

Original Paper

Empirical Academic Platform: Building an Intelligent Infrastructure for Empirical Research in China

Huilin Zhao¹

¹ Shantou University, Shantou, China

Received: March 20, 2026 Accepted: April 11, 2026 Online Published: April 30, 2026

doi:10.22158/rem.v11n1p54

URL: <http://dx.doi.org/10.22158/rem.v11n1p54>

Abstract

Current empirical research faces structural difficulties, and data silos exist in many fields. Also, researchers spend much time on repetitive labor in low level, and scientific resources are wasted because of this. These issues can limit the efficiency of academic innovation and they make achievement transformation difficult in current system. We propose to build an empirical academic platform which has the core function of data integration and automatic standardization of empirical results, with dynamic update, and we use big data and artificial intelligence, plus econometric modeling so that the entire process of empirical research can be automated. Then we describe the theoretical framework, core architecture, technological path, and application scenarios of this platform, and we design a Chinese pilot program which is government led, with university co-construction, and enterprises also participate. We also analyze the challenges and response strategies that the platform faces. It is worth noting that research suggests empirical academic platforms can promote the return of empirical research from method driven to problem driven. They can reshape academic research paradigms. Also, technical support can be provided for the construction of Chinese philosophy and social science system and it has significant theoretical value and practical significance.

Keywords

empirical research, Intelligent platform, Data integration, Automatic verification, pilot scheme

1. The Question of the Times: Efficiency Dilemma and Paradigm Revolution in Empirical Research

When we entered 21st century, empirical research shows its advantages. It can use data support and rigorous logic, so it becomes mainstream research paradigm in social science fields such as economics, sociology, management, and political science. Also, it promotes standardization and internationalization of social science research in China, and it can serve as a carrier for academic innovation and knowledge production. But beneath the prosperous surface of empirical research, we can find some deep and

structural dilemmas. In fact, these problems become prominent now and they are urgent questions that we should address.

Low-level repetitive work is now very common in research. Researchers spend much energy on mechanical tasks, for example data cleaning and model debugging, and this squeezes space for theoretical thinking and new exploration. Also, we find that the data resources are fragmented in current situation. We can observe that ‘data islands’ have formed in many fields and this phenomenon can be seen everywhere. There is a lack of effective integration between official databases, commercial databases, and self-built databases from research institutions, and the formats are inconsistent with poor circulation, which constrains the breadth and depth of research. In fact, scientific research resources are wasted and the transformation of achievements is inefficient because a large amount of valuable research data and empirical conclusions are shelved due to the lack of a sharing mechanism. The barriers between academic research, policy decision-making and corporate practice keep most achievements at level of ‘publication’, making it difficult to transform them into practical social value. These difficulties can weaken the core competitiveness of empirical research. Also, they hinder the essential mission of social sciences to respond to the problems of the times and serve social development.

Because of these problems, we propose to build an Empirical Academic Platform (EAP). It is infrastructure for intelligent chemical technology that can integrate data, verify results automatically, standardize outputs, and it also supports dynamic updates with academic sharing. In fact, EAP can use technology to help change how empirical research is produced and we want to create the scientific ecosystem where people can upload data with one click, select methods with one click, generate results with one click, and share achievements with one click. This platform should not replace empirical researchers. Instead, it can free their hands from tedious work so they can avoid transactional labor and focus on creative work like theoretical innovation, policy recommendations, also problem formulation. The goal is to break data barriers and reduce resource fragmentation. Also, we should reduce repetitive labor and avoid resource waste. Then we can build academic sharing network where every aspect of empirical research generates collaborative value. In this way, the research mode can change from individual inefficient exploration to group efficient innovation.

The construction of EAP can bring new changes to empirical research paradigms and it has importance for theoretical development. Traditional empirical research often has a problem of being “method driven” and researchers pay too much attention to technical details and method selection and this moves away from the core objectives of the research. EAP can help research paradigms return to “problem driven” and we should make “what problem to solve” the starting point of empirical research. Then methods can serve problem solving rather than the opposite. At the same time, by establishing a standardized empirical research process and result presentation system, the platform can reduce academic misconduct such as “method abuse” and “result manipulation” and it can build more fair and transparent academic evaluation standards for social science research. This paradigm shift can reshape value orientation of empirical

