

# Judicial Big Data and Big-Data-Based Legal Research in China•

Weimin ZUO\*

Sichuan University Law School

Chanyuan WANG\*\*

Sichuan University Law School

---

## Abstract

The newly established judicial-transparency platforms, like China Judgements Online, have provided access to a new resource—judicial big data—making it possible to conduct empirical, big-data-based legal research. However, as is often the case with new products, these platforms—China Judgements Online, in particular—pose a few problems for big-data-based legal research: insufficient academic depth; immature technical methods; and lack of innovation due to flawed data, strict technical thresholds, and lack of theoretical ambition and ability. In the future, big-data-based legal research should make use of current data resources, continue to promote statistical science and computer science in research, and apply small-data research methods, and in the meanwhile pay attention to the combination of data and theory.

**Keywords:** China Judgements Online, judicial transparency, big data, empirical research, legal scholarship

## 1. INTRODUCTION

Empirical legal research is a new type of legal-research paradigm that originated abroad. It primarily features the analysis and interpretation of data, presenting the mode of experiential research, which is distinguished from the purely theoretical construction or interpretation. The latter, namely the method of traditional legal dogmatics, was and remains the mainstream path of legal research, and almost any other path of research, like sociology of law, has been regarded as extra-legal, and as a “outsider” of legal science, whose rationality was questioned. This refers to the distinction between the internal and external perspective on law, as H. L. A. Hart, in his prestigious *The Concept of Law*, once illustrated: the legal theorist is supposed to adopt an internal, participant’s point of view, whereas the sociologist seemingly approaches law from an external observer’s perspective.<sup>1</sup> However, according to

---

• We would like to thank Prof. Guo Song, Dr Zhang Lianhan, Dr Wang Lusheng, Mo Hao, Tang Qingyu for their generous feedback on this article.

\* Professor at Sichuan University Law School. Correspondence to Weimin Zuo, Sichuan University Law School, Jiang’an Campus, Sichuan University, Shuangliu County, Chengdu, Sichuan. E-mail address: [zuowm@vip.163.com](mailto:zuowm@vip.163.com).

\*\* PhD candidate, Sichuan University Law School.

1. Hart (2012).

some scholars, like Finnish legal scholar, Kaarlo Tuori, legal science has the peculiar dual citizenship that embraces both internal and external perspectives; in other words, it is not just a legal practice, but also a scientific one.<sup>2</sup> And empirical legal research, in our opinion, is a great example of the combination of these two perspectives.<sup>3</sup>

Foreign scholars have thoroughly explored its methodology as well as the practical applications of its methods. In addition, these scholars have conducted empirical research based on big data. Through the relevant articles of the Annual Conference on Empirical Legal Studies<sup>4</sup> and the *Journal of Empirical Legal Studies*,<sup>5</sup> it can be observed that foreign scholars typically use nationwide sample data or full data related to the research theme, or at least local data or large sample data in a relatively broad area or range. With respect to Chinese legal scholarship, although deduction-based legal dogmatics still occupies the mainstream position, the new legal paradigm—namely empirical legal research—has, in recent years, become widely accepted.<sup>6</sup> Articles on empirical legal research have been published<sup>7</sup> and academic seminars, such as the Annual Conference on Chinese Legal Empirical Research<sup>8</sup> and the International Conference on Legal Empirical Studies,<sup>9</sup> have also been held frequently. Given that empirical-research methods have only recently been implemented in the field of law, nationwide and authoritative legal big data has yet to materialize; therefore, domestic researchers tend to collect small sample data in a specific range as research materials.

However, the advent of the data age has generated massive amounts of data that are becoming increasingly available for research purposes. In the judicial domain, thanks to China's vigorous implementation of the concept of judicial transparency in recent years, judicial big data—a new research resource—has emerged. This new resource for empirical research has come into existence due to the development of legal big data. China's unique judicial big data is available on judicial-transparency platforms—with China Judgements Online being the most prominent. In order to fully protect the parties' and public's right to know and right of supervision, the Supreme People's Court (SPC) has disclosed valuable information on judicial-transparency platforms regarding each part of the litigation process, including the trial, execution, broadcast, and adjudication. Among the available platforms, given the types of data it gleans from a significant volume of adjudicative documents, the most detailed and authentic reflection of judicial practice in China can be found on China Judgements Online.

The arrival of big data not only provides new opportunities for empirical legal research in China, but will also likely promote further development of the empirical-research process. For example, the availability of big data has significantly enriched the base

---

2. Tuori (2008), pp. 188–9.

3. We will discuss the two perspectives in more detail on Section 4.

4. The Annual Conference on Empirical Legal Studies of America has been held for the past 13 years, and the 14th session was held in November 2019. CELS 2019 HOME (2019).

5. The *Journal of Empirical Legal Studies*, edited and published by Cornell University, is currently the most important empirical-research publication in English academia. Wiley Online Library (2019).

6. Zuo (2018c), pp. 45–61.

7. See Zuo (2017a); Zuo (2017b); Zuo (2018b); Tian & Lu (2017).

8. The 3rd Annual Conference on Chinese Legal Empirical Research was held in Yunnan University on 23 June 2018. Law School Yunnan University (2018).

9. The International Conference on Legal Empirical Research, organized by Sichuan University, was held in Chengdu from 19 to 21 July 2019. China Law Review (2019).

materials of empirical study—thereby expanding the scope and framework of research topics. Likewise, the technical requirements of big-data processing will encourage the innovation of empirical-research methods, making them more diversified and scientific. With these advances, the research findings of big-data research will be more precise and objective. To reach that state of precision, various aspects of current big-data legal research based on China Judgements Online will require improvement. To make the necessary improvements, researchers need to focus on the problems and actively develop and implement timely solutions. In this regard, we propose the following suggestions in the discussion below.

## 2. THE ORIGIN OF JUDICIAL BIG DATA

The first elements that should be examined are the timing and location of the emergence of legal big data in China. Legal big data is founded upon traditional legal data, which include judicial statistical data produced and published by the government, which are presented in a digitized and structured form. Such data consist of the statistical yearbook, legal yearbook, work report, etc. that are provided by central and local judicial organs, statistical departments, the SPC, and other levels of people's courts. Yi et al.'s exploration of documents and websites that recorded national and local courts' judicial statistical data in 2014 resulted in the conclusion that such data are besieged with "problems of incomplete data, scattered public channels, inconsistent statistical requirements, [and] discontinuous and untimely publicizing patterns, which make it impossible for the data to form mutually interconnected data networks that can be compared with each other."<sup>10</sup>

In addition to the above shortcomings, the most significant problem with the judicial statistical data used to conduct data research is the lack of public access to the materials (i.e. judicial documents) used to calculate the statistics. Without access to the mesoscopic and microcosmic individual data sources that support the published statistical data, researchers are unable to examine the methods used to produce the statistical data. Accordingly, the official statistical data can be used for only macro and rough trend analysis. Moreover, given the inevitability of political, social governance and judicial management influences, it is impossible to achieve complete neutrality and objectivity in the officially produced data. Consequently, these shortcomings make the "official" data inadequate for use in academic research; in other words, traditional statistical data cannot fulfil the data conditions—namely objectivity, accuracy, and specificity—necessary for academic research. Apart from judicial statistical data, the SPC has also released limited case information—the basic facts, decisions, and reasons—in Guiding Cases. In comparison with statistical data, the judicial data disclosed in Guiding Cases more closely approximate the original case information. Nevertheless, the information presents a mere overview of the case; therefore, there is little information available for examination or use by researchers. Furthermore, given that the SPC has posted only 100+ cases so far, the total number of cases included on Guiding Cases is insufficient for proper data research.

