from 2002 to 2003, Hainan’s digital divide index ranking decreased eight positions. p. 36 There was a slight increase in the ↳ Education indicator in 2003, but the small increase did not compensate for the impact of the decline in other indicators’ rankings, especially the Government and Economy indicator rankings’ sharp decline, which was the main reason for the decline of Hainan’s digital divide index ranking. Although the ranking of its Technology indicator dropped, compared to other indicators’ rankings, the Technology indicator ranking was still better, so Hainan was advanced by technology.

During 2003 and 2005, Hainan belonged to the “promoted by Government” type. The substantial increase of the Government indicator ranking and the slight increase of the Technology and Economy indicator rankings in 2004 prompted the digital divide index ranking of Hainan to sharply rebound seven positions.

From 2004 to 2005, despite its Government indicator ranking dropping one position, Hainan still belonged to the “promoted by Government” type. Its Society indicator ranking’s substantial increase made its digital divide index ranking rise one position. Since the Society indicator refers primarily to the ICT application level, we can say that a significant increase in the level of ICT applications was the main reason for the increase of the digital divide index ranking.

From 2005 to 2007 Hainan belonged to the type “promoted by Government and Society.” Although its Technology indicator ranking had a sharp decline in 2005–2006, the strong ranking of Government and Society indicators and the increase of other indicators’ rankings, especially the significant increase in its Economy indicator ranking, made Hainan’s digital divide index ranking rise two positions. The important influence of economic factors on Hainan’s digital divide index ranking started to appear.

From 2006 to 2007 the Government indicator ranking increased slightly, and the Economy indicator ranking continued to show a significant increase. The economic factors had become the most important factors to enhance Hainan’s digital divide index ranking. At the same time, the Education indicator ranking advanced one position, but was still relatively backward. The Society indicator ranking dropped one position, while the digital divide index ranking remain unchanged.

During 2002 and 2007, the Government indicator showed the greatest positive impact on the digital divide index ranking of Hainan, followed by the Economy indicator, but the pulling effect of the Economy indicator was gradually increasing, which meant Hainan was relatively positive in national ICT investment, scientific and educational input and innovation, and economic development. It is noteworthy that Hainan's Education indicator rankings have always been low and the Technology indicator (which mainly refers to a variety of coverage and hardware subindicators) rankings sharply declined since 2006.

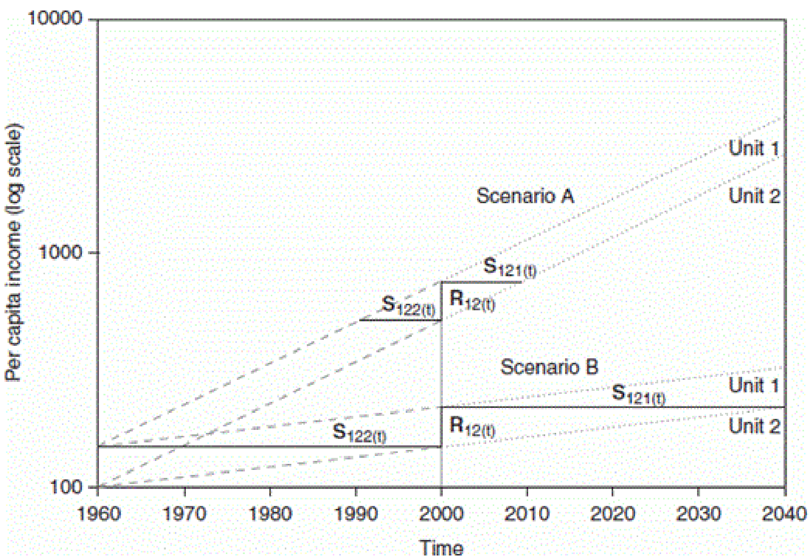
p. 37

Case 3: Using Time Distance Analysis

Typically in research, time is used as a key dimension for analysis and comparison. However, in studies measuring the digital divide, deep information contained in the time dimension has often been overlooked. The TDA approach can put the time dimension into such studies, showing the degree of change and the trends of variables over time, and reflecting the dimensions of the digital divide more clearly while providing the basis for further policy research.