research. Also, it can provide support for standardization and sustainable development of social sciences. In fact, this can help academic research return to exploring truth and responding to essence of the times. At practical level, EAP can bring values to support academic development and it can also help social progress. At academic level, the platform should shorten empirical research cycle. Then researchers can devote more time to new thinking which helps improve quality and efficiency of research output. It is worth noting that at policy level, the platform can integrate multidimensional data to provide real-time policy effectiveness evaluation support for government departments and in implementation of national strategies such as rural revitalization and digital economy it can provide scientific empirical basis for policy formulation. At industrial level, the platform can provide customized services including market analysis and operational evaluations for enterprises, and it can promote deep integration of industry and academia, also we expect it can help business decision-making and industrial upgrading with academic achievements. At educational level, the platform can lower entry threshold for empirical research and it can become an important tool for empirical teaching in university, helping students master core methods and enhance practical abilities, and it can cultivate more composite talents with both theoretical literacy and practical skills. In fact, construction of EAP is a response to the efficiency dilemma of empirical research. Also it is a responsibility to serve the development of the social science era and can inject impetus into academic innovation and social progress.

2. The Current Status and Limitations of Empirical Research Tools and Platforms at Home and Abroad

Foreign countries started to develop empirical research tools early. Now they have formed a mature system centered on professional statistical software and open data ecology, and automated analysis modules are also included. But its evolution always faces a core contradiction. This is the conflict between technological accessibility and research inclusiveness. Professional statistical analysis software (such as Stata, R) has become the cornerstone of empirical research because it can support complex econometric models in principle and provide deep technical support, but the high requirements for programming ability and statistical literacy in such tools can raise the entry threshold. So research resources concentrate towards a small group with professional skills and this is not good for inclusiveness. Open data platforms (such as World Bank Open Data) gather massive macroeconomic and social data and they provide data support for cross country and long-term research. But most platforms only stay at data storage and openness. They have not extended to core empirical analysis processes such as online modeling and result verification. This results in a dilemma of “data supply and analysis demand disconnect”. Although some automated tools can achieve data cleaning and basic regression automation, they are difficult to cover complex causal identification methods such as double difference and synthetic control. Also they lack functional design for cross platform data integration and research result sharing. At academic research level, scholars have paid attention to feasibility of automation, and it is worth noting that Angrist and Pischke (2010) pointed out that standardization of econometric methods provides

methodological basis for automation while Athey and Imbens (2019) sorted out application scenarios of machine learning in causal inference. However, existing research mostly focuses on optimizing single technology. It has not yet formed concept of full process automation platform of “data integration model selection result verification sharing and reuse”. Tool development and academic research show characteristics of “decentralization and fragmentation”. In fact, this makes it hard to build a unified system.

The development of domestic empirical research tools shows a clear characteristic that it “mainly relies on introduction and is only supplemented by single point breakthroughs”, which means we currently lack full platforms developed independently and face serious bottlenecks. This situation is very common. We can see bottlenecks in data integration, tool integration, and also academic collaboration. At data level, commercial databases like Wind and CSMAR and official open platforms have accumulated rich materials, but different platforms use different data formats and interface standards, and also there are departmental barriers and security concerns, so the “data silos” problem is serious. Cross platform data integration needs a lot of labor costs. At analytical tools level, domestic research still depends on introducing foreign professional software. Although some universities developed tools for specific fields, these are designed as “single point functions” and they lack systematic support for full process of empirical research, which makes it hard to meet the integrated needs of complex research in current academic environment. This is a big problem. In academic research, domestic scholars focus more on application and local improvement of empirical methods. This results in some problems. For example, they “emphasize methods over problems”, and also there is serious duplication of research, so it is worth noting that AI intelligent agents also can not solve core issues such as inconsistent standards and poor data flow in current stage. Although Chen Qiang (2014) promoted popularization of econometric analysis tools to lower research barriers, existing research has not touched upon core requirements of “full process collaboration+network sharing”, and it also lacks systematic exploration of architecture design, technical path, implementation mode and operation mechanism. So it is hard to support transformation from “tool dependence” to “ecological empowerment”.

When we look at the current development status both domestically and internationally, existing empirical research tools and related studies still face four core limitations, and these problems exhibit systematic characteristics while being interconnected with each other in a complex way. It is worth noting that technical barriers are the first problem. Professional tools have high requirements that can exclude some researchers. This limits popularity of empirical research and it also reduces participation. Data integration is also difficult because data is scattered across different platforms with format differences and high circulation costs, which can limit breadth and depth of research. Then, in terms of process standardization, we lack unified standards for empirical method selection and result presentation, so it is hard to compare and reuse research results and the validation can also be difficult. Finally, sharing mechanisms are weak. Empirical data and code often exist in isolated forms, and they lack efficient sharing channels and form ‘information islands’. This results in resource waste. Also, academic collaborative innovation is hindered.