The emergence of legal big data is directly related to reforms implemented by the SPC. Based on the judicial-transparency concept emphasized by the 18th Central Committee of

---

10. Yi et al. (2017), p. 69.

the Communist Party of China (CPC), the SPC set up four Internet-based platforms for releasing information, namely China Judicial Process Information Online, China Law Execution Information Online, China Trials Broadcast Online, and, the most well-known and influential, China Judgements Online. Ergo, due to the relative ubiquity of Internet access, for the first time, the Chinese public has access to massive amounts of adjudicative documents.

Establishing China Judgements Online is the SPC's most notable achievement in its efforts to promote online public access to adjudicative documents. On 25 March 2009, the SPC issued its Third Five-Year Reform Plan of the People's Courts (2009–2013), proposing to conduct "research on establishing [an] online publication system for judicial documents and [an] online searching system for case execution information." On 8 December 2009, the SPC issued its Six Provisions on Judicial Transparency, clearly stipulating that, except for mediation cases and cases involving state secrets, juvenile delinquency, personal privacy, and other cases unsuitable for disclosure, adjudicative documents of the people's courts could be released to the public on the Internet. This was the first time that the SPC had adopted a normative document to set forth parameters for the Internet-publication of judicial documents by courts at all levels. On 21 November 2010, the SPC promulgated its Provisions of the Supreme People's Court on the Issuance of Judgments on the Internet,<sup>11</sup> making specific provisions on the principles, scope, and procedures with which the people's courts could disclose judgments online. With its progressive development and implementation, China Judgements Online has steadily carried out the intent of those provisions since their enactment.

On 30 December 2011, the SPC held the first meeting for the Leading Group of Judicial Transparency Work, announcing the need to formulate plans to establish a unified nationwide website for online judgments. A key part of these plans for judicial transparency was the creation of China Judgements Online. More than two years later, on 8 May 2013, the SPC held a seminar on judicial transparency in Liuzhou, Guangxi Province. At this seminar, the SPC sought advice from some courts on the design of a nationwide unified website. Soon after, China Judgements Online entered its design stage.

Capturing the design initiation, the SPC's Leading Party Members' Group discussed and adopted the Report on Establishing China Judgements Online on 22 May 2013. This report clarified that the SPC would establish a platform called "China Judgements Online" to release effective adjudicative documents from courts at all levels. Accordingly, China Judgements Online then entered its implementation stage. On 28 June 2013, the SPC released its first 50 enforceable judgments on China Judgements Online. On 2 July 2013, the SPC enacted the Interim Measures on Releasing Judgments Document Online; this is particularly noteworthy because it was the first institutional document adopted by the SPC to specifically regulate its own procedure for releasing judicial documents on the Internet. According to these measures, all adjudicative documents produced by the SPC, including judgments, rulings, and decisions—except those concerning national secrets, trade secrets, or personal privacy—would continue to be publicized on its official website. More importantly, the Decision of the C.P.C. Central Committee on Several Important Issues of Comprehensively Deepening Reform, adopted in the Third Plenary

---

11. This document has been superseded by the Provisions of the Supreme People's Court on the Issuance of Judgments on the Internet by the People's Courts (2013).

Session of the 18th C.P.C. Central Committee, mentioned that the government would “increase the persuasiveness of legal instruments and press ahead with publicizing court judgments that have come into force,” thereby establishing essential political grounds for online access to adjudicative documents.

After the SPC released its first set of judgments, other courts joined the effort to promote the disclosure of adjudicative documents. On 13 November 2013, the 1595th meeting of the SPC’s Judicial Committee adopted its Provisions of the Supreme People’s Court on the Issuance of Judgments on the Internet by the People’s Courts (2013),<sup>12</sup> regulating that, from 1 January 2014, the people’s courts must publicize effective adjudicative documents on China Judgements Online. This was the first time that the SPC had instituted comprehensive regulations regarding the Internet release of judicial documents—in the form of judicial interpretation—by all levels of people’s courts. In response to the provisions of this judicial interpretation, on 31 December 2013, the people’s courts at all four levels began to upload enforceable judicial documents on China Judgements Online. Since then, China Judgements Online has served as the nationwide platform on which the people’s courts release their effective judicial documents. By June 2015, all the people’s courts had uploaded enforceable judicial documents online, achieving full coverage of case types and courts.<sup>13</sup> The documents uploaded by the courts included verdicts, decisions, notices, and certain parts of mediation documents.

The abundant nationwide data made available by the establishment of China Judgements Online differ greatly from the data used by traditional empirical research in terms of magnitude and breadth. By 23 August 2017, there had been 10 billion visits to China Judgements Online and nearly 32.47 million judicial documents had been publicized on the website. By 12 August 2019, the total number of judicial documents on China Judgements Online had exceeded 74.39 million and the number of visits had climbed to more than 31.2 billion. In other words, both the number of documents and the number of visits doubled in less than two years.

There are three other platforms in addition to China Judgements Online. China Judicial Process Information Online is a platform for parties and their agents to inquire about cases, contact judges, and accept electronic delivery. The information it provides to the public is limited to court address, court notice, member of the judicial committee, etc.—all of which offer little research value. In contrast, the information provided on China Law Execution Information Online and China Trials Broadcast Online is predominantly open to the public. Although the information published on these two platforms is not as comprehensive as the information on China Judgements Online, they may also become potential mining targets for big-data-based research, functioning as supplementary data sources. The China Law Execution Information Online publicizes information on dishonest persons subject to enforcement by all courts. On 1 October 2013, Several Provisions of the SPC on Announcement of the List of Dishonest Persons Subject to Enforcement took effect. After its implementation, courts at all levels started including individuals on the list of

---

12. This judicial interpretation has been annulled by the Decision of the Supreme People’s Court on Abolishing Part of the Judicial Interpretations (The Thirteenth Batch). On 25 July 2016, the 1689th meeting of the SPC’s Judicial Committee adopted the newest Provisions of the Supreme People’s Court on the Issuance of Judgments on the Internet by the People’s Courts (2016).

13. CHINATRIAL (2017).

dishonest persons subject to enforcement. On 24 October 2013, China Law Execution Information Online was opened to the public. As of 18 August 2019, China Law Execution Information Online had publicized a list of 14.95 million cases of dishonest persons subject to enforcement.

China Trials Broadcast Online has permitted all courts to broadcast trials. On 14 April 2016, the SPC promulgated the revised Rules of the Court, providing that, for three types of court trials conducted in accordance with the law, people's courts may publicize live broadcast or recorded broadcast on television, the Internet, or other public media. Those three types are trials that (1) have received significant public attention; (2) will have a great social impact; (3) will be significant to the rule of law with regard to publicity and education. In July 2016, the SPC provided an example of an open court trial and, only months later, on 27 September 2016, China Trials Broadcast Online was officially launched. Nearly three years later, the SPC published more than 3,000 live broadcasts on the site. At present, the total number of live broadcasts nationwide has reached more than 420 million, with more than 18.3 billion visits.