The development trend for every province from 2002 to 2007 can easily be obtained by comparing the difference between the digital divide indexes of each province and the national average, as the calculation of time distance is based on interpolation and compound growth rate. For example, assume that after 2007 the growth rate of certain indexes of Sichuan keep the same compound growth rate as between 2002 and 2007. Based on this premise, the strength of input of every index can be calculated as a reference for every province. The concept of time distance provides a new perspective for research on the digital divide. It also can measure the gaps between selected provinces and the national average. Figure 2-6 shows an example of the relationship between growth, efficiency and inequality.

Figure 2-6.



The relations between growth, efficiency and inequality when based on a dynamic concept of overall degree of disparity.

Static relative measure and time distance lead to different conclusions. Now let us take a broader view of the situation. The concept of time distance

p. 38

for a given level of an indicator as one of the dimensions of disparity leads to a different conclusion about the degree of disparity in scenario A and in scenario B, as can bee seen in Figure 2-6. In the 4 percent growth rate for scenario A with the 50 percent static relative disparity (the more developed region has a 50 percent higher value of the indicator) the time distance between the two regions is ten years. In scenario B with 1 percent growth rate and the same static relative disparity the time distance between the compared regions is forty years. It is highly unlikely that people would perceive such situations as equal degrees of disparity, even though the static measure of disparity is 50 percent in both cases. Higher growth rates lead to smaller time distances, and thus have an important effect on the overall degree of disparity. This is based on both static and time distance, as both matter. Static measures alone are inadequate.

Digital Divide Measurement Based on Time Distance Analysis

Actually there exists a significant digital divide between the thirty-one provinces in China. The level and the penetration rate of ICT technology among the provinces are totally distinct. Time distance analysis, which is different from the absolute and relative difference analysis, can show us the digital divide from another perspective.

Digital divide index system. The same index system and weight distribution will be employed again, but will be analyzed through time distance analysis. The index system and weight distribution were presented in case 1.

Time distance comparison of the digital divide index between thirty-one provinces and the national average. According to the indicator system and corresponding data of the digital divide index above, the digital divide index of thirty-one provinces and the national average from 2002 to 2007 were obtained. The procedure is as follows: first, the data from 2002 to 2007 was aggregated, and the national average value of every index in these six years as the standard value was considered. Then, based on the standardized data and the weight of every indicator, the composite digital divide index of every province was calculated, including the indexes of Economy, Education, Government, Technology, and Society. Then one-dimensional and two-dimensional time distance comparisons were made with these indexes.

One-dimensional time distance comparison. The time distance comparison of the digital divide index between every province and the national average from 2002 to 2007 was calculated. It showed that the digital divide is gradually widening. For example, the digital divide index of Beijing, Shanghai, Tianjin, and so on is leading the national average level by three to six years in 2002 and by four to eight years in 2007. The
 majority of backward provinces fall behind the national average by two years in 2002, but it becomes ↳ two to four years in 2007. Therefore, it can be concluded from the general trend of the digital divide index from 2002 to 2007 that the positive and negative digital divide are all getting deeper.

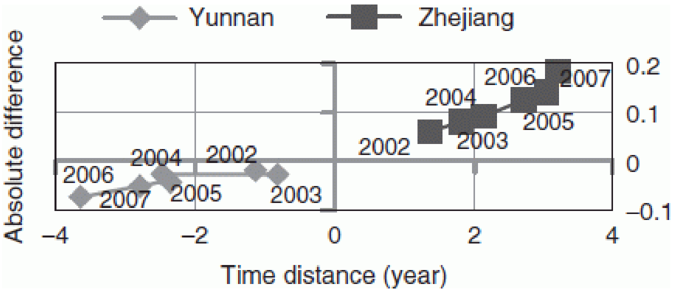
Considering the trend of the time distance between the digital divide index of thirty-one provinces and the national average, it can be seen that three districts have an obvious difference, as shown in Table 2-2. District 1 is a positive area. The aggregative indicator value has kept ahead of the national average level. District 1 is mainly in the eastern part of China, such as Beijing and Shanghai. District 2 is composed of both positive areas and negative areas. The aggregative indicator value is almost equal to the national average, so the digital divide is not so apparent. District 2 is mainly in such provinces as Hainan and Shaanxi. District 3 is negative area. The aggregative indicator is behind the national average level. It should be highly developed in the future. If the current situation continues, it will lead to an increasing digital divide in this district. That will make these provinces continuously lag behind. District 2 is mainly in such provinces as Guizhou and Gansu.

