

Big-Data Measurement-Model Research about Judges' Actual Workload in China[•]

Li YANG*

Shanghai Jiao Tong University

Junlin YI**

Shanghai Jiao Tong University

Hui PENG***

Shanghai Academy of Social Sciences

Abstract

As the growing number of cases is draining the limited court resources in China, how to scientifically measure the reasonable saturated workload of judges has become an urgent issue. This issue is the prerequisite of other important topics such as determination of judges' quotas, measurement of the actual workload of a trial team, performance evaluation of judges, and resource allocation within courts. Data-driven measurement of the actual workload of China's judges depends on various factors such as local economic development, public transportation, case-load in the past, and staffing of assistant positions. Therefore, traditional approaches that depend only on a single element, such as cause of action, do not work well. We proposed a modelling framework based on big-data and machine-learning technology to more accurately measure the actual workload of judges. This framework extracts the core elements of judicial cases, assigns target workload to the cases based on feedback from judges and analyzing case samples to create a standard training dataset, and trains machine-learning models using the data. A preliminary case-weight calculation model is built using the framework. Besides, the model is continuously evaluated and improved by comparing its output with the actual demand in a court through methods such as sampling, questionnaires, and expert evaluation.

Keywords: judicial reform, judges' quota, court-performance evaluation, measurement of judge's workload, big data, law

1. INTRODUCTION

Over the past couple of years, China's judicial reform has completed some fundamental projects contributing to the enhancement of a system of classified management of court staff,

• This work was supported by the National Social Science Fund of China#1 (under grant number 18VSI078) and the National Social Science Fund of China#2 (under grant number 17ZDA130). Also, it is a part of the National Key Research Program (2018YFC0830400).

* Professor of Shanghai Jiao Tong University KoGuan School of Law, China Institute for Urban Governance.

** PhD student, Shanghai Jiao Tong University KoGuan School of Law.

*** Professor, Institute of Legal Studies, Shanghai Academy of Social Sciences. Correspondence to Hui Peng, No. 7, Lane 622, Huaihai Middle Road, Shanghai, China, 200020. E-mail address: penghui@sass.org.cn

a system of judges' accountability, a career guarantee for judges, and a unified management of personnel, funding, and property in local courts under the provincial level. Thereafter, the critical topic of the measurement of judges' workload emerged in order to reform the dynamic administration system for judges' quota.¹ As Chief Justice and President of the Supreme People's Court of China Qiang Zhou pointed out, the reform of the quota system for judges is essentially allocating staff under judicial principles, which is an important mechanism to implement the regulation, specialization, and professionalization of judges, the groundwork for classified management of court personnel, and the cornerstone of enforcing judges' accountability.²

From the above perspective, how to measure the actual workload of judges is essential to the quota system. This is not only an academic research topic, but also an unavoidable and urgent problem in the cause of reforming China's court system and supportive mechanisms.

The beginning of the personnel-quota reform for China's judges can be traced back to 2002 when China's Supreme People's Court stated in the Opinion about Enhancing the Building of Professional Judge Team (the "2002 Opinion") that the personnel-quota system for judges should comprehensively consider factors such as China's national condition, case-load, area of jurisdiction, and status of economic development, and therefore determine the reasonable number of judges in courts of different trial levels.³ In 2004, the court further clarified in the opinion about Piloting Judges' Assistant System in Certain Areas that, under the prerequisite of completing the allocated case-load efficiently, the pilot courts should consider the number of cases and their forecasted changes to be the basic criteria when determining the quota for judges and at the same time take into account other factors such as the quality of the judge team, population size, and the area of the jurisdiction.⁴ Besides, the court also discussed the issue of determining judges' personnel quota and court staffing in the 4th Five-Year Outline of the Program for Reform of People's Courts (the "2015 Outline").⁵

In retrospect, the research of the measurement of judges' quota can be divided into two main stages. The first stage began after the quota system was proposed, consisting of the first, second, and third five-year plans for China's judicial reform, during which only a few related factors were listed. For example, the Supreme People's Court listed the following factors for the determination of judges' quota: China's national conditions, judicial workload, area and population size of a jurisdiction, economic and social development, existing personnel allocation in courts, and the number of collegial panels and presiding judges.⁶ Although some theoretical models exist, they only consider a limited set of factors such as the density of the population, the judicial workload, and the social- and economic-development condition, lacking specific approaches for workload measurement.

The second stage started in 2011 and the research on the quota system became more refined. Two basic categories of measurement models were proposed. In the first category, measurement depends on the amount of judicial workload or workload per judge, where the

1. Wang (2015).

2. Zhou (2016).

3. Supreme People's Court (2002).

4. Supreme People's Court (2004).

5. See Supreme People's Court (2015).

6. See Supreme People's Court, *supra* note 3.

per-judge workload is calculated by multiplying the number of cases and the workload per case. In the second category, researchers tried to consider all possible factors influencing the number of cases, including, but not limited to, the area of jurisdiction, economic development, and the proportion of people assessors and clerks. Based on the above factors, researchers leveraged SPSS (Statistical Product and Service Solutions)⁷ software to build models combining regression analysis and workload measurement.

We found the following areas of improvement for these models. The first improvement is about setting measurement factors. Existing researchers tried to include all possible influencing factors, which is neither realistic nor advisable. Factors such as economic development, area of jurisdiction, and population size indirectly influence the judicial work and their impact is eventually reflected in the courts' workload through litigation procedures.⁸ Therefore, judicial workload is the deciding factor in the measurement of the personnel quota.⁹ Second, when measuring the quota, the existing methods divide the total case-load by the workload per judge. Some models calculate total case-load by averaging different cases; some simply divide cases into those that are adjudicated and those that are withdrawn, ignoring the specificity of individual cases; some approaches do not categorize cases into refined types, failing to set weight coefficients for different types of cases; even for those researchers who did set the weight coefficients per case type, they still failed to give precise methods calculating the coefficients. Third, researchers overlooked the impact of the opening of assistant positions in the process of judicial reform measuring judges' workload. To reasonably estimate judges' workload in the future, researchers should consider the responsibility of judge assistants, their undertaking of work items, and the corresponding reduction in judges' work, instead of ignoring assistant personnel's impact.