Compared to traditional data resources, the above judicial-transparency platforms, especially China Judgements Online, possess important and beneficial features—most notably, their massive data troves. Before China Judgements Online became operational, the judicial data released in the form of work reports by the SPC and Supreme People's Procuratorate in Guiding Cases, legal yearbooks, etc. offered a very limited glimpse into the judicial practice in China. In contrast, the information disclosed by China Judgements Online is unprecedented with regard to both scale and category. In accordance with the SPC's provisions, information on all cases eligible for publication shall be uploaded online within seven working days after they come into force and, in principle, this case information is made available to everyone. This degree of publication is revolutionary for a number of reasons. First, given the release of current and past adjudicative documents, the data volume of China Judgements Online has been increasing rapidly. The resulting information accessibility is on a par with and even surpasses the most accessible judicial data systems worldwide.<sup>14</sup> When China Judgements Online was first launched in 2014, the number of uploaded documents totalled nearly 5.58 million. In 2015, the number reached nearly 9 million and, in 2016, the site contained nearly 10 million documents.<sup>15</sup> As of August 2019, the data volume had exceeded 74.4 million documents. China Judgements Online has now become the largest publicizing platform for adjudicative documents in the world. This unprecedented number of adjudicative documents provides a broad and comprehensive academic resource for future empirical research, which is based on data mining. Such research will then be able to reflect an accurate picture of Chinese justice, especially at the trial stage.

Second, in addition to increasing the volume of data, China Judgements Online has enhanced the richness and specificity of data content. In contrast to traditional legal data, which are shallow and general, judicial-transparency platforms release detailed texts and videos that serve as records of both the processes and decisions involved in individual cases.

---

14. European countries have not yet implemented a comprehensive online collection of all adjudicative documents. In most countries, all adjudicative documents of the Supreme Court are available on the Internet and the adjudicative documents of the high court are partially open to the public, while the courts of first instance publicize only a small number of documents. See He et al. (2013), pp. 51–6.

15. Ma, Yu & He (2016), p. 203.

The detailed presentations of the primary references in the cases (i.e. judgments and decisions) render it possible for multi-angle and in-depth data research.

Third, the online platform has improved the objectivity and non-reactivity<sup>16</sup> of the data. Compared to statistical yearbooks, work reports, and other structured data that have undergone “fine processing” that reflect the value preferences of their data-publishing bodies,<sup>17</sup> data on China Judgements Online are original case texts, uploaded directly by the trial court in accordance with legal provisions. Therefore, the value preference of data-publishing bodies has been diluted and the research based on the unaltered data can maintain its objective nature. In addition, once the adjudicative document has been uploaded, researchers can freely choose to download it and its content or form will not change according to the researcher’s observation. This means that data on China Judgements Online are also non-reactive.

Fourth, the platform generates improved data-mining opportunities. The information contained in China Judgements Online, China Trials Broadcast Online, and China Law Execution Information Online is not strictly data, or quantitative data as some researchers call it,<sup>18</sup> which makes it difficult for researchers to perform mathematical statistics and analysis. However, researchers can use data-science methods such as labelling and coding to convert the case information contained in the documents and trial videos into quantitative data for research.

Lastly, the platform permits the personalization of data collection and analysis. The mesoscopic and microcosmic data contained in the current judicial-transparency platforms and their accessibility make it possible for researchers to select the scope and method of data collection according to their own research ideas and preferences. In this way, the researchers can obtain new legal data that are different from official data and perspectives, thus allowing the design of a personalized and distinctive research model.

In our opinion, the new type of judicial-transparency data found on China Judgements Online constitutes the headstream of legal big data in contemporary China.<sup>19</sup> The launch of these platforms, which has propelled the Sunshine Justice, also created unprecedented opportunities for the legal research of Chinese data. The extensive development of Chinese big-data-based legal research was initiated by the unified online access to judicial documents. Before adjudicative documents were released online, big-data-based legal research was essentially non-existent in China, and empirical legal research was based mainly on small data—that is, data collected by researchers on a local scale or in a specific area to conduct “workshop-styled” research. The emergence of China Judgements Online, the nationwide, public, and detailed big-data legal platform, has enabled researchers to use scientific methods in statistical and computational science, transforming massive documents into data and obtaining legal big data that can be differentiated from official

---

16. Non-reactivity means participants will not change their behaviour when they know they are being observed by the researchers.

17. Zuo (2018c), p. 143.

18. Chang & Cheng (2018), p. 75.

19. To note, in addition to official data platforms such as China Judgements Online, there are, at present, various case-retrieval tools such as Faxon, Wolters Kluwer, Pkulaw, CaseShare, Itslaw, LawFAQ, etc., which are also based on the information of adjudicative documents. However, compared with the official platforms, these tools have problems such as lack of authority, an insufficient number of documents, and high fees, which make them difficult to use as the main source of big-data-based research for legal researchers.



big data. This provides unlimited possibilities for Chinese legal studies and enables empirical legal research to truly move towards big-data empirical legal research. To clarify, a rich sample size alone does not necessarily make an empirical study a big-data study; rather, the uniqueness of big-data research is rooted in the fact that relevant research methods and techniques have been revolutionarily changed by the innate characteristics of big data. In other words, a study in which data-processing methods matching the features of big data are applied can be accurately referred to as a real big-data study.

### 3. ANALYSIS OF BIG-DATA RESEARCH BASED ON CHINA JUDGEMENTS ONLINE

It should be noted that, in recent years, attempts have been made to use big data or a large amount of data from platforms such as China Judgements Online directly for conducting legal research.<sup>20</sup> Nevertheless, the ongoing legal research on big data in China is still in its initial and exploratory stage. Generally speaking, there are several shortcomings. First, the existing research results more closely approximate popular science and are, therefore, not sufficiently academic as a whole. Some of the research focuses on heuristics and there are peripheral discussions on how to use big data to carry out legal research, rather than using big data as a means of research. For example, some scholars examine the ethical norms faced in legal big data<sup>21</sup> and some have provided instructive insight on how to conduct big-data-based legal research.<sup>22</sup> Although there are some studies using big data, there is an overarching absence of sophisticated data-processing models and rigorous theoretical interpretation systems. Most studies contain only simple categorization statistics of various data, based on which researchers raise questions and develop solutions.<sup>23</sup> There are few studies that can present deep analysis and offer theoretical explanations of the general phenomena shown by big data. In addition, there exist studies that are dressed with the big-data “outerwear” but lack the necessary internal elements as they simply list and describe the big-data phenomena. Moreover, some studies even lack a systematic understanding of basic empirical-research methods in criminal justice.<sup>24</sup>

---

20. E.g. Zuo and Zhang conducted an analysis of criminal-defence rates based on 54,409 criminal judgments of first instance in Sichuan Province from 2015 to 2016. See Zuo & Zhang (2019), pp. 167–89. Feng used the 135 judgments obtained by searching the keyword “constitution” in the “legal basis” field on China Judgements Online as samples and analyzed the phenomenon of constitution invoking and its functions in judgments in China. See Feng (2017), pp. 44–59. Qiao et al. organized and analyzed a total of 586 administrative judgments on expropriation and demolition made by the Higher People’s Courts in China from 2014 to 2015, and conducted a comparative analysis of the roles of all three elements (public interest, reasonable compensation, and due process) of the levy cases in restraining the power of local governments. See Qiao & Mao (2018), pp. 68–90.

21. See Xu (2017), pp. 130–49; Gu (2017), pp. 161–7.

22. Bai discussed the possibility and limitations of using big data for adjudicative prediction and the question of how to take samples scientifically in the era of big data. See Bai (2017), pp. 95–100; Bai (2015), pp. 29–35. Hu discussed the “deepening of legal research methods” in the era of big data. See Hu (2015), pp. 108–13. Zhang studied how legal research in the era of big data should “develop interdisciplinary research and applications.” See Zhang (2016), pp. 52–61.