Table 2-2. Three Districts with Obvious Differences in their Digital Divides

Area	Time distance in 2007	Trend of digital divide	Main provinces
District 1	More than 0	Positively widen	Beijing, Shanghai, Tianjin, Guangdong, etc.
District 2	More than –1, less than 1	In the range of plus or minus	Hainan, Shanxi, Jilin, Xinjiang, etc.
District 3	Less than 0	Negatively widen	Neimeng, Jiangxi, Hebei, Xizang, etc.

Two-dimensional time distance comparison. Two provinces (Zhejiang and Yunnan), which respectively represent the well developed and the poorly developed in terms of the digital divide, were chosen for deeper analysis. The broken line graph of the 2D time distance comparison is shown as Figure 2-7. The developing trend is obvious. The two polygonal lines are distributed in the first quadrant and third quadrant. If the digital divide of a district is less than the national average, the absolute difference is positive, and the time distance is also positive. So the digital development of all the provinces in the first quadrant is better than the national average.

Figure 2-7.



2D polygonal lines show the time distance comparison of observed provinces (Yunnan, Zhejiang), 2002–2007; positive figures mean ahead of the national average.

It can be seen from the polygonal line that Zhejiang is in the first quadrant, and Zhejiang developed well from 2002 to 2007, with its digital divide index and increasing rate always above the national average. So the development p. 40 track of Zhejiang shows a kind of linear increase and the same for the time distance and absolute difference.

The absolute difference of the digital divide index of Yunnan is below the national average by 0.05 from 2002 to 2005, and its fluctuation is not quite obvious. However, removing consideration of the slight reduction of the time distance from 2002 to 2003 and from 2004 to 2005, the time distance dimension is increasing rapidly. The digital divide between Yunnan and the national average is growing. Because the digital divide index in 2007 shows negative growth, this leads to the rapid increase of the time distance. In order to understand the specific reasons for the changes in the digital divide, a detailed analysis is presented here only of Zhejiang and Yunnan.

Time distance analysis of Zhejiang. The digital divide index consists of five sub-indexes: Technology, Economy, Government, Education, and Society. And these indexes consist separately of several specific actions each. Therefore, the key elements of the digital divide and the reasons which lead to the change of time distance can be shown clearly by comparing the time distance between the five subindexes of Zhejiang and the national average.

The digital divide index of Zhejiang led the national average by 1.36 years in 2002, mainly due to the Technology and Economy indexes. The Technology index led the national average by 3.07 years and the Economy index led by two years. This is the main reason that the digital divide index of Zhejiang led the national average. But the performance of its Education index and Government index is poor. The Education index and Government index are respectively 0.67 and 0.55 years behind the national average. Therefore, in 2002 the leading position of Zhejiang mainly depended on its Technology and Economy indexes. At the same time, the Education and Government indexes still should be improved.

The digital divide index of Zhejiang in 2007 exceeded the national average in 2002 by more than 3.19 years. It is again due to the Technology and Economy indexes. The Government index fell behind the national p. 41 average compared with the prior situation. Therefore, although the advantages of Zhejiang are its Technology and Economy indexes, the Government index should be emphasized more, in order to improve its position.

Time distance analysis of Yunnan. The digital divide index of Yunnan lagged the national average by 1.17 years in 2002. The root reason is its Education index, which fell behind the national average by 6.02 years. This enormous weakness led to the large gap between the digital divide index of Yunnan and the national average. The Technology, Government, and Society indexes are each behind their own national averages, but they are all above the digital divide index. The only leading index is the Economy index, which performed the best in all indexes in 2002, with a lead of 0.3 year over the national average. Therefore, the Education index mostly enlarged the gap between Yunnan and the national average in 2002. More attention needs to be given to improving it.