Therefore, the advancement of China's personnel-quota system for judges, especially the research of measuring judges' workload, relies on answering the following questions. How can the work items of the judges be scientifically measured? Under the theory of separating core work and supportive work, how can all work items and unit time per item be represented via empirical study? How can the impact of assistant roles on judicial work items and the core work of judgment be evaluated? How can the weight coefficients of different types of cases be determined? How can the amount of judicial work after excluding the supportive work undertaken by assistants be measured? How can the reasonable workload of each judge after considering the impact of the assistant positions be estimated, and therefore the quota for judges be derived?

2. BACKGROUND OF DYNAMIC ADMINISTRATION FOR JUDGES' QUOTA AND MEASURING JUDGES' WORKLOAD

With the number of cases multiplying, conflicts between the growth of cases and the insufficiency of judges become more serious. China's judges have been overloaded for a long

7. SPSS is widely used statistical-analysis software in the field of social science. Apart from statistical analysis, it also provides features such as data management (including case selection, file reshaping, and data derivation) and data documentation (storing a metadata dictionary in data files).

8. More correlation analysis between legal indicators and social-development indicators in different jurisdictions can be found in Zhu (2007), pp. 64–70, and pp. 58–63 provide nationwide analysis.

9. Guo (2013).

time and the situation is getting worse as the amount of case-load continues to grow while judicial productivity falls behind. It is important to measure the saturated workload of judges in a scientific and reasonable manner and to accurately represent judges' working situation. This topic is not only concerned with the judges' physical and mental health—a legitimate humanistic concern by itself—but is also crucial to the sustainable development of the cause of people's courts and the profession of judges. Until now, scholars and legal practitioners have been keenly discussing judges' annual maximal workload, but most research is still in the state of comparative research or theoretical studies. No one has empirically studied the topic in a satisfactory manner; also, the recent pilot programmes published by local courts do not clarify how to measure judges' workload. Therefore, we attempt to propose a framework to analyze and measure judges' workload more accurately, laying the groundwork for determining judges' quota and assistant personnel's proportion in courts. More importantly, we would like to call scholars' attention to the core responsibility of judges, inspiring more research into trial-administration and court-performance appraisal under judicial principles.

It is necessary to study the dynamic administration of China's judge-quota system and the scientific measurement of judges' workload for the following reasons.

2.1 Solving the Challenge of Insufficient Judges

Apart from approaches such as improving judicial productivity by separating simple and complex cases, piloting programmes of punishment reduction considering defendants' confession, reforming trial activities, and diversifying the mechanisms for dispute resolution to mitigate the shortage of judges in China's courts, it is important to optimize the allocation of courts' resource through scientific performance evaluation.¹⁰ The study of dynamically adjusting judge quota can bridge the gap between the workload measurement for different types of cases, enabling the consistent evaluation of cases within a tribunal and in between different tribunals; measuring judges' maximal workload and current work level provides a scientific standard for predicting judges' annual, saturated workload.¹¹

2.2 Evaluating the Performance of Judges

The Supreme People's Court stated in its 2015 Outline that the basic data deciding judges' quota are economic and social development, population size (including temporary residents), the number of cases, and the types of cases, and that other factors to consider include the courts' trial function at different levels, furnishing of supportive staff members for trial, the amount of work done by judges, and conditions ensuring processing of cases.¹² It is crucial to have a measurement standard for judges' workload in order to empirically evaluate their performance and to reform the quota system. Some research and explorations exist in theory and practice about the dynamic administration of judges' quota, but they have limitations because they only depend on a small set of data samples.

10. Su (2010).

11. Tan & Liang (2009).

12. See Supreme People's Court, *supra* note 5.

2.3 Promoting the Informatization of China's Courts

In the context of big-data technology, China's society is having a keen discussion about how to informatize the judicial system and to improve courts' capability, in which the dynamic administration of a judge-quota system can play a leading role.¹³ Thus, approaches for the scientific measurement of judges' actual workload based on big-data technology are of a strong interest in terms of demonstrating the value of judicial information systems and facilitating the informatization of China's courts.

3. RESEARCH THEORY FOR MEASURING JUDGES' ACTUAL WORKLOAD AND EXAMPLES OF LOCAL COURTS

3.1 Literature Review

Currently, China's judges have been overloaded by heavy case-loads for a long time; the situation is getting even worse as the amount of judicial workload continues to grow while the courts' productivity cannot keep up with the pace. Therefore, the scientific measurement of judges' saturated workload, aiming to disclose judges' actual working conditions, is not only a humanistic concern regarding judges' physical and mental health, but also plays a crucial role in the sustainable development of the cause of China's court system and the profession of the judge.

The key to promoting the standardized, specialized, and professional development of a court staffing system is to enhance the classified management for court personnel.¹⁴ There exists consensus in terms of the organizational barriers in court systems restricting judicial capability and people are aware of the necessity to establish a staffing system to meet the characteristics of judicial professions. Yet, the key question of the classified management of judicial personnel is not fully answered: how can the judges' quota be scientifically and reasonably determined? Different answers have been provided by researchers. Li Yang found that the proportion of judges to courts' staff was around 30% to 40%, which is a reasonable proportion and is close to the target value set in pilot programmes.¹⁵ Weidong Chen proposed that insufficient funding of judge assistants was the bottleneck for judicial reform and that the number of assistants should be increased incrementally while the number of judges should be decreased, based on factors such as job accountability and workload.¹⁶ Douyun Chen proposed that the target proportion set in the pilot programmes of the judge-quota system was reasonable, that the number of judges must match the amount of judicial work, and that China should make it a top priority to ensure judges' professional development and the completion of judicial work.¹⁷ Yongsheng Chen thought that the upper bound of judges' proportion (39%), set by the Supreme People's Court and the Supreme People's Procuratorate, should be revisited and revised based on different jurisdictions, case types, and court levels; the upper bound should be lifted in some areas, while