23. In a study sampling articles related to legal empirical research from cnki.net, the statistics calculated by Lei on the topics of empirical research from 2000 to 2016 and the hotspots from 2014 to 2016 show that “countermeasures and improvement” are always important topics. Lei (2017), p. 107.

24. Xiong (2012), p. 6.



Second, the techniques used in current big-data research remain relatively shallow. First of all, the data being collected and mined in the adjudicative documents by the majority of studies are the obvious and “dominant” data that are captured at a specific location in a adjudicative document and can, therefore, be easily extracted (e.g. whether the party has a lawyer, the party’s education level, place of origin, age, coercive measures taken, etc.).<sup>25</sup> The extraction of this kind of information requires the use of the regular expression or the natural language processing (NLP) technique. With regard to data that are more difficult to extract, effective techniques for obtaining hidden and “recessive” data, such as the parties’ claims, evidence, the court’s reasoning, and judgments, have yet to be developed. Instead, researchers are still using methods suitable for small data, namely calculating statistics by hand to mine data.<sup>26</sup> This outdated data-mining method, which is time-consuming and demanding, greatly increases the cost of data research and can only be applied to information mining of small data samples. Given the massive amounts of materials and data in the era of big data, it is difficult for researchers to personally review, count, and analyze data file by file. Moreover, researchers’ use of data-analysis methods has also become monotonous. When facing the collation and analysis of big data, most legal researchers are unable to scientifically and skilfully use methods such as mathematical statistics to conduct quantitative analysis of a problem in a statistical sense, let alone the construction of mathematical models in research.<sup>27</sup> Empirical legal research as a whole still employs descriptive statistical methods such as average number, frequency, and variance as primary tools,<sup>28</sup> which is far from the current technical-research level of other disciplines such as economics and management.<sup>29</sup> Descriptive research plays an important role in depicting the characteristics of empirical phenomena; however, it does not qualify as in-depth research aiming at constructing correlations or even causality between things. Nevertheless, empirical legal research, including big-data-based legal research, is much

---

25. E.g. a previous study of criminal-defence rates examined whether the client had a defence lawyer, whether the defence was entrusted or designated, the number of defence lawyers, and the court level of the case. See Zuo & Zhang, *supra* note 20, pp. 167–89. In Wang’s study, the focus is on the coercive measures of the case, the criminal-defence situation, and the average number of days for the trial. See Wang (2018), pp. 124–47.

26. E.g. when Zhang used the 1,545 adjudicative documents as a sample to study the reference effectiveness of Guiding Cases, he did statistics file by file on indexes like the actual form of effectiveness, the basic content with reference effectiveness, and the number of Guiding Cases referred to. Zhang (2018a), pp. 119–35. In Jin and Shao’s empirical study on environmental-infringement cases, they counted and collected by hand the important case information such as the focus of the dispute and the distribution of the burden of proof. Jin & Shao (2018), pp. 56–9.

27. Zuo, *supra* note 6, p. 51.

28. After statistical analysis of the full-text articles from *Social Sciences in China* in 2010–14, which is one of the three major journals of law, Shen found that, from the perspective of data-processing methods, 90% of legal articles use methods of descriptive statistics such as mean number, frequency, and variance; and 30% use regression analysis, correlation analysis, hypothesis testing, etc.; and none uses advanced mathematical methods such as model calculations. See Shen (2015), p. 103.

29. In a study of 1,126 papers published by authoritative domestic journals in the field of economics from 2012 to 2014, Wang et al. found that the wide application of mathematical methods and mathematical models in research has become an important trend of economic research. Statistical analysis showed that only 165 papers in the samples did not use any mathematical methods, accounting for roughly 15% of the total. Wang & Du (2015), pp. 140–53. After extracting and conducting bibliometrical analysis on 858 papers from 1982 to the end of 2012, management scholars Fan and Lou found that, in the field of public management, the research methods range from deductive to inductive, from theory to empirical, and from qualitative to quantitative; the research normativity ranges from non-standard to gradual normalization; the data-analysis methods range from simple to complex; the statistical variables range from unitary to multivariate; and the statistical methods range from manual to informatization. Fan & Lou (2013), pp. 98–9. In addition, some advanced quantitative-research methods such as path analysis, neural network, data envelopment analysis, and network topology analysis have been applied to public management research. *Ibid.*

more than a tool for describing phenomena, since it also undertakes the missions of revealing the law of the operation process of positive law and explaining the correlations between or even causal relations behind facts.

Third, most research topics and findings lack innovation. The data in the current research often serve as a rationale to support or refute theories, which is far from the core of reasoning. Accordingly, current research typically adopts a purely instrumental application of data—that is, using data-analysis results to verify existing propositions. However, because most research findings are restricted within the framework of the existing theory, the data cannot play a guiding role and the conclusions are expected. Thus, given that such research rarely uncovers objective facts that differ from traditional cognition and rarely addresses thought-provoking problems and theoretical reasoning, it is altogether unlikely that such research would construct a new theory supported by data. Although some domestic scholars have made a few valuable attempts in topic selection or theoretical innovation, such as Wang, who used 3.03 million judgments as a sample to study the symbolic legislation phenomenon in criminal proceedings,<sup>30</sup> and Taiwanese scholar Chang, who found that, in the field of property law, China is closest to the Russian legal family and second-closest to the German legal family on the basis of quantitative legal-research methodology,<sup>31</sup> similar studies are still rare in big-data-based research.

Why do such shortcomings exist? We suspect there are three main reasons: defects in the data, the strict threshold of technology, and the lack of theoretical ambition and abilities. As for data defects, we believe they are the result of the limitation of the data-publicizing channel and the limitation of the data-publicizing range. With regard to channel, China Judgements Online (including China Trials Broadcast Online and China Law Execution Information Online) is a court-centred information-disclosure platform; thus, as a sole publicizing channel, it reflects an incomplete picture of judicial practice. The information released by China Judgements Online is typically either litigation or trial information. The remaining key procedural processes, such as the investigation process by the police, the prosecution process by the prosecutor's office, and the processes before and after the court trial, are not formally or informally included in publically accessible text records, let alone a digitalized database of written documents. Similarly, China Trial Broadcast Online releases only the video data of the trial phase of cases, while China Law Execution Information Online releases only the identifying information of defaulters who have failed to fulfil court orders.

With regard to the limitation of the data-publicizing range, under the framework of the sole publicizing channel and the specific publicizing phase, there remains a lack of data generated from two distinct categories of bias: systematic bias and random bias. Systematic bias refers to the adjudicative documents that are not publicized, such as documents involving state secrets or juvenile delinquency, due to provisions of law.<sup>32</sup> A study by

---

30. Wang, *supra* note 25, pp. 124–47.

31. Chang (2019a), pp. 81–101.

32. Art. 4 of Provisions of the Supreme People's Court on the Issuance of Judgments on Internet by the People's Courts (2016): "Under any of the following circumstances, a judgment rendered by a people's court shall not be issued on the Internet: (1) it involves any state secret; (2) it involves any juvenile delinquency; (3) the case is closed by mediation or the effect of a people's mediation agreement is confirmed; unless it is actually necessary to disclose the judgment for the purpose of protecting the state interests, public interests, and the lawful rights and interests of others;

Zuo et al. on criminal-defence rates found that the juvenile crimes may make up the largest percentage of criminal cases that are not publicized on the Internet by law, while the number of cases involving state secrets, mediation cases, and other cases that are not suitable for publication on the Internet make up a small percentage.<sup>33</sup> According to previous data from the *Law Yearbook of China*, juvenile criminals account for nearly 5–10% of the total number of criminals.<sup>34</sup> Thus, because online adjudicative judgments do not cover all types of cases, there is a systematic bias regarding the information made public.