The digital divide index of Yunnan lagged the national average in 2007 by 3.65 years more than in 2002. And again it is due to the Education index. The whole situation has not been improved since 2002 because of the poor basis of the Education index. It can’t be changed in the short term. The Technology and Economy indexes are both below the national average. It leads to their time distance getting larger in different degrees. The Government and Society indexes have both been improved. Both of them have a little lead over the national average. So the Education index is the key point which should be emphasized to gain further improvement.

Case 4: Using Data Envelopment Analysis

Data envelopment analysis (DEA) is an efficiency evaluation method based on the concept of relative efficiency. As an efficiency evaluation method, DEA is a unified approach to the various evaluation systems used to evaluate the effectiveness of units, based mainly on the input-output indicators evaluation system, which establishes the evaluation model for deriving the efficiency value for each unit.

In case 4, DEA is used to study the digital divide based on five core components: Technology, Economy, Government, Education, and Society. In order to better analyze the impact of the various factors on the digital divide, the efficiency of input-output was divided into two levels for comparison: the analysis of the conversion efficiency of technology and of the conversion efficiency of social applications. Then, based on a production-profitability matrix, all the efficiency values are plotted into two dimensions in four quadrants, divided by the two reference points of the national average efficiency in each year. In each quadrant, representative provinces are selected for study, and for each, recommendations are proposed for improvement.

p. 42

Application of Data Envelopment Analysis

Indicators system's structure. The indicators system follows the same structure as in the previous three cases.

Procedure of analysis: In order to get a better analysis of the impact on the digital divide of the various factors, we divide the efficiency of input-output into two levels to provide a comparison, leading to a further analysis of the conversion efficiency of Technology and Social applications (see Tables 2-3 and 2-4).

Table 2-3. First Level of Input-Output Analysis—Technical Efficiency

Input indicators	Output indicator
1 Economic	Technique
2 Government	
3 Education	

Table 2-4. Second Level of Input-Output Analysis—Social Efficiency

Input indicators	Output indicator
1 Economic	Society
2 Government	
3 Education	
4 Technique	

Based on the input-output models shown in Tables 2 and 3, Economy, Government, and Education are used as indicators of inputs in the DEA model of influence factors; Technology is used as a first-class output indicator to inspect the efficiency of input factors into the related indicators. Social factors are treated as a second output indicator, estimating whether economic policy, education, and development of ICT technology applications in the community are effectively transformed.

Data Collection

The data sources for this analysis are the *China Statistical Yearbook* (2003–2007), *CNNIC Statistical Survey Report on the Development of the Internet* (2003–2008) and the *China Information Yearbook* (2003–2007). Next, we have to convert the Economy, Government, Education, Technical, and Social components into the corresponding indicators of inputs and outputs with integrated values.

p. 43 The geometric mean is used instead of the algebraic average because the geometric mean represents more balanced value judgments, not the development of indexes of inequality.

The index is:

$$\hat{I}_t = \sqrt[n]{\prod_{i-1}^{n} I_t^i}$$

I_t on behalf of the value of the index T factor; *n* on behalf of the number of independent indicator composed of each factor; *i* on behalf of one of the *i*th indicators. For technology, n=10; for the economy n=4; for the government n=5; for education n=3; for society n=7.

Evaluation Results and Analysis

The C^2R model is used to establish the corresponding linear programming model for the provinces and municipalities (decision-making unit) to obtain the decision-making units (DMUs), in order to express more clearly the concept of conversion efficiency.

Based on the adoption of a *DEA efficient* defined model, we can see:

- When technical efficiency or social efficiency is 100 percent, and the remaining variables or the slack variables of the influencing factors are 0, the result is referred to as *technical DEA efficient* or *social DEA efficient*, indicating that when the application of the impact of technology or social factors has reached the maximum input-output, the conversion has produced the best results.