These limitations are intertwined and they lead to the dilemma of low efficiency, high cost, weak collaboration, and poor reusability in empirical research. It is difficult to adapt to the needs of rapid response to real-world problems in the new era of social sciences.

Existing research and tools provide certain technical foundation for empirical research on automation, but they still have the limitation of “single function optimization” and cannot solve the core problems regarding “full process automation” and “data network collaboration”. This is the starting point of our study. We try to fill the systematic gap in existing research. In fact, the full concept of Empirical Academic Platform (EAP) is proposed for the first time. Four core functions are integrated, including data integration, automatic verification, result standardization, and academic sharing, so that an automated empirical research ecosystem with full functions can be built. Based on national conditions of China, we design a collaborative implementation model with government leading, university co-construction, together with enterprise participation, which can ensure platform combines academic and practical aspects. There are multiple challenges that platform construction may face, such as technological adaptation, academic standards, data security, and long-term operation, and we propose targeted response strategies to provide guarantees for sustainable development of the platform. This study addresses the limitations of existing tools and research. Also, it provides practical solutions for new empirical research paradigms.

3. Core Architecture and Functional Design of Empirical Academic Platform (EAP)

EAP is the first intelligent infrastructure for empirical research in China. It uses “one-stop service” as the core idea and can cover economics, sociology, management, political science, education and other disciplines. Also it can serve the different users such as universities, research institutions, government departments, and enterprises. The core goals include several aspects. We can use the full process automation to improve the production efficiency of empirical research, and this is expected to help researchers save time. Also we should break down the “data silos” so that multi-source resources can be integrated in the platform. It is worth noting that a unified academic standard can be established. At the same time, we build a shared network and this can reduce repetitive labor, which means researchers do not need to do the same work repeatedly. The academic achievements can be connected with policy decisions and enterprise practices. In this way, we can improve the efficiency and optimize the quality of empirical research, and the value transformation can also be achieved.

We use cloud native architecture for the platform, and it relies on distributed computing and big data storage technology, which means the system can provide high concurrency and high availability while service capabilities are scalable to meet different requirements from researchers. Overall architecture follows core logic that four modules should work together (see Table 1). It is worth noting that these include data integration and governance module, automated empirical analysis module, academic sharing and collaboration, and supervision and review module. Each module has its own responsibility and they are interconnected and can form complete empirical research ecosystem covering the full process from

data input to final output and supporting both individual research and group collaboration. Also innovation exploration can be connected with standardized supervision.

Table 1. Core Module Functions and Value Positioning

| Core Module | Core Functions | Value Positioning |
|---|---|---|
| Data Integration and Governance Module | Multi-source data collection, automated cleaning, format standardization, dynamic updating, manual review | The data cornerstone of the platform, ensuring data quality and availability |
| Automated Empirical Analysis Module | Basic statistical analysis, complex causal identification, full-process automation, standardized result output | The technical core of the platform, lowering research thresholds and improving efficiency |
| Academic Sharing and Collaboration Module | Achievement sharing and copyright depository, collaborative research, academic social interaction, thematic activity organization | The ecological core of the platform, facilitating resource reuse and academic collaboration |
| Supervision and Review Module | Data compliance review, academic achievement verification, security protection, behavioral traceability | The security guarantee of the platform, ensuring normative compliance and rights protection |

The data integration and governance module provides basic support for the platform operation, and we collect data in multiple ways including official docking, commercial cooperation, user upload (including user made indicator data), and other methods that can cover both macro data from government departments and segmented data from commercial providers. At the same time, users can upload their own compliance data. In fact, this creates data pool with rich types. In data governance process, we combine automated processing and manual review, where machine learning algorithms identify missing values, outliers, duplicate records, and other problems in the raw data before different formats are converted to standard formats with supporting explanations. Then, a professional team does sampling and review. This is expected to ensure data accuracy and compliance. Also, the platform sets up a dynamic update mechanism that sets different update frequencies for different data types and supports users to update data through manual or automatic methods, which can ensure data timeliness in principle. We design the automated empirical analysis module as the technical core of platform. It can realize the intelligence for whole process from data input to result output, which means users can get final results without complex manual operations and the system handles everything in between by using built-in algorithms and predefined workflows. At basic analysis level, users do not have to write code. In fact, they only need to select data and variables, and then the system can generate descriptive statistics, correlation analysis, entanglement and other results with one click, and at the same time it produces standardized tables and visual charts that can be used in papers and at the causal identification level the platform includes basic methods such as OLS regression and Logit/Probit models. Also, it has mainstream quasi natural experimental methods. These include double difference method, propensity score matching method, breakpoint regression method, and instrumental variable method. Users can choose according to their research questions. Then the platform completes model setting, parameter estimation, and it can do robustness testing. This is expected to ensure reliability of the results. At result