In contrast, random bias refers to the adjudicative documents that are permitted by law to be publicized online but, for various reasons, are not actually publicized. These include appeals and protest cases, and, although they are not yet posted online, they are counted in the number of cases closed that year. Random bias may also include late cases—namely those that are not publicized in time due to work delay.<sup>35</sup> Thus, the degree of random bias is closely related to the implementation of work by individual courts and the staff responsible for data uploading. Due to the above reasons, in terms of the overall number of adjudicative documents, what is publicized by China Judgements Online does not fully capture the features of a full sample, as the number of judicial documents publicized differs greatly from the number of adjudicative documents from closed cases. The problem of missing data is fairly serious. According to the statistics calculated by Ma et al. on online adjudicative documents between 2014 and 2015, as distinguished by provinces, the highest proportion of released adjudicative documents from actual closed cases was 78.14% (Shaanxi); the lowest proportion was 15.17% (Tibet); and the proportion for SPC was 46.13% (roughly the same as the national scale).<sup>36</sup> As of 6 July 2019, the courts in Sichuan Province had a total of 1.43 million adjudicative documents released online in 2017 and 2018, and, according to the work reports of the Sichuan Higher People's Court, the total number of cases closed within the province in 2017 and 2018 was 2.16 million.<sup>37</sup> In sum, the legal big data in China encompasses a large amount of data, official and semi-structured, that captures a limited and angle-specific data set in the national legal domain. Even with its deficiencies, this significant volume of data may be the “big data in reality.” Thus, researchers must, at the outset, understand that their research can only reflect a limited picture of the legal and judicial practice in China, as it is impracticable to apply a sample condition extracted from the limited adjudicative documents online to a target population. The deficiency of the online adjudicative documents in terms of quantity, region, case type, etc. also means that a data

---

(4) it involves divorce proceedings or child support and guardianship of under-aged children; and (5) any other circumstance under which the people's court holds it inappropriate to issue the judgment on the Internet.”

33. Zuo & Zhang, *supra* note 20, p. 169.

34. The selection of this value has been confirmed by multiple sources: (1) According to the *Law Yearbook of China*, in 2004–13, the proportion of criminals under the age of 18 in the whole country accounted for 5–10% of the total number of criminals in the year. (2) A study pointed out that the proportion of juveniles in a certain province of total criminals had dropped from 9.99% in 2009 to 5.00% in the first half of 2013. Yao (2015), p. 112. (3) Based on an interview with authorities in Sichuan Province, in 2014, the number of juvenile cases in C city of Sichuan Province was 1,416 and the number of juvenile defendants was 1,863. This accounted for nearly 11% as to 12,183 criminal cases of first instance and 16,585 defendants in the city.

35. Zuo & Zhang, *supra* note 20, p. 169.

36. Ma, Yu & He, *supra* note 15, p. 208.

37. According to the work reports of the higher court of Sichuan Province, the number of cases closed by all the courts in Sichuan was 1.04 million in 2017 and 1.13 million in 2018, and the total number of cases in these two years was 2.17 million.

study limited in a specific range may not be able to obtain representative full-sample data. Moreover, previous data-research experience has revealed to us that unrepresentative data without adjustment are highly likely to result in erroneous conclusions.<sup>38</sup>

Going back to the second reason for the big-data shortcomings in China, for legal researchers, there exist innate technical thresholds for data collecting, cleaning, processing, and analyzing. The core of big-data legal research lies in the value mining and processing of massive data. The abilities of researchers to grasp and apply related technical methods largely determine the depth and level of research, and the weakness in data application may lead to superficial or even wrong research conclusions. It can be said that the necessary steps in dealing with data have set up innate technical thresholds for big-data-based legal research, but traditional legal researchers find it difficult to master the new techniques of statistical and computer science. Due to the large volumes of adjudicative documents on the website, it is impossible for researchers to download documents manually, file by file. Thus, given the need to collect large amounts of data, researchers have begun to use methods such as crawler software to obtain data. Nevertheless, since obtaining data from the China Judgements Online system using web-crawler systems causes website overload, which affects access by normal users, the special operation and maintenance support team established by the SPC implemented a verification-code process to enact an anti-crawling function of the system software. The inevitable strengthening of the anti-crawling technology will make it more and more difficult for researchers to actualize the process of quickly obtaining large amounts of data.<sup>39</sup>

The acquisition of documents is merely the starting point of big-data research. The data obtained from the adjudicative documents by the crawling software are typically unlabeled and unstructured. The data often contain a lot of “dirty data,” such as duplicated or blank documents, that must be cleaned. This cleaning process involves filtering or modifying incomplete, erroneous, or duplicate data and is a necessary part of accurate and effective data mining. Data cleaning can be achieved through deleting or ignoring missing values—the easiest method that is accompanied by a loss in sample size and weak statistical power. More elaborate data-cleaning methods include, for example, interpolation, mean-value interpolation, and outlier analysis.<sup>40</sup> The proper application of these methods requires researchers to have a statistical background.

Once cleaned, the data must undergo an additional transformation into structured data. Certain variables in the adjudicative documents, such as information on the parties, information on the defenders, the year of judgment, and the level of trial, are fairly clear given their uniform textual expression, while other valuable information may not be easily captured, as personalized expressions of other variables may exist. Therefore, researchers need to first define and code the variables in a detailed way. For example, in an empirical

---

38. A well-known example is the prediction of the US presidential election by Literary Digest in 1936. After sending votes to 10 million people, Literary Digest took the sample of the 2.4 million votes returned by participants. Under the circumstances that these votes had not been adjusted, weighted, or interpreted, Literary Digest predicted that Alf Landon would beat Franklin Roosevelt, the incumbent. However, Roosevelt overwhelmingly defeated Landon. The reason for this false prediction is that, during the polling process of Literary Digest, some participants were systematically biased and their samples did not represent the target population.

39. Beijing Youth Daily (2019).

40. Zhao, Bian & Cong (2017), pp. 222–4.

study in which Cheng et al. researched labour-contract disputes, they paid special attention to a key variable—whether the labourer won. The study analyzed and listed five potential judgment results when the labourer seeks relief, determined that four of the five potential judgments resulted in a win for the labourer, and used that determination in statistical calculations.<sup>41</sup>

Next, researchers must analyze the structured content of the processed data. In this step, most researchers still predominantly use descriptive data analysis—a process that involves empirically describing the features of the research object and adopting traditional speculative deduction for the analysis of causality. Few researchers have been able to use statistical software and statistical analysis methods such as regression-discontinuity design, a difference-in-differences model, and matching, to conduct an accurate quantitative analysis of data resources. More in-depth, big-data-based legal research also involves the use of machine-learning and algorithm applications. For example, through data-correlation analysis, researchers may find a correlation between data in a large amount of scattered data and then form the data into a data set in order to depict a developing pattern or trend of a certain thing or event.