- When technical efficiency or social efficiency is 100 percent, but the remaining variables or slack variables of the influencing factors are not 0, the result is referred to as *technical DEA weakly efficient* or *social DEA weakly efficient*. However, in this study, as in the model, there is a relatively large number of variables as to the decision-making units for provinces, so there are no examples of the *DEA weakly efficient* case in which the optimal solution of 1 when only the variable value is 1, the value of other variables are 0, and slack variables and the remaining variables are 0.

- When technical efficiency or social efficiency are less than 100 percent, it is referred to as *technical non-DEA efficient* or *social non-DEA efficient*, indicating that the impact of technology or social factors do not achieve the desired output results, which shows that the applications of technology or society need p. 44 more development. At that point, the remaining variables and slack variables need to be ↳ analyzed to find the root causes of the result being non-DEA efficient.

Analysis shows that in recent years, technical efficiency is slightly higher than social efficiency, that is, the input of the influencing factors is better able to transfer into technology, while the application in the transformation of society is less certain. At the same time, the efficiency levels are increasing each year, and social efficiency levels are increasing faster than technical efficiency levels. In fact, for all provinces, the annual values of efficiency levels have increased in varying degrees, as a result of actively taking measures to enhance the level of ICT technology and the corresponding level of social application.

For technical efficiency of output, the top provinces are Beijing, Jiangsu, Shanghai, Shandong, and Anhui. However, for efficiency of the social output, the top provinces are: Beijing, Shanghai, Guangdong, and Tibet. Thus, it can be seen that Beijing and Shanghai are leading the country in the technological and social application of conversion efficiency; Guangdong province, although a little behind on technical efficiency, has a high efficiency of social applications. Jiangsu, Shandong, Anhui, and other provinces, although they show a good use of government, education, economic, and other factors that raise the level of technology, do not have the same levels of application in this community. This shows that strengthening the ICT infrastructure should be followed by strengthening the community’s understanding and application of technology, in order to get the full benefit of the infrastructure construction.

Analysis of the Relative Position of Technical Efficiency and Social Efficiency

The relationship between the efficiency scores obtained from both the technical efficiency and social efficiency assessments can be explored by means of a *production-profitability matrix* as proposed by Boussofiane, Dyson, and Thanassoulis. We use social efficiency as abscissa and technical efficiency as the longitudinal coordinates, and the average point of technical efficiency and social efficiency as the criteria for the classification, to divide the figure into four quadrants. These useful scores can also be plotted in two dimensions.

The overall trends show that the criteria for classification gradually move to the right, and the technical and social efficiency of the average is increasing on a year-by-year basis. For all provinces, technical and social efficiency correspondingly increased more quickly. Judging from the details, each quadrant has its own characteristics. Two characteristic provinces from two of the four quadrants are selected and compared. The
p. 45 remaining variable ↳ s_1^{*-} shows the output of the economic component; s_2*- refers to the government component, s_3^{*-} to the education component, and s_4^{*-} to the technical component.

Representative province of the first quadrant: Zhejiang. Zhejiang Province has usually been located in the first quadrant (except for 2002); the quadrant is characterized by relatively high technical and social efficiency values. From the point of view of the actual numbers (except for social efficiency in 2002), all of its technical and social efficiency measures are more than 75 percent, and efficiency at all levels is ranked at the top, that is, the technical and social conversion efficiencies of the applications are relatively good. However, through analysis of the remaining variables, it can be seen that there are still factors constraining the efficiency of technical efficiency and social efficiency. From the point of view of technical efficiency, the remaining variable is only s_2^{*-}, showing that the main factor constraining transformation is Government. Government should intensify its policy efforts to invest in the information industry and in technical research and development so as to enhance the rate of technology transfer to achieve the desired results. From the impact factors of social efficiency, although there are some changes in six years, the main variables remaining are s_1^{*-} and s_4^{*-}, which means the major influencing factors are the Economic and Technical aspects. Although the implementation of technology is improving the infrastructure, its availability did not sufficiently increase the consumer economy, so the application of ICT technology alone is not enough. Available technical facilities are idle in this situation, so that the efficiency of social applications has not achieved the best results. These areas should be targeted for future improvements, so that consumers, through more channels and means of using ICT applications, improve their utilization rate of the infrastructure.

