

## Research Article

# Empirical Analysis of Beneficial Equality of the Basic Medical Insurance for Migrants in China

Jieying Gao <sup>1</sup>, Dongxiao Chu <sup>1</sup> and Tao Ye <sup>2</sup>

<sup>1</sup>*School of Finance, Capital University of Economics and Business, Beijing 100070, China*

<sup>2</sup>*School of Banking and Finance, University of International Business and Economics, Beijing 100029, China*

Correspondence should be addressed to Tao Ye; [yetao@cueb.edu.cn](mailto:yetao@cueb.edu.cn)

Received 15 May 2021; Accepted 7 June 2021; Published 17 June 2021

Academic Editor: Daqing Gong

Copyright © 2021 Jieying Gao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The fairness of the benefits of basic medical insurance for the migrants is drawing increasing attention. This paper examined the beneficial equality of the basic medical insurance for the floating population in China using the “2014 National Internal Migrant Dynamic Monitoring Survey.” The Heckman model was employed due to selection bias among inpatients, and the random forest algorithm of machine learning was used to analyze the importance of factors affecting the hospitalization decision-making, hospitalization consumption, and reimbursement proportion of the floating population. The results show significant differences in the fairness of basic medical insurance benefits among different income groups, and the highest-income group benefits the most. In contrast, the higher-income group benefits the least. Further verification by introducing the commercial medical insurance indicated that the differences among different income groups did not disappear but reduced the degree of difference among the groups. Although China’s healthcare reform has progressed greatly, the study’s findings confirm that the government’s fair medical insurance can lead to unfair problems and the phenomena of low-income groups subsidizing high-income groups under the equalized basic medical insurance system. Adjusting the design of equalized medical insurance and allowing different income groups to pay different premium levels according to the payment level may be more conducive to the fairness of benefits based on achieving universal health coverage in China.

## 1. Introduction

Population mobility has become the main feature of China’s economic and social development. China has the largest migrant population globally, which increased significantly from 50 million in 1990 to 236 million in 2019 based on the census data. The migrant population has become the main source of China’s labor market, influencing the country’s economic growth and social stability significantly [1]. However, due to their self-capacity and social policies, the migrants have worked and lived in terrible environments, engaged in unsanitary and dangerous work, and are vulnerable to injuries and disease, resulting in their health status not being effectively guaranteed [2–4]. Medical insurance could improve people’s medical utilization and health status by reducing the economic threshold for medical treatment and increasing medical services availability [5]. In recent years, many studies

discovered that patients with medical insurance had a higher utilization rate for outpatient, inpatient, and healthcare [6–9], while patients without medical insurance might give up or delay treatment due to the heavy financial burden [10–13]. Additionally, evidence shows that patients with medical insurance could improve their health outcomes [14–16].

Like many other countries worldwide, basic social health insurance (SHI) systems have been gradually established since the 1980s in China. One of the primary goals of the SHI in China is to ensure that citizens should not be denied quality healthcare services due to financial hardship [17, 18]. In 2015, more than 95% of the Chinese population was covered by three basic social health insurances, including the New Rural Cooperative Medical Scheme (NRCMS) for rural residents, Urban Resident Basic Medical Insurance (URBMI) for urban workers in the informal sector and unemployed urban residents, and Urban Employee Basic Medical Insurance

(UEBMI) for employees in the formal urban sector [19]. The SHI has played an important role in increasing health service accessibility, reducing economic burden, and improving health [20, 21]. The proportion of out-of-pocket (OOP) payments among total medical consumption decreased from 60% in 2000 to 28% in 2016 [22]. Many studies showed that different medical insurance systems had achieved different implementation effects. For example, the NRCMI increased outpatient and inpatient services utilization but did not reduce out-of-pocket expenditures [23–25], the UEBMI and URBMI improved the utilization of preventive medical services, and the UEBMI could additionally reduce the OOP ratio and improve people's health status [19].

The original intention of the basic medical insurance system was to ensure that every insured person's basic medical needs are effectively met. The fairness of benefits is also the basic principle of the basic medical insurance system design. Therefore, every insured person who pays the same premium should enjoy the same proportion of reimbursement compensation, namely, the same costs and the same benefits. The equalization system should ensure that all insured persons also have equal opportunities to benefit and even persons in low-income groups enjoy equal basic health services. However, under the equalized medical insurance system, the more people use health services, the more medical insurance compensation they receive, and the more they benefit. Some studies indicated that basic medical insurance compensated different income groups differently, which benefited the wealthy more [26–28]. The basic medical insurance should balance the welfare improvement and efficiency loss by medical compensation and not result in excessive medical demand due to the insured persons' medical security improvement [29, 30]. Therefore, from the perspective of fairness, basic medical insurance should effectively make up for the medical shortage faced by the low-income groups and balance medical resource utilization among different income groups rather than being more inclined to the rich.