13. Liu (2019).

14. See Supreme People's Court, *supra* note 5.

15. Yang (2016).

16. Chen (2019).

17. Chen (2014).

the number can be decreased in rural regions of Western China; besides, scholars in places such as Inner Mongolia, Qinghai, Guizhou, and Yunnan emphasized the special characteristics of ethnic areas and advocated a dynamic quota threshold for judges or at least for minority judges.¹⁸ Ruihua Chen thought that a fixed proportion would essentially make the system depend on whether there are judicial vacancies.¹⁹ Fei Feng thought that the quota system should be examined against whether and to what extent the quota system satisfied the original requirements of China's judicial reform instead of focusing on meeting a specific target for judges' proportion, and that the system should be also justified in terms of what paths it is paving for the future reform of judicial practices.²⁰ To resolve the above issues, researchers studied the scientific measurement of judges' workload. For example, Jing Wang and others picked a sample of 55 civil judges in basic-level courts and leveraged methods such as participant observation, questionnaire, interviewing, and video-recording to classify and quantify the amount of judgment work.²¹ Based on the result, they proposed the separation of core judicial work and supportive work, and further advocated that, under the existing litigation procedure and judicial organizations, the quota for judges should depend on the core work and the quota of judge assistants should depend on the supportive workload.²² Weimin Zuo thought that the basic data for calculating judges' quota were the number of cases or, in other words, the judgment workload, because of a subtle interaction among the organizational structure of a court, the function structure of the judges in the court, and the number of cases accepted by the court.²³ Xiangdong Qu recommended a workload-measurement model to estimate the quota-based core factors including case type, work task, task frequency, and task complexity.²⁴

3.2 Examples of Judges' Workload Measurement in Local Courts

In practice, the performance evaluation in China's courts are based on tentative quantitative analysis and manual adjustments. Quantitative analysis (weight-coefficient assignment, internal estimation, linear regression, etc.) is conducted via sampling methods (interviewing experts, questionnaires, judgment data extraction, and browsing the statistical yearbook). Additionally, performance evaluation is manually adjusted based on inputs from experienced case-handling staff in terms of different types of cases and causes of action, and is assisted with statistical methods. Here, we analyzed three typical examples of weight-coefficient calculation: Shanghai, Jiangsu, and Guizhou courts.

3.2.1 Shanghai

The calculation of weight coefficients in Shanghai is via the "2 + 4" mode; the "2" here means two basic factors: cause of action and litigation procedure; the "4" represents four

18. Chen & Bai (2016).

19. Chen (2018).

20. Feng (2015).

21. Wang et al. (2015).

22. *Ibid.*

23. Zuo (2017).

24. Qu (2016).

variables for calculation: length of court session time, word count of transcripts, the number of trial days, and word count of legal documents.²⁵ By comparing the four variables in different cases and the variables' proportions in cases, Shanghai courts determined the weight coefficients applying to different types of cases. The calculation consists of four steps. First, collect high-priority data from the case itself, including the trial time, legal documents, word count of transcripts, and length of court sessions. Second, calculate normal weight coefficients. Basically, within a specific time horizon, the courts calculate the averages for the four variables mentioned above for all cases and then use the average weights as a baseline; thus, the weights of a specific case in the same time horizon can be derived by comparing it to the average. For example, if the baseline number is 1 and a certain case's variable average is calculated as 1.5 times that of the baseline, the weight of that case is 1.5. Third, set adjustable weight coefficients. In some cases, judges' work increases due to counterclaims or addition of third parties. An adjustable coefficient is therefore configured to increase the weight coefficients accordingly. The adjustable coefficient is calculated by comparing all cases having the above elements with other cases that do not have such elements; say, if the adjustable coefficient of a counterclaim is 2.05 and the cases without counterclaims have a coefficient of 1.2, then the adjustable coefficient is 0.85, the delta of the two coefficients. Fourth, set the fixed weight coefficients. After calculating the normal weight coefficients, Shanghai courts assign a fixed weight to simple cases or those cases in which special procedures are applied, which is irrelevant to the cause of action.²⁶ Take simple batch cases as an example; their fixed weights are calculated by how the cases are closed—that is, judgment, mediation, or withdrawal. Based on the fixed weight coefficients published by Shanghai courts, the weight for simple batch cases is 0.18, cases for mediation weight 0.09, and the weight for withdrawn cases is 0.05.²⁷

The above approach has some drawbacks. First, it relies on a small set of data instead of big data of the overall context, so the accuracy is affected by the quality of the basic data and requires the data to be highly structured. Second, the approach is not portable because it does not include all factors influencing the overall case-handling work and the weight calculation is complex. Third, the approach only used historical data, failing to foresee new types of cases or causes of action. Finally, the approach lacks the ability to evolve by adapting to new circumstances, as the prototype design was done in 2008 and has not been updated since.

3.2.2 Jiangsu

Jiangsu proposed a next-generation system for judge-performance appraisal and case weighting by considering complex dimensions, including both fixed weights and adjustable weights. A early-phase design for the user interface has been piloted in some courts of Jiangsu province, receiving positive feedback. Jiangsu's reform has the following advantages: (1) heads of the courts highly valued the measurement of judges' workload; (2) the dimensions in the system became more and more refined; (3) pilot programmes covered

25. Chinacourt.org (2015).

26. *Ibid.*

27. Cui (2015).

a broad range of jurisdictions within the region; (4) the user interface was implemented and delivered effectively. With that being said, Jiangsu's system strongly relies on small data—a problem similar to that of Shanghai courts. Jiangsu courts used a questionnaire to cover a wide range of audiences, collecting data for statistical analysis, but they did not conduct data mining on top of the data; also, big data and AI technology were not used.

3.2.3 Guizhou

Guizhou's approach relied on external and internal data about a jurisdiction's economic development, population, number of cases, and the types of cases. Besides, when determining the quota for judges in courts of different trial levels, Guizhou combined other factors, such as the court level, staffing of judge assistants, and conditions ensuring case processing. Guizhou courts' approach has the following advantages. First, heads of Guizhou High People's Court gave a high priority to the reform of the personnel quota. Second, Guizhou proposed some innovative concepts in an early phase, laying the theoretical groundwork for policy-making. Still, Guizhou's approach has some shortcomings: (1) collaboration was insufficient between the courts' divisional leaders and other government institutions; (2) the courts lacked external data and their internal data are not structured enough; (3) some of the modelling dimensions are difficult to measure; (4) Beige Data, the company implementing the system, needs to understand more about the courts' domain knowledge.