output level, the platform generates bilingual empirical reports that include textual explanations and standard tables, together with visual charts, and the textual part covers core content such as research questions, data sources, method selection and result interpretation in detail which can help readers understand the whole analysis. The table format follows core journal publishing standards. It supports multiple formats for export and can meet the needs of different usage scenarios such as academic papers and research reports.

The academic sharing and collaboration module is the key part of the platform. It can help change empirical research from individual work to group collaboration. In achievement sharing part, users can upload data and methods related to empirical analysis, and they can also upload various reports to the platform, which means these resources can finally form a standardized knowledge base that can be used for empirical research. Other users can obtain these resources through multi-dimensional search. They can cite and comment on them. Also the bookmark function is provided. It is worth noting that we use blockchain technology in platform. It can provide copyright certification for achievements and protect intellectual property rights of users. For collaborative research, users can create project teams and assign different role permissions to members, and real-time synchronization of research progress can be realized while the system tracks modification records through the whole process, which means we build in online discussion functions so team members can communicate ideas. In academic socialization part, users can follow peer researchers. They can view research results and communicate privately. In fact, the platform holds online academic seminars and it provides empirical research training. These activities can strengthen the academic exchange atmosphere and they can promote cooperation across different disciplines and cross institutional boundaries so that the overall research level can be improved.

The supervision and audit module can provide full security support for platform operation. It is expected to help achieve data compliance and result standardization, and we also try to make sure that user behavior can be kept legitimate during the whole platform operation process, because this is important for maintaining the system security. In data review process, we check the authenticity and compliance. In fact, user uploaded data should be verified, and false data, confidential data and data that infringes rights should not be uploaded by users because these data types can cause serious problems for the platform security and user privacy protection. Sensitive data is anonymized using technology so that it can be used safely. In results review process, a professional academic team conducts review for uploaded empirical results, and they focus on verifying the accuracy of methods used and check if the interpretation of results is rational, which can help avoid academic problems. Academic misconduct such as plagiarism is prohibited. Non compliant results will be returned and users should modify them. Also, for serious violations, we can cancel the account and report to the unit. In terms of security protection, the platform adopts encryption technology. It can help ensure the security of data storage and transmission, and user account security is protected through multiple identity authentication mechanisms, and we establish a log audit system to record operational behavior so that full traceability and supervision can be achieved, and we can build a solid defense line for the healthy and sustainable operation.

This platform has formed its own competitiveness through distinct advantages. It is worth noting that the whole process is automatic, which can reduce the difficulty for the empirical research and make scientific research more universal in principle. Also, we integrate multi-source data to break “data silos” and this provides support for research work. In fact, unified academic standards run through entire process and can help improve result quality. Then, sharing mechanism is built to promote resource reuse, and it can reduce repetitive labor and can avoid resource waste. At the same time, multiple security measures are used to balance data security and intellectual property protection, establishing a solid compliance bottom line, and modular design supports flexible expansion of functions and can be upgraded according to disciplinary development and user needs, so it can adapt to the long-term development trend of empirical research.

4. Technical Path and Feasibility Analysis of Platform Construction

We build EAP based on the core logic of ‘multi technology collaboration and full process empowerment’, and we integrate big data, artificial intelligence, measurement modeling, and blockchain together to form a supporting system that can provide technical foundation for the whole process in different application scenarios. At the same time, we verify technology maturity and we use cost control methods and funding schemes. Then feasibility of platform construction and operation can be ensured. It is worth noting that technological path focuses on ‘stable support’, ‘intelligent efficiency’, and we also consider ‘secure compliance’ as important goal, where cost and resource planning is tailored to actual needs and can be adjusted according to different situations. This can provide full support for platform implementation.

