Trends in Social Sciences and Humanities Research
Print ISSN: 2959-992X
Online ISSN: 2959-9938
DOI: https://doi.org/10.61784/tsshr3181

DYNAMIC EVALUATION OF OPPORTUNITY INEQUALITY AND POLICY OPTIMIZATION VIA THE HST-GBM MODEL

Han Wu*, RuoShi Wang, YuQi He
School of Statistics, Tianjin University of Finance and Economics, Tianjin 300222, China.
Corresponding Author: Han Wu, Email: 18722176929@163.com

Abstract: In the context of China's rapid economic development and structural transformation, inequality of opportunity has become a major obstacle to upward mobility and the achievement of common prosperity. This study proposes an innovative Hybrid Spatiotemporal Gating-Boosted Model (HST-GBM), integrating LSTM, XGBoost, fixed-effects models, and gated recurrent units, to address the limitations of traditional methods in capturing nonlinear, heterogeneous, and intertemporal dynamics. Using CFPS panel data from 2010 to 2022, the study examines the evolution of opportunity inequality and its key drivers. First, the model effectively quantifies the relative impact of individual effort and environmental factors, revealing a long-term trend of declining inequality and increasing influence of personal effort. Second, the analysis shows that opportunity structures are highly responsive to policy changes, with effort-related impacts intensifying during reform windows. Finally, the study offers targeted policy recommendations that emphasize adaptive algorithms and effort-oriented approaches. Overall, this research contributes by advancing methodological innovation, enriching the theoretical understanding of dynamic inequality, and providing empirical support for policies aimed at promoting social equity and long-term stability.
Keywords: Inequality of opportunity; Common prosperity; LSTM; XGBoost; HST-GBM

1 INTRODUCTION

As China undergoes rapid economic development and social transformation, inequality of opportunity has become a pressing issue, limiting social mobility and hindering the goal of common prosperity. Despite ongoing reforms in education, income distribution and social welfare, structural disparities across urban–rural, regional and social dimensions persist. Emerging forces like the digital economy have further widened these gaps, highlighting the need for dynamic assessments of inequality and its policy implications.

Scholars have long debated the sources and mechanisms of opportunity inequality. Roemer first laid the theoretical foundation by distinguishing "circumstance" from "effort"[1], and later refined it into the more systematic "type–effort" decomposition within theories of distributive justice[2]. In the Chinese context, Zhang, Lin, and Li analyzed inequality of opportunity in basic education, showing how access and institutional arrangements shape outcomes[3]. Building on this, Chen and Zhou emphasized new forms of digital inequality, revealing how technological disparities reinforce stratification[4]. From the perspective of common prosperity, Shi, Chen, and Fang highlighted that low-income groups face structural disadvantages in opportunities, which accumulate throughout the life course[5]. Similarly, Guo and Li examined the link between educational opportunity inequality and social mobility, finding that structural disparities persist across cohorts and regions[6].

Quantitative approaches further deepened this research: Wan, Zhang, and Tang re-estimated income opportunity inequality with machine learning, showing that circumstance factors often outweigh effort in explaining disparities[7]. Complementing this, Liu and Han investigated the interplay between social mobility and educational equity, providing evidence that opportunity inequality is closely tied to income distribution and broader economic gaps[8]. Beyond income, Liu and Zhao identified new trends in educational inequality and proposed governance strategies to reduce institutional barriers and enhance fairness[9]. Finally, Xue extended the classic "ascription–achievement" model, demonstrating how effort interacts with ascriptive factors in occupational attainment, thereby signaling the need for more flexible and dynamic analytical frameworks[10].

Despite these advances, most prior studies adopt static perspectives or rely on single-model approaches like XGBoost, limiting their capacity to capture evolving structural patterns. To address this, this study employs longitudinal microdata from the China Family Panel Studies and proposes a Hybrid Spatiotemporal Gating-Boosted Model (HST-GBM). The model integrates Long Short-Term Memory networks for time-series learning, gating mechanisms for adaptive variable selection and XGBoost for enhanced prediction and interpretability. By incorporating fixed effects and spatiotemporal embeddings, it enables multi-level modelling across time and regions. Compared to traditional methods, HST-GBM improves accuracy, robustness and causal inference, though at higher computational cost.