In summary, research into the measurement of judges' workload is still in an early stage, and a scientific, reasonable, and effective system to reflect judges' workload is lacking. As measurement models and data-collection approaches become more advanced, researchers have started to focus on measuring judges' workload. Workload measurement is the key to opening the door for performance evaluation, enabling the scientific allocation of courts' resources, guaranteeing the development of judicial professions; also, it will help society to understand why judges are overloaded. Thus, measuring judges' workload is indispensable in the implementation of the quota system.

4. ANALYSIS OF THE SUPREME COURT'S MEASUREMENT FRAMEWORK FOR JUDGES' WORKLOAD

Before proposing any new approach to measure judges' workload, we need to answer the core question: which factors influence or decide the number of cases heard by courts? Answering the question will help us to construct a modelling framework, such as the three-stage process in Figure 1, to guide the collection of key factors key factors via big-data technology. Fortunately, the Supreme People's Court has provided a framework to answer this question in its authoritative documents related to judicial workload. Here, we review the key ideas presented in the documents, identifying the core elements in the reform of judges' quota.

First, in the Opinion about Enhancing the Building of Professional Judge Team (the "2002 Opinion"), the Supreme People's Court proposed a plan to implement the judge-quota system, considering the following factors: China's national condition, case-load, area of jurisdiction, population size, economic-development situation, etc.²⁸ At the same time, since

28. See Supreme People's Court, *supra* note 3.

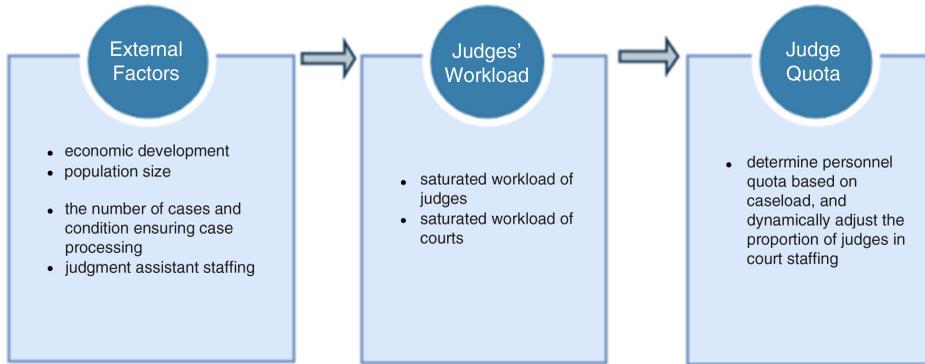


Figure 1. Modelling framework for judges' annual maximal actual workload.

there are already a large number of judges in China and judicial institutions are overstaffed, the Supreme People's Court intended to limit the quota of judges within the courts' existing personnel size.²⁹

Second, in the Opinion about Pilot Programs of Judge Assistants in Certain Local Courts (the "2004 Opinion"), the Supreme People's Court set the goal of implementing the classified management of judicial staff, stating that the primary factors to consider when determining judges' quota are the number of cases and the trial workload, and that other factors include judges' quality, organization, area of jurisdiction, economic development, and population size.³⁰

Third, in the Reply of the Supreme People's Court on the Opinions and Suggestions from Netizens III (the "2009 Reply"), the Supreme Court stated that the primary criterion determining court staffing is workload and other influential factors include the economy, location, population, and trial levels of the people's courts; on top of these criteria, the court advised that the personnel-quota system should be designed under the principle of classified personnel management, taking into account the characteristics and workload of courts in different levels.³¹

Fourth, in the Opinions of the Supreme People's Court on Comprehensive Deepening of Reform of People's Courts—The 4th Five-Year Outline of the Program for Reform of People's Courts (the "2015 Opinion"), the court proposed the goal of regularization, specialization, and professionalization of court staff.³² The primary data determining judges' quota for all courts are the social development of the jurisdiction, the size of the population (including temporary resident population), the number of cases, and the types of cases.³³ Other factors consist of the courts' function at different trial levels, judges' workload, supporting staff members, and conditions ensuring case processing.³⁴ Moreover, because of the severe attrition of judges, the Supreme People's Court emphasized in its 2015 Opinion that a transition plan

29. *Ibid.*

30. See Supreme People's Court, *supra* note 4.

31. See Supreme People's Court (2009).

32. See Supreme People's Court, *supra* note 5.

33. *Ibid.*

34. *Ibid.*

should be formulated during the reform of the quota system, ensuring that outstanding judges could still remain at the forefront of justice.³⁵

Through comparison, we noticed some changes in the factors identified by the Supreme People's Court influencing judges' quota (Table 1). In the 2004 Opinion, the court divided the "comprehensive factor" in the 2002 Opinion into "basic factor" and "comprehensive factor," rephrasing the judgment workload to become a basic factor, and kept other comprehensive factors (area of jurisdiction, population size, and economic-development level). Entailing the 2004 Opinion, in its 2009 Reply, the court continued the same thought except that two basic factors in its 2004 Opinion were combined together to become one single basic data factor (workload) and that two comprehensive factors in its 2009 Reply (court level and court's characteristic) substituted for the factor of judges' quality and judicial organization in its 2004 Opinion.³⁶ In February 2015, the Supreme Court rephrased the "basic factor" and the "comprehensive factor" in its 2009 Reply to "basic data" and "supportive data;" "workload"³⁷ was renamed as "judges' workload" and became secondary data; some comprehensive factors in the 2009 Reply (such as economy, territory, and population) were upgraded to "basic data" and were rephrased as "economic and social development" and "population (including temporary residents)."³⁸ From the repeated adjustment of the terms and their corresponding modifiers for the terms, we can tell that the Supreme Court is very careful about describing the factors influencing the quota system. Take the item "population," for instance; it was used together with "the area of jurisdiction" in the 2002 Opinion, but it became a stand-alone factor in the 2004 Opinion and later, in the 2015 Opinion, "population" was rewritten as "the amount of people," with a supplement modifier to include temporary residents.³⁹ Up to 2015, the Supreme Court achieved a hierarchical vision of the various factors influencing judges' quota.