4.1 Core Technology Support System: Multi Technology Collaboration Empowers the Entire Process

EAP system is built on four core technologies (see Table 2). Each is integrated. In fact, these technologies cover data processing, empirical analysis, security assurance and other key links, and we can use them to build a full technical support chain.

(1) Big data technology. As the foundation of data integration and storage, we adopt a distributed storage solution. It can adapt to different data types. Structured data is stored in relational databases, and semi-structured data is managed by document databases. Unstructured data is put in distributed file systems. In fact, we use data sharding technology to optimize storage structure based on disciplines, themes, research directions, and other dimensions. Query and processing efficiency can be improved in this way. Also, a streaming computing framework is used to receive real-time data, clean it, and then store it. This can ensure data timeliness. At the same time, we combine big data visualization technology to transform empirical results into intuitive charts and optimize user experience.

(2) Artificial intelligence and machine learning technology. This is main driving force in automated empirical research. In fact, in the data governance, algorithms can process missing values and identify outliers and this is the first step. Then in model selection, we use automated machine learning technology to recommend suitable empirical methods based on the research questions and data features and we also include a causal inference toolkit which can improve the accuracy of causal recognition. It is worth noting

that in parameter optimization, algorithms such as grid search and Bayesian optimization are used to help find the optimal parameter combination, and the empirical results are expected to be reliable.

(3) Measurement modeling engine. We need to support empirical analysis for academic research. So we use a modular design. It includes descriptive statistics and regression analysis. Also causal inference is supported. We base algorithm logic on professional statistical software and it can be validated by academia. In fact, we should establish mechanism to update algorithms and it can incorporate latest research results in field of econometrics, which means platform methods can maintain scientific nature and stay up-to-date in principle.

(4) Blockchain technology. We should strengthen the defense line of data security and copyright protection. It is worth noting that we adopt alliance chain architecture to build a trusted network, and generate hash values for results uploaded by users to be stored on the chain, so the intellectual property rights can be traced in principle. In fact, we use a hybrid architecture for data privacy protection. Also, combined with zero knowledge proof technology, data is expected to be “available but invisible”. Then we use smart contracts to pre-set profit distribution rules and this can help incentivize data sharing.

Table 2. Application Scenarios and Supporting Value of Core Technologies

| Core Technology | Key Application Scenarios | Core Supporting Value |
|----------------------------------|---|---|
| Big Data Technology | Multi-source data storage, real-time data processing, empirical result visualization | Supporting massive data storage, efficient circulation and intuitive presentation |
| AI & Machine Learning Technology | Intelligent data governance, automated model selection, parameter optimization, automatic robustness test triggering | Lowering research thresholds and improving the efficiency and accuracy of empirical analysis |
| Econometric Modeling Engine | Standardized empirical method encapsulation, frontier algorithm iteration, multi-model adaptation | Ensuring the academic standardization and frontier nature of empirical analysis |
| Blockchain Technology | Copyright depository of achievements, data privacy protection, shared revenue distribution, empirical result traceability | Safeguarding data security and intellectual property rights, and incentivizing academic sharing |

4.2 Technical Feasibility Verification: Migration and Adaptation of Mature Technologies

The core technologies required for EAP have a mature application foundation, and successful practices in related fields provide direct reference for platform construction, fully verifying the technical feasibility. In fact, big data technology is mature. Existing official platforms and commercial databases can manage massive data in integrated way and we can migrate their distributed storage and streaming computing solutions to EAP. For artificial intelligence and machine learning, automated machine learning platforms can realize automated model construction. Also, causal inference toolkits are used in academic research. It shows that automated empirical analysis is feasible in principle. At level of metrological modeling, the algorithm logic of professional statistical software is mature. It provides solid foundation for developing the platform’s metrological modeling engine. Blockchain technology has achieved data security certification and copyright protection in scientific research data sharing platforms and its technical

solution can adapt to the needs of EAP. Also, domestic technology enterprises provide distributed computing, artificial intelligence, blockchain and other services. These can serve as the foundation for platform technology support, and it can ensure the stability and scalability of the platform.

4.2.1 Cost Estimation: Phased Control of Resource Investment

We estimate the cost of EAP based on principle of on-demand investment and efficient utilization. In fact, it is planned in two stages including initial construction and long-term operation and this approach can in principle help us allocate resources in reasonable way so that we can avoid waste.