Representative province of the third quadrant: Heilongjiang. From the point of view of the actual numbers, during the period 2002 and 2007, Heilongjiang Province, whose application of technical and social efficiency has been lower than the national average, held its technical efficiency basically at the 60 percent level with fluctuations, and an upward trend was not apparent. Social efficiency was even lower, on average only about 45 percent, and except for 2003, when the figure increased a bit more than a basic fluctuation from the average, there was no obvious increase. From the point of view of technical efficiency, the remaining variable is only s_2^{*-}, showing that the main factor constraining transformation is Government; from the components of social efficiency, the main variables for the remaining factors are s_1^{*-} and s_2^{*-}, indicating the major influencing factors are Economic and Government. A major factor impacting technology is the government. Lacking sufficient local technology infrastructure investment results in a low
p. 46 rate of technology transfer, the technology infrastructure is not adequate without the ↳ corresponding technology in support, and that, coupled with the awareness that ICT spending is weak means it cannot bring about a high level of social application. The most urgent tasks are first to strengthen technology-related facilities, and second, within the existing technology level, to strengthen economic development and improve people’s consumption level and their level of awareness of new technologies, in order to fundamentally solve the problem of informatization.

The Future: Challenges and Opportunities

Strengths of Digital Divide Measures

Most current digital divide and e-readiness measures are basically descriptive and comparative. They are good for describing and comparing the status between and within economies and regions, and showing changes over time. The value of this is not to be underestimated. It allows us to compare countries and regions, indicate areas of strengths and weakness, and provide time-series data for interpretation. The use of composite indicators to create ordinally ranked (i.e., ranked in order of priority) sets of economies provides policy makers with relevant directional guidance.

Studies of this type have been widely used internationally, and developed by both governmental and private bodies. Indeed, there are many models of such studies, not all of which are compatible. They have been very useful in drawing attention to the nature and size of the digital divide, and helping focus energies on overcoming it. This is very significant and useful.

Such studies, however, are not as strong at answering the questions of *how* and *why* changes occur; at showing causality; at identifying the relative importance of factors; and at presenting methods to predict the future. They do not always make evident the internal mechanisms of informatization in ways that can be tested and validated. Taking this to the next level requires developing approaches that are both explanatory and predictive, and that will require a different way of thinking about the theory of informatization.

Information technology in society is a complex adaptive system. It needs to be seen comprehensively. It is an integrated system of technologies, networks, industries, content, policies, laws, regulations, and social factors. In a narrow sense, it is linked to productivity and growth; in its broad sense, it shapes the foundation of all social relations. All these aspects need to be seen together.

Emergent systems like information networks often involve such a high degree of causal complexity that traditional modes of analysis may not be adequate. Thus some shy away from this area simply because it is difficult to operationalize. This type of analysis is not economics, but the application of social science
p. 47 statistical tools to an array of economic, social, cultural, ↳ political, and other data to identify relationships and effects. There is much to be gained from rising to this challenge.

Limitations of Digital Divide/E-Readiness Measurements

There is no perfect way to measure either the digital divide or e-readiness, due to the limitations of dealing with proxies (indicators). They both involve a complex analysis of inhomogeneous dynamic adaptive systems. Most existing approaches are subject to certain inherent methodological limitations. It is possible that these approaches have reached the limits of their potential.

Two general critiques are paramount: conceptual imprecision and issues of measurement. Without clear definitions of concepts, scientific discourse is difficult if not impossible, and any resulting theory will lack clarity and precision. This conceptual imprecision (e.g., information, digital divide) may be responsible, to some extent, for the subjective interpretation, implicit value judgment, or ideological claims, which we often find in the literature on the process of social informatization.

According to the *OECD*, in general composite indicators suffer from the following weaknesses, which reduce their analytic value:

- Conceptual imprecision
- Studies not comparable across time or place
- Choice of indicators and weights subjective
- Data/indicators often not comparable
- Computation ad hoc
- Lack of concurrent validity

- Built on different values
- Sensitivity to different weighting and aggregation techniques
- Problem of subjective indicators (e.g., well-being)
- Problem of handling missing data

This approach is built on empirical data, and involves sophisticated mathematical procedures, so it is scientific, but it generally does not proceed in the traditional scientific model of hypothesis testing (observation, hypothesis, experiment, theory, and prediction).