Based on the change history of the above four authoritative documents regarding the quota system, we reached the following conclusion: "the number of cases" (or case-load), repeatedly emphasized by the court, is the most important factor in deciding judges' quota; other datapoints only serve as supplements or expansions to case-load. Table 2 clearly shows that, among the four documents from the Supreme Court, "the number of cases" has always been a fundamental factor frequently ranked at the top. Though listed as the second to last factor in the 2015 Opinion, the importance of the number of cases is not lowered; rather, it was actually considered to be the eventual factor able to quantitatively represent all other data. Without the number of cases, it is difficult to manage courts in a data-driven approach because other data cannot be measured easily, which is inconsistent with the Supreme People's Court's reason to promote the quota system in the first place. Unfortunately, although the term "basic data" has been frequently referred to by the Supreme Court, by media, and in practice, this concept is abstract and ambiguous, waiting to be interpreted

35. *Ibid.*

36. See Supreme People's Court, *supra* note 3; Supreme People's Court, *supra* note 4; Supreme People's Court, *supra* note 31.

37. "Workload" is a basic factor in the Supreme People's Court's 2009 Reply.

38. See Supreme People's Court, *supra* note 5.

39. See Supreme People's Court, *supra* note 3; Supreme People's Court, *supra* note 4; Supreme People's Court, *supra* note 31.

Table 1. Naming changes in the Supreme People’s Courts’ authoritative documents

Factors/documents	Society	Case ^a	Court	Judge	Assistant staff	Guarantee
2002 Opinion	China’s national condition Area and population of jurisdiction Economic development	Case workload	No	No	No	No
2004 Opinion	Area of jurisprudence Economic development Population	The number of cases, judgment workload	Court organization and judgment workload	Judge’s productivity	No	No
2009 Reply	Economic growth Area Population	Workload	Court’s level, The characteristics of a court	Workload	No	Characteristics of courts
2015 Opinion	The economic and social development, and population (including temporary residents) in the jurisdiction	Number of cases, type of case, judge’s workload	Court level and function	Judge’s workload	Assistant staffing for trial work	Conditions ensuring case processing

^aThe factor of cases includes the number and the type of cases, but it can also be converted into the workload of the judges or courts. Thus, the workload can be put into all three factors (case, court, and judges).

Table 2. Ranking changes in the Supreme People's Courts' authoritative opinions

Factors/documents	Society	Case	Court	Judge	Assistant staff	Guarantee
2002 Opinion ^a	Ranked first, third, and fourth in comprehensive factors	Ranked second in comprehensive factors	Same as on the left	Same as before	No	No
2004 Opinion	Ranked third, fourth, and fifth in comprehensive factors	Ranked first in basic factors	Ranked second among comprehensive factors	Ranked first among comprehensive factors	No	No
2009 Reply	Ranked first, third, and fourth in comprehensive factors	Ranked first in basic factors ^b	Ranked fourth and fifth in comprehensive factors	No	No	Ranked fifth among comprehensive factors
2015 Opinion	Ranked first and second, in basic data	Ranked third and fourth in basic factors	Ranked first among supplement factors	Ranked second among supplement factors	Ranked third among supplement factors	Ranked fourth in supplement factors

^aThough the factors are all considered comprehensive factors in the 2002 Opinion, their order is different, which can be considered as the comparative ranking within the same category.

^bIn the 2009 Reply, workload is the only basic criterion.

by local courts based on their own condition. Therefore, the Supreme Court's opinions can only serve as high-level guidance.

In summary, among the three basic factors, the former two (economic development and population size) positively correlate with the third (the number of cases in a jurisdiction). This is aligned with the general observation that the number of cases heard by a court is correlated with the social-development indicators in terms of economy, urbanization, and population.⁴⁰ More specifically, there is multicollinearity among the three factors so they influence measurement models in a combinative way, concealing their independent influence, thus affecting the models' overall accuracy. Among the basic factors, the number of cases accepted by a specific court is the direct or deciding factor to the personnel quota for judges; and the type of a case, namely whether the case is of a simple or complex type, also plays an important role.

5. MEASUREMENT MODEL FOR JUDGES' WORKLOAD BASED ON DYNAMIC QUOTA MANAGEMENT

Without the system of classified management for court staff, the reform of China's personnel-quota system will go back to square zero.⁴¹ So, the judicial personnel-quota system and the classified management of court staff are closely related. Besides, the quota system plays a critical role in the advancement of the reform for the comprehensive mechanisms supporting China's court system. In the new round of top-down reform actions, the quota system is the cornerstone for promoting the system of judicial responsibility—an important mechanism to allocate courts' human resources under judicial principles and to ensure the standardization, specialization, and professionalization of judges. However, there are no detailed instructions and reference methods about how to implement the quota system; neither are there enough doctrinal discussions or piloting mechanisms. Thus, the personnel-quota system started to change from an official reform plan to an academic topic for discussion.⁴² After reviewing the existing research approaches, we proposed a big-data-based framework to build models measuring judges' workload based on calculating weights for different types of cases.

5.1 Case-weight-measurement Framework Based on Big Data

Big-data-based approaches are different from the traditional statistical-analysis approaches used in social-science disciplines. Traditional approaches generally start from a certain assumption and then establish indicators and models for verification, so the conclusion is generally easier to understand.⁴³ Yet, due to the dependency on a predetermined assumption, it cannot easily be adapted to new scenarios if the research object changes structurally; also, it has a higher requirement for data quality.⁴⁴ On the other hand, big-data approaches rely on the principle of discovering knowledge based on a large amount of data, so they do not require rigorous prerequisite assumptions and also have a higher tolerance of

40. Zhu, *supra* note 8, p. 59.

41. Dong & Huang (2018).

42. Wang (2017).

43. Liu & Yin (2017).

44. *Ibid.*

data quality; as more data are fed into the model, the model will iterate continuously and optimize its algorithm to eventually approximate the reality.⁴⁵ Here, we propose a big-data-based, supervised-learning framework that consists of the following stages: data collection for case elements, assignment of target workloads, model training, model evaluation, and feature engineering.