In initial construction stage, the cost is mainly about platform building. It is worth noting that it includes technology development, server setup, data integration, and personnel training. For technology development, we need to build cross disciplinary R&D team including engineers and econometricians, and our work focuses on architecture design, module development, also algorithm optimization should be considered. Server setup needs hardware equipment and computer room facilities, and they should meet requirements of massive data storage and high concurrency access, which means the configuration must satisfy demands from large-scale data processing and simultaneous user access at the same time. Data docking involves cooperation with official and commercial databases, and standardization work should be done. In fact, personnel training aims to provide professional guidance to pilot unit users and technical support teams, and it can help the platform be implemented and used after launch.

The long-term operating costs are mainly about how to keep platform running continuously and stable, and we should consider all the aspects that can affect the platform performance when we make the budget plan for system in next few years. In fact, we can divide these costs into four main parts: server maintenance, data updates, personnel compensation, and technology upgrades. Server maintenance includes hardware depreciation, the data center operation and maintenance, and network bandwidth guarantee. It is worth noting that data updates involve data procurement, cleaning work, and interface maintenance and it can ensure data timeliness. Personnel compensation covers technical maintenance, data review, academic review, and operational management teams. Technology upgrades focus on algorithm optimization, functional expansion work, and security protection upgrades and this part should adapt to the development trend of empirical research.

4.2.2 Fundraising: Diversified Models Ensure Sustainable Operation

In fact, the construction and long-term operation of platform need stable funding. We propose a funding plan combining government subsidies and research funding support, and market-oriented operation can also be used. It can form a stable funding guarantee system.

Government subsidies should be main funding source in initial stage. We can seek policy support from multiple departments to cover construction costs. Also, research funding is obtained through working with key universities and research institutions to apply for national level research project funding, which can support technology development and pilot promotion. After the platform operates stably, market-oriented operation generates revenue through value-added services, including commercial data customization, authorization of advanced empirical analysis functions, and also enterprise consulting

services, and these can cover long-term operating costs and form a sustainable operating model. It is worth noting that this model should prioritize public welfare and supplement the market.

In general, EAP technology path can depend on mature technology system and cost planning that is scientific and reasonable. Also, the funding channels are diverse and can support the development. It is worth noting that whether we consider the technology implementation and resource investment, or the long-term operation is considered, it can show sufficient feasibility and this provides a foundation for the platform landing and promotion.

5. Application Scenarios and Value Realization of the Platform

EAP uses data integration, automatic verification, and also standardized sharing as core features, which can help build connection between academic innovation and social applications, and we cover four main areas: academic research, government decision-making, enterprise practice, and talent cultivation. It is worth noting that the value is not from simple addition of single functions. Instead, it comes from reconstruction of production and circulation logic in empirical research, and customized solutions can be provided for different subjects. It promotes the return of academic research to innovation essence. Also, we help various sectors achieve scientific decision-making and value maximization. In fact, this can form a virtuous ecology where academic empowerment of practice and practice feeding back academia happen at the same time.

In academic research, EAP has a key value. It can free the research productivity and can help academic innovation, so the quality and efficiency can be improved. The platform uses full automation. In fact, researchers do not need to do repetitive work like data cleaning and model debugging, so they can focus on creative work including theoretical innovation and problem posing, and the research cycle can be shortened. We build an ‘empirical research knowledge base’. It can break the isolation of academic achievements, so researchers can retrieve and reuse datasets and research methods from others, and they can also check the empirical conclusions. This can avoid repetitive research and improve resource use efficiency. At the same time, we establish unified empirical standards and strict academic review mechanism in platform. It is worth noting that this can standardize key aspects like method selection and result presentation, and curb academic misconduct such as method abuse and result manipulation. This provides guarantee for the development with standardization and high quality in academic research, and promotes a research atmosphere of ‘innovation first, sharing, collaboration and mutual learning’ in the community.

EAP works as a key bridge. It connects data and decisions in government and enterprise settings. For government departments, platform can integrate multidimensional data and this provides empirical evidence for policy formulation. Also it can assist in designing differentiated policy plans. Then after policy implementation, we evaluate the effects using mainstream causal inference methods. It is worth noting that these results can be used as a reference for policy adjustment and optimization. When sudden public events happen, the system uses real-time data processing to analyze different response measures,

and this can improve the timeliness of emergency decision-making. For enterprises, the platforms provide research support for market entry by integrating market data and user information. In daily operations, we can evaluate advertising effects. Also in industrial upgrading and technological innovation, analyzing industry trends and policy orientations provides empirical references for transformation directions. This helps enhance core competitiveness and can support the deep integration of industry, academia, and the research.