Using this framework, the study tracks the evolution of opportunity inequality and evaluates the impact of policy reforms across various stages of reform. The main contributions are threefold: first, it develops an innovative machine learning model that overcomes the limitations of static or single-model methods; second, it identifies the dynamic mechanisms underlying opportunity inequality; and third, it provides evidence-based, policy-relevant insights to support China's pursuit of common prosperity and long-term social stability.

2 DATA PREPROCESSING AND DESCRIPTIVE STATISTICS

Before proceeding with the analysis, we first describe the data sources and then detail the preprocessing steps applied to ensure data quality and consistency.

2.1 Data Source

This study primarily utilizes data from the China Family Panel Studies, covering seven waves conducted biennially from 2010 to 2022. The CFPS is administered by the Institute of Population and Labor Economics at the Chinese Academy of Social Sciences. All relevant datasets were obtained from the official website and compiled for the corresponding years. A summary of the data is provided in Table 1.

Table 1 Overview of Data Sources

Data Name	Data Type	Data Description	Source
CFPS 2010–2022 Data	Mixed Panel Data	National longitudinal household survey data from 2010 to 2022	Institute of Population and Labor Economics, Chinese Academy of Social Sciences
China Statistical Yearbook 2010–2022	Mixed Panel Data	National statistical yearbook data from 2010 to 2022	National Bureau of Statistics of China

2.2 Data Preprocessing

Given the presence of missing values, outliers, and inconsistent formats in the dataset used for this study, these issues may negatively affect data quality and model inputs. Therefore, comprehensive data preprocessing is essential to ensure that the dataset meets analytical requirements, thereby improving the accuracy and reliability of the modeling results.

This study addresses various data issues as follows: Outlier handling was performed using the 3σ rule to remove observations beyond ±3 standard deviations from the mean. For missing data, variables with high missing rates were excluded. Linear interpolation was applied to time-series data, while for cross-sectional data, K-nearest neighbors (KNN) clustering was used to group similar observations, and decision trees were trained within each group to predict missing values. The “years of education” variable was binned into ordered categories based on the standard duration associated with the highest degree obtained. Categorical variables were converted into numeric form using ordinal encoding for ordered variables and label encoding for binary or unordered ones. Standardization methods varied by indicator type: Min-Max was used for driving indicators, negative Min-Max for constraints, and Z-score for balance indicators. Household expenditure data were transformed into proportions of total spending. Finally, data from the China Family Panel Studies (2010–2022) were merged using household IDs, with duplicate entries containing complementary information consolidated through matching and sorting procedures.Table 2 reports the descriptive statistics.

Table 2 Summary of Descriptive Statistics

Variable	Sample Size	Mean	Std. Dev.	Min	Max
Age	43,613	43.264	14.162	10	95
Regional GDP per Capita	43,613	71,100.614	26,671.835	45,000	190,000
Education Expenditure	43,613	2,606.846	4,900.858	0	80,000
Subjective Health Score	43,613	3.091	1.299	1	5
Medical Expenditure	43,613	3,171.046	5,547.973	0	155,000
Number of Unhealthy Members in Household	43,613	0.458	0.679	0	5
Educational Attainment	43,613	2.148	1.301	0	7

3 MODEL DEVELOPMENT AND EMPIRICAL ANALYSIS

To achieve both structural interpretation and dynamic identification of opportunity inequality, this study develops a dual empirical approach that integrates indicator system construction and mechanism modeling. By distinguishing between contextual factors and individual effort variables, the study introduces the HST-GBM algorithm, which possesses strong capabilities for temporal modeling.

3.1 Model Description

This study proposes the Hybrid Spatio-Temporal Gated Boosting Model (HST-GBM) to capture the dynamics and generation mechanisms of opportunity inequality while balancing predictive performance, structural interpretability, and spatiotemporal generalization. Traditional methods like XGBoost and LightGBM are effective for static data but limited in capturing temporal dependencies and spatial heterogeneity. Conditional Inference Forests, while strong for causal analysis, face challenges in dynamic modeling and scalability. HST-GBM addresses these limitations by integrating a Long Short-Term Memory (LSTM) network for encoding temporal features, a gating mechanism for dynamic feature selection, and XGBoost for non-linear prediction. Additionally, the model incorporates panel fixed effects to account for unobserved heterogeneity and a gated residual-enhancing mechanism to adjust prediction errors.

A dynamic weight-adjustment layer ensures adaptive model fusion. Hyperparameters are fine-tuned using grid search with early stopping to optimize performance. This integrated approach provides a unified framework for multi-scale prediction of opportunity inequality, combining the strengths of deep learning and structured machine learning methods.