The first stage is data collection, whose responsibility is collecting case elements from the data sources in judicial systems. For those datasets that already exist in databases, we used database query technologies to extract, transform, and load the data from the data sources. Besides, lots of unstructured text data are not stored in databases, including data related to the length of court sessions, trial times, legal documents, and other procedural documents. To collect case-element information from such unstructured texts, we designed an information-extraction system based on technologies such as named-entity recognition,⁴⁶ knowledge graphs,⁴⁷ and log event collection. Besides, the post-processing data can be visualized for quick query and modification, streamlining the data-analysis process. Finally, when there are inconsistencies among data from different data sources, this stage normalizes and standardizes the data to resolve conflicts.

Next, output values (i.e. the estimated judges' actual workload) are assigned to the training data. Basically, under supervised learning, the training data for the learning model are a set of examples and each example consists of a pair of input values and the desired output value.⁴⁸ The model learns rules or patterns from the training data and is able to predict values under testing data not seen before. In this stage, to quantitatively estimate the output values, approaches such as participant observation, questionnaires, and interviews are used. In short, the assigned output values, together with the input data collected previously, become the training datasets for the model training in the next stage.

The third stage is model training and here we leveraged the supervised-learning algorithms to predict the weights for different types of cases based on their input case elements. The essential goal of supervised learning is to train the models' generalization capability, learning rules from the training datasets and then using the rules to predict results on the testing datasets. In our research, we used two simple, explainable algorithms—decision tree and linear regression—to demonstrate the capability of the framework; still, this model can be extended to support other algorithms. When training the model, training datasets prepared by the previous two stages are fed into the model and each pair of the training data has input values and a targeted workload. The model continuously compares its prediction with the target values and adjusts the algorithms until the predicted results are within a small error range; other model-tuning parameters include tree depth, the maximal number of tree branches, and the number of iterations of the regression algorithm.

45. *Ibid.*

46. Named-entity recognition (NER) is a common task in Natural Language Processing (NLP); its responsibility is tagging entities in text with their corresponding type. See nlpprogress.com (2019).

47. A knowledge graph models a specific domain created by experts of the subject matter. With the help of machine learning, a knowledge graph provides a common interface for your data, allowing you to create smart multilateral relations on your databases. See Poolparty.biz (2019).

48. Ashley (2017), p. 138.

The fourth stage is model evaluation. Generally, with sufficient data for training and validation, the model with the most accurate prediction on the validation data is the best.⁴⁹ The validation dataset is generated by expert review of case samples collected at a ratio of 7:2:1 in terms of the number of civil, criminal, and administrative cases. Though learning models are generally compared by their error rates, error is only one of the criteria.⁵⁰ Explainability is also important, as knowledge extraction should be checked and validated by experts.⁵¹ Therefore, apart from generating metrics about error ratios, it is necessary to review the results with judges and other legal experts, leveraging their expertise and knowledge to appraise whether the model is reasonable. Admittedly, the accuracy of any model relies on the quality and quantity of the input data, which always has room for improvement; yet, the model will eventually approximate the reality as it runs more iterations on more data.

Finally, we have a separate feature-engineering stage to select the most important features for model training. Features are the variables denoting the attributes of the input data⁵² and, in our framework, case elements are features. As numerous case elements can be collected from the judicial data sources, it is not desirable to feed all inputs into the model, because models become more complicated and more expensive as the number of inputs grows.⁵³ Therefore, feature-selection methods are leveraged to select a subset of key features from the original inputs. One intuitive approach is to run the model-training phases multiple times with different subsets of input features and identify which features have the biggest impact on the results.⁵⁴

5.2 Reform of Judges' Quota Based on Annual Average Workload

After explaining the stages constituting the big-data framework for learning the case weights, we will further explore five different dimensions of the above framework in the context of the dynamic administration of judges' quota.

First, the direct factor deciding judges' quota is the number of cases accepted by a specific court. Currently, some courts consider judges' quota to be the same as the proportion of judges in the courts' staff. A rigorous proportion leads to the following problems: (1) scepticism due to the lack of scientific and reasonable criteria supporting the proportion; (2) difficulty in applying the same proportion to other places; and (3) inflexibility adapting to changing conditions. Beyond this, the challenging question for courts to answer is how to measure judges' workload. Against this backdrop, we think the reform of the quota system should focus on the methodology to calculate the number of judges rather than finding the specific proportion figures. We observed that there is a positive correlation between the number of court cases and the following factors: economic development and population size (including temporary residents) in a jurisdiction. Or, in other words, these factors have multicollinearity that affects the measurement result in a combined and mixed manner, and conceals each individual factor's independent impact, thus reducing the accuracy

49. Alpaydın (2010), p. 40.

50. *Ibid.*, p. 477.

51. *Ibid.*, p. 478.

52. *Ibid.*, p. 87.

53. *Ibid.*, p. 109.

54. Xu & Hou (2018).

and explainability of models. Actually, the direct or decisive factor that affects the demand for the number of judges is the number of cases accepted by a specific court; the type of cases (complex or simple cases) matters as well. In short, in the judicial reform of the judges' quota, the authoritative documents should be analyzed and summarized, including the Supreme Court's 2002 Opinion, 2004 Opinion, and 2015 Opinion.

Second, build the model measuring judges' annual saturated actual workload. Dividing the case-load by the annual maximal workload of a judge will give us the quota of judges in a court. Here, a judge's annual maximal actual workload is the upper bound of the number of cases that can be fulfilled by the existing judicial resources within the allocated annual legal working hours of a judge. To reasonably estimate the number of judges and estimate the workload, we can leverage the following approaches: (1) data analysis (i.e. to collect and summarize datapoints for case types, way of closing a case, transcripts of collegial panels' discussion, the number of case files, and other basic information); (2) interviewing (to collect information about work items and their required time at various litigation phases, including pre-trial hearing, trial, serving legal documents, mediation, document preparation, and verdict); (3) measuring judges' workload by typical case sampling (to sample different legal procedures such as summary procedures by a single judge, summary procedures by a collegial panel, ordinary procedures, to analyze the time occupied by each procedure, comparing their similarities and differences, and to examine the time spent on difficult cases, understanding the actual workload hidden beneath the surface of legal documents). Also, to avoid the Hawthorne Effect—a type of reactivity in which individuals would modify their behaviour when aware of them being observed—we mainly relied on data analysis; interviewing was only used to obtain the work time for those items hard to quantify.