EAP has low threshold and practical characteristics. It can be used as a new tool for empirical research teaching in universities, and we find it helps cultivate composite talents. In fact, students do not need professional programming knowledge and they do not need to know statistics in advance, and the platform can complete the whole empirical analysis process from data selection to result output through “one click operation”. This reduces the entry threshold and students can understand the core logic of empirical methods. In teaching practice, students can choose research topics by themselves, use the data resources and model library in the platform to conduct empirical research, and participate in the entire process from problem posing to result interpretation. This strengthens the combination of theory and practice. It also helps cultivate new thinking and research design abilities. At the same time, the platform allows students to participate in the entire process of data processing and analysis, helps them establish the data awareness, and master the core methods of data interpretation. They can enhance data literacy necessary in digital age. This is expected to lay a solid foundation for future academic research or career development, and can promote alignment between talent cultivation in universities and social needs.

6. Challenges and Response Strategies Faced by the Platform

EAP is a system engineering project crossing technology and domain boundaries, and it also involves multiple academic subjects, so construction and promotion of such platform can face multiple challenges in four dimensions: technology integration, academic adaptation, data governance, and long-term operation. These challenges are intertwined and can influence each other, so we should focus on core positioning of the platform and adopt systematic response strategies with clear targets, and we should promote continuous improvement of the platform when solving difficult problems. It is worth noting that we can achieve organic unity of technical feasibility and academic standardization. Also, operational sustainability should be ensured.

There are technical challenges in our research. Platform should support collaborative research in different fields and keep adaptability. On one hand, we need to integrate big data, artificial intelligence, econometric modeling and blockchain, which can help achieve automated complex causal inference and secure data sharing. But it is hard to combine empirical methods from different disciplines, and balancing data sharing and privacy protection is also difficult. On the other hand, when user base and data volume grow, the platform should in principle handle high concurrency access and storage pressure, while different disciplines and users have differentiated needs, which can place higher demands on stability and scalability. For response strategies, we establish collaborative innovation mechanism between

industry, academia and research and we can use advantages from all parties to carry out core technology research and development. It is worth noting that we adopt step-by-step implementation and gradual optimization R&D logic. First we tackle core challenges. Then we improve details. We rely on cloud native architecture and distributed computing technology to enhance elastic scalability and use modular design to achieve on-demand loading of functions, and we establish real-time performance monitoring mechanism to solve performance bottlenecks in time.

We face a key challenge at academic level. It is increasing acceptance in the academic community and ensuring new value. Some researchers have resistance to automated empirical research because they fear it can reduce professionalism of research and there are big differences in empirical paradigms among different disciplines which makes it hard for platform standardization process to adapt to special needs. At the same time, automated processes can lead to homogenization of empirical results. Some users may rely on the platform too much and overlook theoretical construction. This affects depth of results. It is worth noting that our response strategy focuses on “empowering rather than replacing”. We strengthen publicity through academic conferences and special lectures. Also we clarify the value of the platform in freeing up the hands of researchers. It can help them focus on theoretical innovation. We establish an academic advisory committee and involve renowned experts from various disciplines in platform design. Then we customize specialized functional modules. This can enhance disciplinary adaptability. In fact, we carry out pilot demonstrations to validate value of the platform with high-quality academic achievements. At the same time, we establish incentive mechanism and set up “innovation achievement zone”. It guides researchers to deepen theoretical thinking and expand methods on empirical basis of the platform and avoid homogenization.

Data level faces several challenges. We should consider data integration quality and also security assurance for compliance, because these two aspects can affect system performance in practical applications. Multi-source data comes from different places. It has different formats and quality is not the same, also update frequency varies. This makes integration difficult, and when users upload data there can be authenticity issues and accuracy problems. In fact, data copyright ownership is complex. This can lead to intellectual property disputes, and when we share sensitive information there is risk of privacy leakage. For response strategy, we establish standards for multi-source data integration. We should sign agreements with data providers to clarify quality requirements and update frequency, so that data source can be managed. Then we improve governance mechanism of “automation+manual review”. It is worth noting that we review user uploaded data and establish data quality evaluation system, which can ensure overall quality of data. We should clarify data copyright ownership. Blockchain technology can provide copyright certification for user achievements, and this method should protect rights of data owners. For sensitive information, we use data anonymization and encrypted storage. Access control should also be used. These technical means should comply with relevant laws and regulations on data security and personal information protection, which can ensure data processing compliance.