As for the third reason for big-data shortcomings in China, domestic researchers have not fully demonstrated the theoretical ambitions and abilities necessary to conduct high-quality big-data research. Given that the data are incomplete and certain technical tools are absent, domestic research at this time is primarily theory-oriented; that is to say, researchers apply data as instruments to validate existing theories without making an effort to use data research to discover new phenomena or create new theories, even though these possibilities are achievable in big-data research. The origination of empirical legal study once bridged the gap between the discourse of traditional legal dogmatics and the context of judicial practice; now, the emergence of judicial big-data resources make the prospect in reality clearer and more meticulous. In the face of massive and free big-data resources, we are more likely to obtain new information and knowledge, greatly expand the scope and field of legal research, and produce data-oriented academic research. However, domestic researchers are stuck in the initial phase of big-data use. They may have not yet realized the value of big data to be mined. Or perhaps they have failed to master the scientific methods of processing big data and, therefore, lack the confidence and abilities to re-examine legal practice and challenge authoritative legal theory with big data. This relative shortage of ambition and ability makes it hard for the domestic big-data research community to escape its current predicament.

#### 4. HOW TO CARRY OUT RESEARCH BETTER USING JUDICIAL BIG DATA

Despite these problems, there is no doubt that big-data-based legal research will become the legal-research paradigm of the future, which is why scholars must stick firmly to this path. Specifically, future big-data-based legal research should work to advance the following four features.

---

41. Cheng & Ke (2018), pp. 13–4.

First, researchers should accept the objective defects of the existing data resources and make the best use of the limited data available. Given the various objective restrictions, it is far too ambitious at this time to expect access to big data that contains a full sample of domestic justice; in other words, Chinese legal researchers may be faced with large amounts of data instead of full data for a long time. Nevertheless, a large amount of data is still important material for legal research; thus, it is worthy of great attention and full use. In other words, although samples that researchers can obtain through platforms such as China Judgements Online cannot directly represent the attributes of the target population, such a lack of representation may be ideal of certain research goals. For example, when conducting research based on China Judgements Online, under the premise of knowing the data-deviation situation, researchers can properly narrow the range of research and limit the research objects to ensure that relatively complete and representative data in a certain field, a certain region, or a certain category can be collected. Thus, full-sample data research focusing on specific regions, types, and problems can be carried out.

To demonstrate, divorce disputes are settled largely by means of mediation. Due to the fact that mediation documents are not usually publicized, big-data-mining reports on marriage need to be treated with caution. Moreover, even if the research focuses on a specific range, researchers cannot obtain the perfect data for the research goal, as systematic and random biases cannot be eliminated from any data sample. However, if the incomplete data are adjusted and corrected by certain technical means, the problem caused by the lack of data can be effectively solved by data-processing and analyzing methods reasonably designed by researchers. First, researchers can generalize the results of the study by comparing data within the sample to the target population. For example, researchers Wang et al. once used an obviously non-random sample—a non-probability sample consisting of American users of Xbox (a Microsoft game console), among which males made up 93% and the young—between 18 and 29 years old—made up 69%. Researchers adjusted the non-random-sampling process during the evaluation of this seemingly unsatisfactory sample with the help of the post-stratification technique, which involves the grouping of the sample by the auxiliary information of the target population and then weighing the results. In a nutshell, researchers divided the sample population into groups based on response tendency (e.g. if all males have the same response tendency and all females have the same response tendency, then post-stratification based on gender can produce unbiased assessment conclusions). In the end, this study correctly predicted the result of the 2012 US presidential election.<sup>42</sup> Second, researchers can also integrate multiple data sources to fill in the gaps between data. Specifically, when the complete data required by the study are not available through China Judgements Online alone, researchers can consider turning to other data sources, such as traditional judicial statistics, unofficial case-retrieval tools, and databases established by unofficial organizations. For example, when facing the systematic bias on criminal judgments online, Zuo et al. looked up the statistical results of non-Internet cases from data sources other than China Judgements Online, and then took into account the weighted estimations of non-Internet cases in addition to the known number of online cases. After formula calculations, Zuo et al. finally obtained the overall defence rate in the province that

---

42. Wang et al. (2015), pp. 980–91.



they were studying.<sup>43</sup> Foreign scholars Ansolabehere and Hersh applied a more complex and detailed data-integration process in their study. They linked the voting records from Catalist with the social-survey data into a larger primary data source, and then analyzed the correlation between voting behaviour and voters' attributes based on the primary data sources. Both of these two basic underlying data sources were indispensable for the study.<sup>44</sup>

Second, researchers should continue to promote the deep applications of statistical and computer science in their studies. In terms of data collection, mining, sorting, and analyzing, legal research requires mature statistical methods and data-science methods. As mentioned above, in the process of data selection and collection, researchers can make full use of statistical tools to adjust the non-full-sample data, in order to restore the full-sample data estimated with maximum precision and also assess the validity and authenticity of big data or large amounts of data. In the data-mining phrase, regular expression is still the most widely used method. This method shows strong accuracy in dealing with normed expressions in adjudicative judgments, such as automatically extracting information like the number and identity of defenders—information expressed with high consistency. However, when regular expression is confronted with diversified expressions, it becomes powerless, since methods of expression in reality cannot be exhausted. For example, although confession may not appear directly as a word (“confess”) in a document, it may be the driving force behind a diverse expressions such as “seized and turned over to police by family members.” In such a case, researchers need to use the NLP technique.<sup>45</sup> More methods and techniques should also emerge for analyzing and judging the correlation and causality among data.<sup>46</sup> For example, Zhang adopted a quantitative comparative-law approach, using cutting-edge statistical methods and concrete and numerous criteria (170 to be exact) to study one legal field (property law), and eventually produced a dendrogram (i.e. a legal family tree) consisting of 128 jurisdictions, with which he analyzed the similarity between these jurisdictions.<sup>47</sup> In the field of data analysis, the transition from “soft science” to “hard science” can only be realized by turning the subjective and assumptive causal analysis into a more objective and scientific correlation study. It is worthwhile for future researchers to attend to the fact that the machine-learning methods, which are statistically relevant though quite different, are now emerging and being used in big-data analysis. When existing analytical and statistical tools can no longer meet the needs of big-data processing, artificial intelligence, a new and evolving technical tool, has entered the stage. Through intelligent sifting and algorithm analysis of huge amounts of data, artificial intelligence can achieve significant improvement in the performance of massive-data analysis. For example, in Blumenstock et al.'s study, they created and trained a machine-learning model that can predict the answers given by 1.5 million users in the survey.<sup>48</sup>

Third, researchers should adhere to the integration of big-data and small-data research methods. Big-data legal research should be carried out in a variety of ways, rather than just

---

43. Zuo & Zhang, *supra* note 20, pp. 167–89.

44. Ansolabehere & Hersh (2012), pp. 437–59.

45. Wang, *supra* note 25, pp. 124–47.

46. See the six methods of causal inference introduced by Chang: experiment, breakpoint regression, double difference, instrumental variables, matching, and event research. Chang (2019b), pp. 135–41.

47. Chang, *supra* note 31, pp. 81–101.

48. Blumenstock, Cadamuro & On (2015), pp. 1073–6.



relying on pure big-data interpretation. Big-data analysis has its own advantages when it comes to a general description, while it may depreciate and even entirely ignore the value of individuals. In addition, big data tends to neglect the influence of the background (i.e. the politics, society, and judicial system) behind the data. Therefore, big data cannot be considered a “thick description” when analyzing samples and does not sufficiently reflect the depth of the research object. For a long time, legal empirical research based on small data has always been the mainstream research method, as researchers generally obtain qualitative data by the means of surveys and in-depth interviews. The meticulousness and usefulness of small data permit an extreme degree of data mining and analysis towards individuals, and can create significant academic value different from that of big-data research. Therefore, big-data research will not replace small-data research.