Third, study the factors influencing judges' workload model. Due to limits in funding, human resources, and technical analysis, it is hard to get the complete set of datapoints. Therefore, the samples used in research have to be a subset. Still, to measure judges' workload, just counting the number of cases is insufficient; data from all aspects such as the working environment and unit work time should be recorded so that enough datapoints are collected from the interviewees for the purpose of big-data analysis. Note that data-driven thinking is not an end in itself, but a means to surface the real problem, and therefore just the starting point of the research. Moreover, some dynamic influencing factors need to be considered: (1) the ideal maximal saturated workload generated by a model is the theoretical upper bound of the judges' work and a reasonable workload ought to be adjusted to fall below this upper bound to avoid overloading judges and draining the pool of judicial professionals; (2) an individual judge's workload may vary due to factors such as expertise, experience, family conditions, parental conditions, and job attitude; (3) external factors to the courts (such as the economic environment and the number of cases) also affect the judges' workload. Thus, some flexible buffers are required on top of any predetermined quota.

Fourth, conduct more research about big data and the dynamic adjustment of judges' quota. Specifically, the research can be conducted in the following five perspectives. The first perspective is about case elements. We care about the elements influencing the length and difficulty of a case, and these elements can be retrieved via data collection and processing structured or unstructured data. Currently, we have preliminarily collected about 30 elements and we do not plan to collect too many. If too many elements are included, it will become infeasible to analyze an individual element's impact on the results

via the method of control variables, thus losing the model's explainability. Until now, we have collected the following elements: (1) court-hearing elements: trial transcripts, and the number and length of court sessions; (2) document elements: the total word counts of texts such as judgment opinions, evidence, the reasoning section in an opinion, documents from the parties, holding, and the sources of law; the number of statutes and statutory codes cited; seizure of property; identification and evaluation; settlement; and court's examination; (3) judgment elements: case types, causes of action, trial time, reason for case closure, case numbers, case-type codes, litigation procedure, the way of case closure, the object in dispute, the number of people involved, appealed or not, the number of appeals, the existence of incidental civil actions, submitted to judicial committee or not, small-claim procedure or not, and the type of trial-supervision procedure. The second perspective is about the target data. Based on different case types (criminal, civil, or administrative), we sampled a few cases to conduct expert evaluation in terms of the length of the trial time. Third, we used machine learning to train models based on the standard base dataset (case elements and target data). Fourth, in terms of self-adaptive learning, measurement models are applied to the whole judicial datapoints of a providence and continue to be updated through iterations, so the measurement of accuracy of judges' saturated workload change over time. Still, the measured value will become closer and closer to the real value in the long run as time goes by. Note that the approximation process is incremental and depends on the quantity and quality of the input data, so it will not finish in a single iteration and requires continuous improvements. The fifth perspective is about applying the model's results to other areas, not limited to the analysis of judges' saturated workload based on big data. Our research framework enables the visualization of the evaluation result if courts have such a kind of requirement; for instance, visualization can include mobile applications displaying performance management and systems appraising the performance of judges and courts. Nevertheless, we recognize that these applications rely on the prerequisite input of the data's quantity and that quality can meet the standard required by machine learning.

Sixth, derive the quota for judge assistants based on the quota for judges. Basically, the staffing of judge assistants will reduce judges' workload, as the assistants will take care of supportive tasks such as reviewing the submitted materials, legal research, citation checking, time scheduling, and drafting legal documents.⁵⁵ Judge assistants are part of a judge-oriented team; after the quota for judges are determined based on the measurement of the core judgment workload, the quota for judge assistants can be derived proportionally based on the quota for judges. It is neither necessary nor desirable to measure the workload of the judge assistants separately. First, the piloting of judge assistants is still in progress, so relevant workload data are very limited, which does not fit well with big-data analysis; also, the working model between judges and their assistants is not yet fixed. Second, calculating the workload of judge assistants independently would overlook the fact that judges and assistants work as a team and that the impact of judge assistants is eventually evaluated based on the judges' improved capacity in the core judgment work.

55. Zhu, *supra* note 8, p. 197.

6. CONCLUSION

The prerequisite of scientific analysis is selecting approaches based on the nature of the topic. For the reform of China's personnel-quota system for judges, the primary question to answer is: how many judges do we need? Traditional legal research is limited by qualitative analysis. So we decided to seek a different approach, borrowing ideas and methodology from quantitative disciplines such as economy and statistics. At the same time, we realize that no model is perfect, which is especially true when it comes to the quantitative methods in interdisciplinary research. With that being said, a model does represent some aspects of the reality enabling abstraction of the aspects to study the research object more accurately.

As for the modelling of judges' workload, measurement of judges' quota can only be comparatively accurate to the extent to which data are collected. Still, such a kind of measurement is a more reliable and accurate representation of the real demand for the number of judges under dynamic situations, compared to qualitative analysis and simple data comparison based on intuition. Still, we realize that the application of our model is not unlimited because the determination of judges' quota is tied to various aspects of the judicial system and therefore no model can be studied in silos. Judicial reform requires supportive mechanisms to facilitate the establishment of the quota system and we can only truly answer the challenging judicial question of how many judges is enough after the supportive mechanisms are in place.

Judicial reform must start from the essential characteristics of judicial power, dividing judicial work into two functions: judicial function and non-judicial function.⁵⁶ On top of this separation, we proposed methods measuring the judge's workload in a quantitative manner, through participant observation, questionnaires, and interviews. Moreover, by separating the two judicial functions and measuring their workload accordingly, we hope our research can not only provide empirical support for the determination of the number of judicial personnel and the proportion between judges and assistants; we also would like to trigger researchers' attention to judges' core responsibility and hope that more people will join in studying the administration of court personnel and performance evaluation under judicial principles. With the quota matching the workload, the number of judges and assistant staff matching their responsibility, problems such as unbalanced workload or more cases but fewer staff can be avoided; at the same time, outstanding judges could concentrate on the forefront of judicial work. Eventually, a judge-centred resource-allocation framework focusing on the fulfilment of judicial tasks will come to fruition.