Toward a General Model

The global process of identifying indicators, establishing accepted definitions of indicators, uniformly collecting data, and storing it and making it accessible in a consistent and timely manner is well underway. p. 48 At the same time, it is necessary to be mindful that information needs a theoretical ↳ framework – that is the difference between description and explanation. And the role of experimentation and hypothesis-testing for suggesting causal relationships between variables is a critical factor that cannot be ignored if theory is to be grounded in data linked to the world we live in and be used as a foundation for policy and economic choices.

That said, and with all due recognition of the challenges, difficulties, and limitations of a large-scale and meaningful information metrics program at the national level, useful contributions from these experiences are possible for directing economic and social development in real time. A good example of this is well underway in China.

Exploration of New Models

What is called for is the development of a general theory of informatization, of which the analysis of the digital divide and e-readiness are just particular applications. That is, the digital divide is not informatization. Informatization is an approach to understanding and resolving the digital divide. They are each specific applications of the idea of informatization, designed to answer different questions. Informatization is not just about measuring the digital divide or e-readiness, it is a general theory of the transformation of society by information technologies, of which they are both subsets.

Thus far, the study of informatization shows a lack of a coherent theoretical approach. China has an implicit experimental approach—"stepping on stones to cross the river" (try, evaluate, proceed or retry)—which may ultimately lead to theory from the "bottom up," building theory based on experience through a "natural experiment." However, this kind of inductive approach, while it may allow for some qualified predictions, does not necessarily explain the mechanisms of why or how effects occur (causality), or help us to model informatization in specific circumstances.

Statistical Methodologies

We have presented four statistical approaches to understanding the digital divide. They are very useful tools that add to our collective understanding, and continuously move us toward better models. The discussion in this section does not advocate for or against the use of any particular statistical approach to analyzing data. What is important is the ability to test theory-based hypotheses, discern causal relationships, and permit at least probabilistic predictions. To this end, approaches that reduce the level of subjectivity with respect to selection of indicators and the weighting of factors are to be preferred. The goal is to "let the data speak for itself." This will likely require statistical tools and computer programs capable of analyzing very large amounts of data. And the results may be multidimensional, not just one-dimensional rankings.

p. 49

Going Forward

The study of informatization needs to develop testable hypotheses, and move from an inductive approach to a deductive, experimentalist approach. But evidence of progress (or the opposite) must be measured against initial goals, so first the ultimate goals must be defined. They could be:

- Economic prosperity
- High quality of life
- Harmonious society
- Satisfaction/Happiness

Each of these possible goals is different, but aspects of them are not mutually exclusive. What is critical is that the chosen goal(s) must be carefully defined, and then specified as measurable/quantifiable components. Based on experience, it is possible to formulate initial assumptions to test. Then going forward, as policies or investments are made in IT-related factors, the process will need to:

- Define the goal
- Look at past data
- State the hypothesis to be tested
- Formulate the question related to achieving the goal
- Present a prediction as to the outcome
- Specify the main components (factors) related to the goal
- Quantify the components
- Identify and quantify the relevant indicators
- Apply appropriate mathematical tools (evaluate relevant methodologies)
- Implement
- Measure
- Evaluate
- Revise

The use of testable hypotheses is the path to determining if it is possible to develop a generally applicable theory of informatization, which will apply to both developed and developing nations, indicate the relevance of particular factors under specified conditions, and (based on an analysis of complex cross-causalities) permit a probabilistic prediction about the consequences p. 50 ↳ of particular policy choices. The paradox appears to be that the more extensive and detailed the information used to analyze any particular situation, the more the outcomes will relate only to that particular situation. So any general theory will have to be adapted to the specifics of the case at hand, and the goals of the users. However, it is the development of such a general theory that is the key to further future advancement.

Notes

1. This chapter is a condensation of a much longer paper of the same title presented at the Beyond Broadband Access workshop, which includes extensive tables showing the underlying data and calculations. That paper is available on request from the authors.