3.2 Construction of the Indicator System

Based on Roemer's "circumstance–effort" framework, this study classifies opportunity inequality into contextual and effort-based factors. To capture their dynamic effects, a structured indicator system is developed, with two-period lags included for both types of variables.

3.2.1 Construction of environmental indicators

This study examines how social attributes and structural conditions contribute to opportunity inequality by constructing a set of environmental indicators. Gender is defined as a constraining factor, as women often face systemic disadvantages in education, employment, and household resource allocation. Age, drawing on the life-cycle theory, is also considered a constraint due to declining access to opportunities over time. Hukou status reflects institutional stratification under China's dual urban-rural system; individuals with rural hukou face structural disadvantages in education, healthcare, and social security.

In contrast, regional GDP per capita is treated as an enabling factor, reflecting access to economic resources. Urban residence indicates relative advantages in education and employment, and is therefore also considered enabling. Household education expenditure reflects investment in human capital and contributes to reducing inequality. Self-reported health status, medical insurance coverage, and healthcare expenditure all serve as indicators of health security and are categorized as enabling factors. Finally, the number of unhealthy household members reflects health burdens that constrain opportunity structures and is treated as a constraining indicator.

3.2.2 Construction of effort-based indicators

Effort-based indicators include educational attainment and current employment status. Education reflects human capital accumulation; under social mobility theory, higher education increases access to resources and opportunities, making it an enabling factor. Employment status captures labor market integration, which not only improves economic conditions but also facilitates upward mobility, and is therefore also considered an enabling factor.

3.2.3 Weight assignment using the generalized entropy method

To quantitatively assess opportunity inequality across provinces and over time, this study adopts the Generalized Entropy Index (GEI) to decompose its structural sources. We classify the influencing factors into two dimensions: contextual (environmental) and effort-based (individual behavior), and aggregate variables within each dimension using weighted combinations. These composite indicators are then embedded into the GEI framework to measure and decompose inequality by source. The GEI allows sensitivity adjustment through a parameter α; for this study, we use $\alpha = 1$ to balance interpretability and model tractability.

Finally, we define a combined inequality index as the weighted sum of inequality attributed to environmental and effort-based factors. The resulting structural weights indicate that contextual and individual factors contribute nearly equally to overall opportunity inequality, accounting for 50.48% and 49.52%, respectively. This suggests that both structural conditions and individual efforts jointly shape the evolution of opportunity inequality across regions and over time.The results are presented in Table 3.

Table 3 Indicator Weight Allocation

Variable Type	Weight
Environmental Indicators	50.48%
Effort-Based Indicators	49.52%

3.2.4 Construction of lagged indicators

This study uses one-period and two-period lagged variables to explore the time-lag effects of environmental and effort-based factors on opportunity inequality. Environmental variables typically have delayed effects, while the cumulative impact of effort-based factors may emerge later. A Pearson correlation heatmap (Figure 1) shows that some lagged variables are weakly correlated with their original counterparts, suggesting potential non-linear relationships and confirming the need for lagged variables in the analysis.

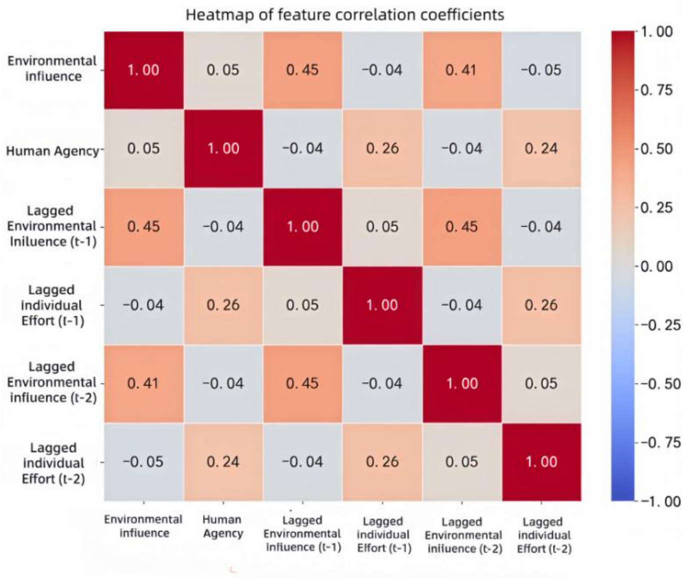

Figure 1 Heatmap of Feature Correlation Coefficients

4 EMPIRICAL ANALYSIS AND RESULTS

This section analyzes CFPS data using the HST-GBM model to examine the evolution and drivers of opportunity inequality from 2010 to 2022. It assesses the relative impacts of environmental and effort-based factors, tracks annual fluctuations and their policy roots, and identifies the causes of major changes. The findings offer empirical support for understanding inequality mechanisms and inform future policy and research.