Given the present conditions—namely incomplete data and failure to fully use analytical methods—current big-data research should be combined with small-data research to verify the research conclusions. On the one hand, the elaborated thinking and methods of small-data research can refine the big-data research. On the other hand, the abundance of big-data resources can enhance the scientific nature of small data. These two complement each other and jointly increase the value of research. This combination requires researchers not only to pay attention to the hidden information behind the adjudicative documents,<sup>49</sup> but also to look beyond these documents, actively and purposefully collect individual data, and conduct relevant interviews to verify and collect the information behind big data. For example, in an empirical study on the rule of illegal-evidence exclusion, Zuo learned that courts launched investigations into evidence legality in only 40–50% of the cases in which the defendant submitted an application for such an investigation. In other words, in approximately half of all cases in which a defendant asked the court to investigate the legality of evidence, no investigation was conducted. After interviewing the judges, Zuo found that an important reason behind this figure is that the judges at this stage do not want the defendants to apply for illegal-evidence exclusion; therefore, in practice, they tend to persuade the party to revoke or not to file an application. Even though some parties submit applications anyway, the judge has the ultimate discretion regarding legality investigations.<sup>50</sup> In several other empirical studies, Zuo adopted a similar face-to-face-interview research method.<sup>51</sup>

Fourth, big-data research should be combined with theoretical research. Data research is not the same as empirical research on data statistics. Factual descriptions without theoretical depth and data investigations departing from abstract theories are no different from boring investigation reports that academic researchers hope to avoid. Therefore, we should examine the tension between data and theory. It should first be noted that empirical data cannot be directly embedded in legal theory, for the empirical social-science emphasis on the dimension of truth and legal science focuses more on normative correctness; for example, the recognized “truth” in court is ultimately determined by criteria provided by legal norms rather than an empirical world.

---

49. E.g. when studying the judicial relief of ecological destruction, Zhang Zhongmin realized that the ecological-damage cases in the sample were largely obscured and could not be directly singled out by factors such as cause of action. Therefore, he applied the method of carefully reading the adjudicative documents, and finally obtained accurate data after the sorting-out process. Zhang (2016b), pp. 111–24.

50. Zuo (2015), p. 2015.

51. See Zuo (2016), pp. 159–73; Zuo (2018a), pp. 238–60.

Can data materials gained from an external perspective be transformed into an internal-perspective knowledge on law? According to Tuori, “the results of social science can become part of legal discourse,” since “legal discourse does also deal with extra-legal reality, but views it through normative lenses.” Further, he added that facts may “appear in legal discourse as factual premises of court decisions and the interpretative standpoints of legal dogmatics.”<sup>52</sup>

To be specific, Fischman, highlighting the important impacts that empirical research may exert on legal-theory development, once said that important empirical research “can guide legal reform,” “describe . . . legal phenomena that participants in the legal system find important,” or “contribute . . . to the development of theories.”<sup>53</sup> Likewise, Chang et al. enumerated the four roles that empirical legal research may play. First, it may function as the factual basis of normative argumentation. Second, it may be used for measuring the effect of positive law. Third, it may be used to describe legal arguments and legal phenomena. Fourth, it may explore the behavioural patterns of relevant legal practitioners (such as judges and lawyers).<sup>54</sup> In other words, the empirical facts relied upon by empirical research are used not only to illustrate legal phenomena or patterns, but also to potentially provide sufficient and objective evidence for the subsequent theoretical construction.

Big-data research serves similar functions. It provides us with not only the outcomes of legal decision-making, but also in-depth information on cases such as coercive measures taken, defence situations, evidence, and reasons for the judgment, etc. More importantly, legal empirical research based on judicial big data possesses a unique value that differs from traditional empirical research, given that its features, such as massiveness, continuity, authority, and neutrality, lend toward a better comparison with previous empirical research and open up the cutting-edge-problem domain in legal scholarship. Let us take the unprecedented size of data as an example. This feature frees legal research from traditional research objects and materials, and enables certain types of research to become possible, such as studies of rare events, the discovery of nuances, etc. If researchers can collect and analyze big data scientifically, they are much more likely to acquire objective cognition and discover hidden truths and the laws behind it. Therefore, they may verify or refute existing theories and even make an original theoretical breakthrough. For example, Spamann et al. revealed new phenomena through experimental data—that judges from common-law countries are actually less likely to be influenced by precedent than judges from civil-law countries, and that judges often improperly take into account some factors unrelated to law.<sup>55</sup> Though Spamann et al. have not offered a theoretical interpretation of this unexpected experimental result, the findings deviate significantly from common sense and traditional cognition, and will inevitably contribute to in-depth tracking and discussion by academia. Furthermore, Stremitzer et al. developed a new theory on the strength of the discovery of new phenomena—as opposed to the traditional “shoot for the moon” approach (“even if you miss it you will still land among the stars”); the study showed that, when rules are too demanding, the opposite effects to those desired by the rules will arise and medium

52. Tuori, *supra* note 2, pp. 188–9.

53. Fischman (2013), pp. 157–8.

54. Chang & Cheng, *supra* note 18, p. 75.

55. Spamann (2019).

rules will yield better results in practice.<sup>56</sup> These studies show us the potential of data to jump beyond the limitations of existing viewpoints and theoretical frameworks, and directly reveal the hidden face of the objective world.

Throughout the history of academic development, the backdrops of all astonishing achievements are made up of the inconspicuous, yet indispensable, work of scholars making their way brick by brick and step by step. When big-data legal research, the newly rising legal-research paradigm, opens up a promising door for legal scholarship, it also establishes difficult challenges intended to motivate entrants to rise to the challenges. Therefore, in the face of the unprecedented opportunities for legal scholarship provided by big-data platforms, such as China Judgements Online, researchers may wish to keep up with the new research paradigm, seek improvement from frustration, ignite innovation in changes, and contribute to the prosperity of big-data-based legal research together.

## REFERENCES

- Ansolabehere, Stephen, & Eitan Hersh (2012) “Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate.” 20 *Political Analysis* 437–59.
- Bai, Jianjun (2015) “A Little Influence of Big Data on Legal Research.” 1 *Peking University Law Journal* 29–35.
- Bai, Jianjun (2017) “The Possibility and Limit of Adjudication Prediction in the Era of Legal Big Data.” 10 *Exploration and Free Views* 95–100.
- Beijing Youth Daily (2019) “The Data of China Judgements Online is Sold at a Price: Crawling by the Crawler Program May Constitute Infringement.” <https://tech.sina.cn/i/gn/2019-08-02/detail-ihytcitm6291286.d.html> (accessed 12 August 2019).
- Blumenstock, Joshua, Gabriel Cadamuro, & Robert On (2015) “Predicting Poverty and Wealth from Mobile Phone Metadata.” 350 *Science* 1073–6.
- CELS 2019 HOME (2019) “About the Conference on Empirical Legal Studies,” <https://cels2019.cmc.edu/> (accessed 12 August 2019).
- Chang, Yunchien (2019a) “The Causal Revolution of Quantitative Legal Empirical Research.” 2 *China Law Review* 135–41.
- Chang, Yunchien (2019b) “The Position of the Book of Things of the Chinese Civil Code in the Property Law of the World—Quantitative Comparison Method.” 1 *Global Law Review* 81–101.
- Chang, Yunchien, & Jinhua Cheng (2018) “Methodological Coordinates of Legal Empirical Research.” 6 *China Law Review* 73–89.
- Cheng, Jinhua, & Zhenxing Ke (2018) “The Federal Practice of Chinese Legal Power—Taking the Field of Labor Contract Law as an Example.” 1 *The Jurist* 1–17+192.
- China Law Review (2019) “[Agenda] International Conference on Legal Empirical Research, 2019,” [https://wemp.app/posts/d565dabc-b604-42a8-99a4-07c42533351f?utm\\_source=latest-posts](https://wemp.app/posts/d565dabc-b604-42a8-99a4-07c42533351f?utm_source=latest-posts). (accessed 3 September 2019).
- CHINATRIAL (2017) “China Judgements Online Has Exceeded 10 Billion Total Visits,” <http://www.chinatrial.net.cn/news/9974.html> (accessed 12 August 2019).
- Fan, Bainai, & Xiaojing Lou (2013) “Statistical Analysis and Evolution Path Research of Public Management Research Methods in China.” 2 *Journal of Public Management* 94–100+41.
- Feng, Jianpeng (2017) “Constitutional Citation and Its Function in Judicial Judgements in China: An Empirical Study Based on Publicized Judgement Documents.” 3 *Chinese Journal of Law* 44–59.
- Fischman, Joshua B. (2013) “Reuniting Is and Ought in Empirical Legal Scholarship.” 162 *University of Pennsylvania Law Review* 117–68.