We face two main challenges in operation level. The first is long-term cost control and the second is increasing user engagement. Platform server maintenance, data updates, personnel salaries and other operating costs can be high, and it is hard to achieve sustainable operation when we only rely on government subsidies and research funding. At the same time, if users only use the platform for specific needs and lack sustained motivation, the ecosystem construction and long-term development can be affected. For the response, we explore a market-oriented operation model that should prioritize public welfare and use market as supplement. While we ensure the public welfare attributes, revenue can be generated through customized services, commercial data authorization, value-added services and other commercial cooperation. Also, we should establish a diversified funding mechanism that can integrate government subsidies, research funding, social donations and other resources, and then we optimize cost structure to reduce operating expenses. We optimize user experience continuously and establish a normalized feedback mechanism, which can help us respond to needs promptly, and we also build incentive systems such as point redemption, achievement promotion, academic certification and other rewards to enrich academic communities, online seminars, training courses and other services. In this way, we create a platform ecosystem with practical value and good academic atmosphere, which can enhance user stickiness and activity.

7. Research Conclusion and Prospect

Current empirical research faces some structural problems. For example, there are “data silos” and low-level duplication of work, which wastes resources and makes achievement transformation hard. We propose to build an Empirical Academic Platform (EAP) to handle these issues. We construct the theoretical framework. Also, the core architecture and implementation path are discussed in this paper. EAP is an intelligent chemical infrastructure including data integration, automatic verification, result standardization, and academic sharing. In fact, it can help empirical research return from “method driven” to “problem driven”. Also, technology is used to improve research productivity. We can build a collaborative ecosystem at the same time. The platform has four core modules. They work together and has a good performance in intelligence, standardization, security, and scalability, so it can meet needs in academic research, government decision-making, enterprise practice, and talent cultivation. We have mature core technology support and diversified funding schemes, and the pilot model of “government led, university co construction, enterprise participation and social collaboration” can be used, so the construction and promotion of EAP should be feasible. It is worth noting that its implementation can provide technical support for construction of the Chinese characteristic philosophy and social science system. Also, it can help implement the innovation driven development strategy.

In future, EAP should develop further to support research. We can examine three main directions in this work. At technical level, we should strengthen the integration of artificial intelligence with econometrics and introduce new causal inference algorithms, and blockchain can be used for full process traceability in research while we explore how emerging technologies such as metaverse affect empirical research

models. It is worth noting that these technologies are still new. At the application level, we should break the boundaries of social sciences and extend to natural sciences. Also, scenario based functions including policy simulation and enterprise risk assessment can be deepened according to practical needs. Then we connect with international databases and academic resources. This can help promote platform internationalization in global context. At the ecological level, we improve the academic sharing and collaboration mechanism. We can build a global empirical research community and establish talent training bases with universities, and we should participate in formulating global empirical research standards based on platform practice, and this can enhance China's academic discourse power. In fact, the results can be used as a reference for future development.

Intelligentization and standardization are trends in empirical research we cannot avoid in development of the times, and data sharing is also necessary. Building EAP is a complex task that needs multi-party collaboration among the government, universities, research institutions, enterprises, and also researchers who are the main users. We call on the government to provide policy and funding support, and include the platform in national key planning so that the construction can get enough resources and attention. Universities and research institutions can contribute data and academic wisdom, and they rely on platforms to carry out new research and talent cultivation, which can help improve the overall research capability in the long term. Enterprises use their technological and market advantages, participate in research and development operations, and market promotion. Researchers should use platforms and share achievements. Then we can promote the innovation of empirical research paradigms together. It is worth noting that the task ahead is arduous. However, with joint efforts of all parties, EAP can become the core infrastructure for empirical research in China, and it can promote the development of academic research with good quality. Also, we can contribute Chinese solutions and wisdom to global empirical research.

References

- Angrist, J. D., & Pischke, J. S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2), 3-30.
- Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685-725.
- Chen, Q. (2014). *Advanced Econometrics and Stata Application* (2nd ed.). Beijing: Higher Education Press.

Acknowledgements

This article has received funding from the Ministry of Education's Humanities and Social Sciences Research Youth Project (23YGC790197) and the Shantou University Research Start up Fund Project (STF23007).