4.1 Analysis of the Impact of Indicator Importance

This section analyzes the importance of each indicator in the model, identifying key variables that explain inequality of opportunity. By quantifying their impact, it reveals the main drivers of opportunity inequality and guides future policy interventions. The distribution of indicator importance is shown in the table 4 below.

Table 4 Distribution of Indicator Importance

Indicator	Effort Factors	Environmental Factors	Effort Lagged One Period	Effort Lagged Two Periods	Environmental Lagged One Period	Environmental Lagged Two Periods
Importance	70.26%	26.81%	0.57%	0.53%	1.14%	0.66%

The results from Table 4 show that individual effort, with a weight of 70.27%, is the dominant factor in explaining inequality of opportunity, surpassing environmental factors (26.82%). This indicates that personal investments in education and skills development play a central role in shaping opportunities. Environmental factors, though important, have a more indirect and constrained impact, influenced by demographic, geographic, and health conditions. Lagged variables show minimal impact, with diminishing effects over time. Overall, while individual effort remains crucial, environmental factors and past conditions still exert long-term influence. Policies should focus on enhancing education and skill development while optimizing environmental conditions to create a more inclusive society and reduce opportunity inequality.

4.2 Trend Analysis of Equality of Opportunity

According to the results of the model analysis (as shown in Figure 2), the level of equality of opportunity in China exhibited a clear upward trend from 2010 to 2022, with the mean value rising from 0.418 in 2010 to 0.494 in 2022. This change indicates substantial progress in promoting fair opportunities across social groups during this period. Notably, significant fluctuations in 2014 and 2018 suggest that policy shifts and socioeconomic factors had a phased impact on the state of equality.

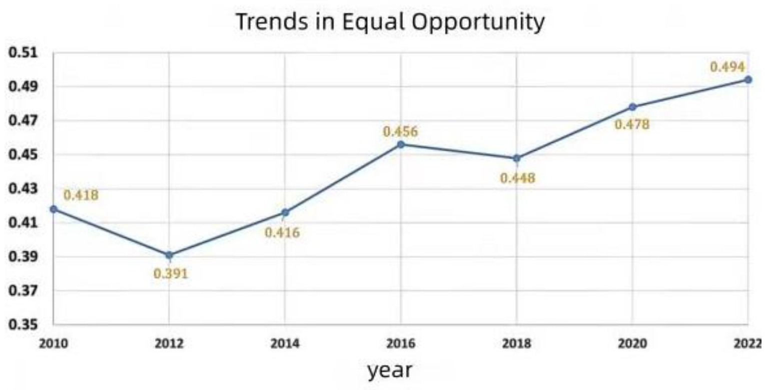

Figure 2 Trends in Equal Opportunity

These findings suggest that over the past decade, inequality of opportunity in China has been significantly reduced, reflecting the effectiveness of government policies. In particular, reforms in education, improvements in the social security system, and adjustments in income distribution have facilitated a more equitable allocation of social resources, gradually narrowing opportunity gaps between different social groups. This progress has not only enhanced social equity but also expanded the space for upward mobility across society.

4.3 Rate of Change Analysis

This section analyzes the rate of change in inequality of opportunity and its key influencing factors—namely, environmental conditions and individual effort—to reveal temporal trends and the dynamic effects of each variable over time (as shown in Figure 3). Rate of change analysis offers a more intuitive view of the relative speed and direction of variation among different variables, thereby deepening the understanding of the long-term mechanisms driving opportunity inequality. By calculating the annual change rates of each feature, the study further identifies both the phased and directional characteristics of their influence on inequality of opportunity.