REFERENCES

- Alpaydın, Ethem (2010) *Introduction to Machine Learning, Second Edition (Adaptive Computation and Machine Learning)*, Cambridge, MA: The MIT Press.
- Ashley, Kevin (2017) *Artificial Intelligence and Legal Analysis: New Tools for Law Practice in the Digital Age*, Cambridge: Cambridge University Press.
- Chen, Douyun (2014) "The Classified Management of Court Staff and the Practical Perspective of the Reform of Judicial Power." 4 *China Law Rev.* 215–21.
- Chen, Ruihua (2018) "Theoretical Reflection on the System Reform of the Specified Number of Judge." 3 *Contemporary Law Rev.* 1–14.

56. Guo (2017).

- Chen, Weidong (2019) "Litigation Explosion and Courts' Response." 3 *Jinan Journal (Philosophy & Social Science Edition)* 13–22.
- Chen, Yongsheng, & Bing Bai (2016) "Limits in the Reform of Personnel Quota System for Judges and Prosecutors." 2 *Journal of Comparative Law* 21–48.
- Chinacourt.org (2015) "Shanghai Completed the Project of Case Weight Calculation and Applied It in Practice," <https://www.chinacourt.org/article/detail/2015/05/id/1614376.shtml> (accessed 21 September 2019).
- Cui, Yadong (2015) "Solve the Challenge 'More Cases than Court Personnel' by Measuring Maximal Saturated Workload." 1 *Shanghai Judicial Think Tank Legal Reports* 1–15.
- Dong, Bangjun, & Shanshan Huang (2018) "The Alienation Risk and Future Path of the Judge Quota System and Future." 1 *Journal of Hubei University of Police* 14–21.
- Feng, Fei (2015) "The Target and Strategy of the Reform of Judges' Quota." 5 *Contemporary Law Rev.* 140–8.
- Guo, Renhan (2017) "Embedment and Extraction: A Study on the Interest Measurement of the Reform about the Judicial System for Specified Number of Judge Personnel." 2 *Journal of Beijing Administration Institute* 29–34.
- Guo, Song (2013) "Performance Appraisal and Justice Administration." 4 *The Journal of Jiangsu Administration Institute* 123–31.
- Liu, Taoxiong, & Decai Yin (2017) "Big Data Era and the Reform of Research Paradigm in Social Science." 6 *Theoretical Exploration* 27–32.
- Liu, Xiaoyun (2019) "Proposal on the Establishment of Scientific Adjustment for Courts' Staff," http://sohu.com/a/225279801_161795 (accessed 11 August 2019).
- Nlpprogress.com (2019) "Named Entity Recognition," http://nlpprogress.com/english/named_entity_recognition.html (accessed 21 September 2019).
- Poolparty.biz (2019) "What Is a Knowledge Graph?," <https://www.poolparty.biz/what-is-a-knowledge-graph> (accessed 21 September 2019).
- Qu, Xiangdong (2016) "Case-Oriented Staffing and Measurement Model of Judges' Quota." 5 *Modern Law Science* 160–80.
- Su, Li (2010) "Trial Management and Society Management: How Courts Can Effectively Solve the Challenge 'More Cases but Fewer Staff.'" 6 *China Legal Science* 176–89.
- Supreme People's Court (2002) "Opinion about Enhancing the Building of Professional Judge Team." <https://www.chinacourt.org/article/detail/2002/07/id/8900.shtml> (accessed 21 September 2019).
- Supreme People's Court (2004) "Opinion about Piloting Judges' Assistant System in Certain Areas." <https://www.chinacourt.org/article/detail/2004/03/id/109924.shtml> (accessed 21 September 2019).
- Supreme People's Court (2009) "Reply of the Supreme People's Court on the Opinions and Suggestions from Netizens III." <https://www.chinacourt.org/article/detail/2009/12/id/387391.shtml> (accessed 21 September 2019).
- Supreme People's Court (2015) "Opinions of the Supreme People's Court on Comprehensive Deepening of Reform of People's Courts—The 4th Five-Year Outline of the Program for Reform of People's Courts." <https://www.chinacourt.org/article/detail/2015/02/id/1557989.shtml> (accessed 21 September 2019).
- Tan, Suigui, & Sanli Liang (2009) "A Reflection on the Establishment of the Autonomous Judicial Administration: on Issues and Solutions of the Localized Judicial Administration in China, Northern Legal Science." 3 *Northern Legal Science* 69–80.
- Wang, Jing, Xueyao Li, & Zhiyang Xia (2015) "How to Decide the Personnel Quota for Judges: Classification and Measurement based on Civil Caseload." 3 *Law and Social Development* 29–40.
- Wang, Ruilei (2015) "Judicial Reform after the 4th Plenary Session of the 18th CPC Central Committee." 5 *Journal of the Chengdu Municipal Party College of CPC* 108–12.
- Wang, Yaming (2017) "New Study of the Personnel Quota System for Judges." 5 *Research on Rule of Law* 66–77.
- Xu, Weiqiang, & Qilin Hou (2018) "Machine Learning Based Model for Fraud Protection in Consumer Financial." 10 *Modern Management Science* 51–4.

- Yang, Li (2016) “The Critical Issue of China’s Judicial Reform: Example of the Pilot Reform in Shanghai.” 1 *Social Sciences in China Rev.* 53–68.
- Zhou, Qiang (2016) “Advancement of the Modernization of Trial System and Trial Capability Under the Principles in General Secretary Xi Jinping Important Speech Series.” 7 *Journal of Law Application* 1–3.
- Zhu, Jingwen (2007) *Report on China Law Development: Database and Indicators*, Beijing: China Renmin University Press.
- Zuo, Weimin (2017) “Where Does Time Go: An Empirical Study on the Working Time of Criminal Judges in Grassroots Courts.” 5 *Contemporary Law Rev.* 174–83.