56. Stremitzer (2019).

- Gu, Liping (2017) "Collection and Disposal of Citizen Privacy Data in the Era of Big Data." 9 *Academic Journal of Zhongzhou* 161–7.
- Hart, H. L. A. (2012) *The Concept of Law*, 3rd edn, Oxford: Oxford University Press.
- He, Fan, Na Lin, Jianwen Yang, & Fei Long (2013) "An Overview of Foreign Adjudicative Documents Online." 5 *Legal Information* 51–6.
- Hu, Ling (2015) "The Impact of the Rise of Big Data on Legal Practice and Theoretical Research." 4 *Journal of Xinjiang Normal University (Edition of Philosophy and Social Sciences)* 108–13.
- Jin, Hai, & Bo Shao (2018) "An Empirical Study of Environmental Infringement Cases of Several People—Taking 499 Adjudicative Documents from China Judgements Online as Samples." 18 *Environmental Protection* 56–9.
- Law School Yunnan University (2018) "The Agenda of the 3rd Annual Conference on Chinese Legal Empirical Research," <http://www.law.ynu.edu.cn/info/1058/2242.htm> (accessed 3 September 2019).
- Lei, Xinhong (2017) "Legal Empirical Research in China from the Perspective of Methodology Evolution." 4 *Chinese Journal of Law* 97–116.
- Ma, Chao, Xiaohong Yu, & Haibo He (2016) "Big Data Analysis: China Judicial Adjudicative Documents Online Public Report." 4 *China Law Review* 195–246.
- Qiao, Shitong, & Wenzheng Mao (2018) "Judicial Control of Administrative Expropriation: An Analysis Based on the Judgements of High Courts." 4 *Tsinghua University Law Journal* 68–90.
- Shen, Tingting (2015) "Analysis on the Characteristics of Use of Scientific Data in Humanities and Social Sciences: An Empirical Study Based on the Sample Papers from Social Sciences in China." 3 *Journal of Academic Libraries* 101–7.
- Spamann, Holger (2019) "Judging around the World: Experimental, Clickstream, and Writing Evidence from the Lab" Presented at 2019 International Conference on Legal Empirical Studies, Chengdu, China, 20–21 July 2019.
- Stremitzer, Alexander (2019) "Aspirational Rules" Presented at 2019 International Conference on Legal Empirical Studies, Chengdu, China, 20–21 July 2019.
- Tian, He, & Yanbin Lu (2017) *Empirical Research*, Beijing: Social Sciences Academic Press (China).
- Tuori, Kaarlo (2008) "Two Challenges to Normative Legal Scholarship." 53 *Scandinavian Studies in Law* 188–9.
- Wang, Lusheng (2018) "On the Symbolic Legislation of Criminal Proceedings and Its Consequences—Based on the Natural Semantic Mining of 3.03 Million Judgements Big Data." 6 *Tsinghua University Law Journal* 124–47.
- Wang, Qingfang, & Derui Du (2015) "The Methods and Orientation of China's Economic Research: An Analysis Report Based on 1126 Papers from 2012 to 2014." 3 *Nankai Economic Studies* 140–53.
- Wang, Wei, David Rothschild, Sharad Goel, & Andrew Gelman (2015) "Forecasting Elections with Non-representative Polls." 31 *International Journal of Forecasting* 980–91.
- Wiley Online Library (2019) "Journal of Empirical Legal Studies," <https://onlinelibrary.wiley.com/loi/17401461> (accessed 12 August 2019).
- Xiong, Qihong (2012) "The Transformation of Criminal Procedure Law Research Viewed through Chinese Journal of Law." 5 *Chinese Journal of Law* 3–7.
- Xu, Ming (2017) "The Privacy Crisis in the Era of Big Data and Its Tort Law Response." 1 *China Legal Science* 130–49.
- Yao, Wanqin (2015) "The Status of Juvenile Delinquency in China and the Correction of Preventive Countermeasures." 4 *Chinese Youth Social Science* 112–6.
- Yi, Feifei, Chao Ma, & Haibo He (2017) "The Disclosure of Judicial Statistics in China: Status and Recommendations." 2 *China Review of Administration of Justice* 56–72.
- Zhang, Hua (2018) "On the Reference Effectiveness of Guiding Cases—Based on the Empirical Analysis of 1545 Publicizing Adjudicative Documents." 2 *Journal of Gansu Political Science and Law Institute* 119–35.

- Zhang, Jiyu (2016a) "The Main Challenges and Opportunities for China's Judicature in the Age of Big Data—Also on the Judicial Needs of Jurisprudence Research and Talent Cultivation in the Age of Big Data." 6 *Law and Social Development* 52–61.
- Zhang, Zhongmin (2016b) "Judicial Relief for Ecological Destruction—Analysis of Samples Based on 5792 Environmental Adjudicative Documents." 10 *Law Science* 111–24.
- Zhao, Yifan, Liang Bian, & Xin Cong (2017) "Research Summary on Data Cleaning Methods." 12 *Software Guide* 222–4.
- Zuo, Weimin (2015) "'Hot' and 'Cold': An Empirical Study on the Application of Illegal Evidence Exclusion Rules." 3 *Studies in Law and Business* 151–60.
- Zuo, Weimin (2016) "An Empirical Study of the Operational Status of the Judicial Committee." 3 *Chinese Journal of Law* 159–73.
- Zuo, Weimin (2017a) *Chinese Legal Empirical Research*, 1st edn, Beijing: Law Press China.
- Zuo, Weimin (2017b) *Chinese Legal Empirical Research*, 2nd edn, Beijing: Law Press China.
- Zuo, Weimin (2018a) "China's Response to the Litigation Explosion: An Empirical Analysis Based on the Trial Practice of the W District Court in the Past 30 Years." 4 *China Legal Science* 238–60.
- Zuo, Weimin (2018b) *Chinese Legal Empirical Research*, 3rd edn, Beijing: Law Press China.
- Zuo, Weimin (2018c) "Towards Big Data Based Legal Research." 4 *Chinese Journal of Law* 139–50.
- Zuo, Weimin, & Lianhan Zhang (2019) "Criminal Defense Rate: Analysis of Differentiation and Its Economic Factors—Taking Sichuan Province's 2015–2016 First-Instance Judgements as a Sample." 3 *Chinese Journal of Law* 167–89.