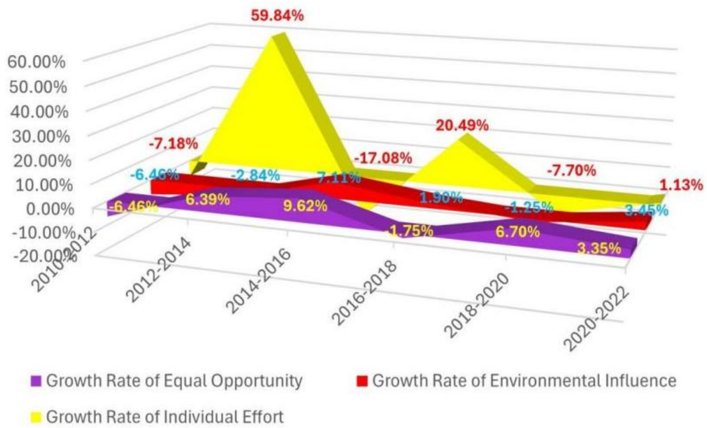

Figure 3 Three-Dimensional Line Plot of Change Rates Across Key Variables

The rate of change in opportunity equality has fluctuated over time, with notable increases during 2014–2016 (9.62%) and 2018–2020 (6.70%), likely reflecting major policy reforms—especially the deep structural changes in 2014. In contrast, the 2020–2022 period (3.35%) showed more moderate growth, suggesting stabilization in policy impact. Environmental factors remained relatively stable, indicating that deep-rooted structural conditions like family background and hukou continue to influence opportunity access. This points to a more mature policy system that mitigates but has not yet fully transformed structural inequality.

In contrast, individual effort showed significant variation, peaking in 2012–2014 (59.84%) and 2016–2018 (20.49%), highlighting the increasing role of personal initiative—particularly in education and employment—in shaping opportunity. These shifts underline the effectiveness of policy incentives in enhancing individual motivation and social mobility.

Overall, from 2010 to 2022, opportunity inequality in China has declined steadily. Effort has become the primary driver of improved access, while environmental influences have weakened but persist. The evolving opportunity structure reflects the combined effects of policy reform, institutional support, and personal agency.

5 CONCLUSION

This study constructs a Hybrid Spatiotemporal Gating-Boosted Model (HST-GBM) to address the complexity of variable interactions and nonlinear structures in high-dimensional data. Empirical findings reveal a significant shift in the drivers of opportunity inequality in China between 2010 and 2022, with individual effort surpassing environmental factors as the dominant influence. This shift reflects the growing effectiveness of institutional reforms and policy interventions in enhancing the role of personal initiative. The overall level of opportunity inequality has declined steadily—especially after 2014—while change rate analysis highlights the system's high responsiveness to policy, with individual effort increasing notably in certain years and environmental effects remaining relatively stable.

Based on these findings, it is recommended to expand the application of adaptive models like HST-GBM in public policy evaluation to improve the analysis of complex social dynamics. Policymakers should continue to promote effort-oriented strategies by improving access to education, strengthening vocational training, and supporting human capital development. In parallel, reforms aimed at education equity, basic public service equalization, and urban–rural integration should be advanced to reduce structural disparities. Lastly, a responsive policy feedback system should be developed to monitor evolving effects of key variables and enable timely, targeted policy adjustments for enhanced social equity.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Roemer J E. Equality of Opportunity. Cambridge: Harvard University Press, 1998.
[2] Roemer J E. Theories of Distributive Justice. Cambridge: Harvard University Press, 2012.
[3] Zhang N, Lin J, Li J. Research on inequality of opportunity in basic education. China Industrial Economics, 2020(08): 42-59.
[4] Chen M, Zhou Y. New progress in digital inequality research. Economic Review, 2022(04): 123-139.
[5] Shi X, Chen T, Fang S. Common prosperity from the perspective of equal opportunities: Evidence from low-income groups. Economic Research Journal, 2022(09): 99-117.
[6] Guo Q, Li Y. Educational opportunity inequality and social mobility research. Nanjing Journal of Finance and Economics, 2022(01): 1-14.
[7] Wan X, Zhang C, Tang L. Re-estimation of income opportunity inequality in China: New findings from machine learning. Journal of Quantitative & Technical Economics, 2024(01): 192-212.
[8] Liu X, Han Y. Social mobility and educational equity: Theory and evidence. Social Science Research, 2024(01): 110-128.
[9] Liu J, Zhao J. New trends of educational inequality in China and governance strategies. Journal of Educational Economics Review, 2025(02): 45-62.
[10] Xue Y. How much does effort matter for occupational attainment? An extension of the "ascription–achievement" model. Population & Economics, 2025(04): 89-101.