Little work has been done on fairness within different individuals' incomes after the implementation of the basic medical insurance system. More attention has been paid to the comparison before and after the implementation of the policy. Additionally, it was found that there were very few studies on the equity of medical insurance benefits for urban residents. However, there is almost no literature on equity for the floating population, which plays a huge role in China's urbanization. This paper's primary focus is to discover whether the migrants have access to equal basic medical insurance benefits and whether the poor subsidize the rich under the current basic medical insurance system among the floating population. But fairness is a concept that includes value judgment, and different standards may result in different evaluations. It is necessary to define fairness in the evaluation of system effects. Many factors affected the unfairness of basic medical insurance benefits, including personal health, age, and social factors, for example, income, education, and insurance. Social factors were the main cause of unfairness [31–34], and the difference in medical compensation caused by social factors is unfair. Among these

factors, which ones are more important? Previous literature used  $p$  values to judge the significance of influencing factors, and few literature about the utilization of health-care used machine learning methods to discuss the importance of variables [35, 36]. Random forest is a tree-based ensemble learning model, and it is widely used to solve classification and regression problems. The importance of variables can be ranked according to the best variable selected in the decision tree as the classification node [37]. Based on the random forest algorithm, this paper explores the relatively important factors that affect the insured's hospitalization decision-making behavior and hospitalization reimbursement behavior and discusses the fairness of medical insurance benefits for the floating population from the perspective of income differences. Furthermore, according to the Health China 2030 planning outline, a multilevel, complete, and mature medical security system must be supplemented by commercial health insurance. There was little study done on commercial health insurance regarding the equity of benefits from basic medical insurance for the floating population. Another goal of this study is to explore whether commercial health insurance affects the beneficial equality of basic medical insurance for the floating population.

Based on the above foundation [34], this paper defines fairness as follows: the floating population with different income levels should enjoy the same health rights and medical compensation right, which could be explained using the incidence of disease and hospitalization, the total expenditure of hospitalization and the accessibility to being hospitalized, and the reimbursement ratio of medical insurance among the insured migrants with different-level incomes. The main aim of this study proceeds as follows: First, based on the fact that medical insurance is characterized as mainly compensating and reimbursing hospitalization expenses, this paper examines whether the insured migrants with different income levels who have been hospitalized equally benefit from the basic medical insurance or not in two ways: hospitalization expense compensation and hospitalization possibility. Second, it discusses the differences in the insured groups' health status in terms of the incidence of diseases requiring hospitalization, which can influence the insured's beneficial equality. Thirdly, this paper analyzes the utilization of medical services and the applicability of medical insurance for the insured to discover the main possible reason why the fair benefit of basic medical insurance is impacted. Finally, the insured people who also have commercial medical insurance are added to the basic medical insurance groups. The study explores the impact of commercial health insurance on the equality of benefits for the insured with different income levels by analyzing those who have enrolled in the basic medical insurance and commercial health insurance simultaneously.

## 2. Materials and Methods

*2.1. Data Source.* The data in this paper is obtained from the China Migrants Dynamic Survey in 2014. According to the

principle of randomness, the survey data adopted a stratified, multistage, and scale proportional PPS sampling method. Samples were obtained from 31 provinces (autonomous regions or municipalities) and the Xinjiang Production and Construction Corps, where the floating population was relatively concentrated. A questionnaire survey was conducted among 200,937 respondents who were 15–59 years old from nonlocal district (county or city) households who lived in the place of inflow for more than one month. The questionnaire included the migrant population's health, medical insurance, reimbursement, detailed demographics, household income, and expenditure information. This study mainly examined the fairness of the benefits of basic medical insurance for the floating population. After deleting the samples with free medical care, no basic medical insurance, and rural women's reimbursement subsidies that account for 100%, a total of 173733 respondents with basic medical care were chosen. Because of the limited use of data, only the survey in 2014 included the commercial medical insurance data, and there is no such related data in other years. Therefore, this study had to choose the survey in 2014, however, which does not affect the final conclusion. The problem in this study pertained to the fairness of the benefits of basic medical insurance for migrants. The key outcome variables listed in Table 1 were the total medical expenditure of hospitalization and reimbursement amount for basic medical insurance according to prior literature [1, 19, 38]. Table 1 shows that the mean age of the migrants was about 29 years, 63% were male, and 77% were married, and the majority of residents were from rural areas (85%) and had educational qualification of junior high school or below (66%). On average, the cost of hospitalization for the insured with basic medical insurance was about 9342 yuan, and the proportion of reimbursement through medical insurance exceeded 50%.

## 2.2. Methods

**2.2.1. Heckman Model.** On studying the difference in basic medical insurance benefits of the insured with different income levels, there was a phenomenon that many insured received no reimbursement when they suffered from hospitalization. One reason was that there might be deductibles, and the other was that the insured might also give up the reimbursement opportunity due to cumbersome reimbursement procedures or the problem of reimbursement in different places. The above phenomenon might result in the sample's random error, which no longer obeys the normal distribution, and the ordinary least squares (OLS) estimation would lead to estimation bias. Accordingly, the Heckman selection was used to obtain unbiased estimates [39]. The Heckman two-stage model is suitable for solving endogenous problems caused by sample selection bias. The Heckman sample selection included two parts of estimation. The first step was to select the estimation process to estimate the factors that affected the insured's medical behavior, shown as follows:

$$\Pr(Y_i = 1 | \text{Income}, X, Z) = \phi \left( \sum_{j=1}^5 \text{Income}_{ij} \vartheta_j + X_i \beta + Z_i \gamma + u_i \right), \quad (1)$$

where  $Y_i$  is a dummy variable indicating whether the hospital medical insurance reimbursement amount for respondent  $i$  was greater than 0 in the past 12 months. If their reimbursement amount is greater than 0,  $Y_i$  is 1; otherwise it is 0.  $\text{Income}_{ij}$  is a set of dummy variables indicating the income level group of respondent  $i$ . Here, the groups are based on a five-point scale with the possible values of 1 (lowest income), 2 (lower income), 3 (middle income), 4 (higher income), and 5 (highest income), respectively.  $X_i$  is a vector of individual characteristics such as age, sex, marital status, education, and employment status.  $Z_i$  is a vector of city and year dummies, used to control the unobserved heterogeneity in time and geographical environment, where  $\phi(\cdot)$  is the standard normal probability density function.  $u_1$  is a vector of random errors. The second step was to predict the amount for medical insurance reimbursement using a general linear model to estimate the nonzero amount. The reason for choosing the general linear model was that the reimbursement amount and total medical expenditure of hospitalization had a heavy fat tail on the left side. However, their logarithmization was closer to a normal distribution, as shown in Figure 1, which better met the assumptions.

The underlying mechanism takes the following form:

$$\text{Log}(\text{Reimburse\_amount}_i) = \sum_{j=1}^5 \text{Income}_{ij} \vartheta_j + W_i \beta + Z_i \gamma + \varphi_i, \quad (2)$$

where  $\text{Reimburse\_amount}$  is greater than zero, and there is no relation between the residual terms of models (1) and (2). The household registration factor is taken as an exclusion condition to improve the model's estimation validity. It will affect whether the insured patients are hospitalized or not and then influence their reimbursement behavior. However, it will probably directly affect the specific amount of reimbursement. The remaining variables in  $X_i$  are the same as those in  $W_i$ .  $\varphi_i$  is a vector of random errors.

Additionally, this study mainly examines the fairness of the benefits for the floating population in the basic medical insurance system. It also involves influencing factors, such as the health status of the insured and the utilization of medical insurance. The incidence of hospitalization with diseases can replace the health status, that is, the question, "In the past 12 months, how many hospitals have you been in?" If the answer is more than zero times, the incidence is recorded as 1 and 0 otherwise. The applicability of medical insurance is determined by the medical reimbursement ratio. If the ratio is greater than zero, it is recorded as 1, indicating that it is applicable; otherwise, it is recorded as 0. The health status and applicability of medical insurance of the insured can be proposed by probit as follows:

TABLE 1: Definition and statistics for the selected variable.

Variable and type	Definition	Mean	SD	Percent (%)	Overall
Total medical expenditure of hospitalization (continuous variable)	Last hospitalized medical expenses (yuan)	9342.037	14297.87	—	2165
Reimbursement (continuous variable)	Reimbursement amount for basic medical insurance	4793.706	8152.47	—	2165
Income (category variable)	The average monthly total income of household locally and divided into 5 equal parts from low to high: lowest = 0 (reference); lower = 1; middle = 2; higher = 3; highest = 4	—	—	23.36 (0)	35329 (0)
				16.97 (1)	25665 (1)
				21.18 (2)	32033 (2)
				18.49 (3)	27961 (3)
				19.99 (4)	30219 (4)
Working status (category variable)	Employment status: employee = 1; employed = 2; self-employed = 3; other = 0 (reference)	—	—	1.79 (0)	2710 (0)
				57.93 (1)	87599 (1)
				8.77 (2)	13261 (2)
				31.50 (3)	47638 (3)
				20.78 (0)	31421 (0)
Marital status (category variable)	First married or remarried = 1; divorced or widowed = 2; unmarried = 0 (reference)	—	—	77.04 (1)	116496 (1)
				2.18 (2)	3291 (2)
				13.07 (0)	19767 (0)
				53.02 (1)	80173 (1)
				20.54 (2)	31053 (2)
Education (category variable)	Education level: primary school or below = 0 (reference); junior high school = 1; senior high school = 2; specialist college = 3; college or above = 4	—	—	8.66 (3)	13089 (3)
				4.71 (4)	7126 (4)
				36.96 (0)	55887 (0)
				63.04 (1)	95321 (1)
				85.21 (0)	128841 (0)
Age	Range of 15 to 59 years old	28.871	9.574	0.237	2.390
Male (category variable)	Female = 0 (reference); male = 1	0.630	0.483	36.96 (0)	55887 (0)
				63.04 (1)	95321 (1)
Hukou (category variable)	Rural = 0 (reference); nonrural = 1	0.148	0.355	14.79 (1)	22367 (1)
				85.21 (0)	128841 (0)
Com_insurance (category variable)	Insured by commercial medical insurance, com_insurance = 1; other, com_insurance = 0 (reference)	—	—	95.37 (0)	144192 (0)
				4.63 (1)	6996 (1)

Note. The numbers in parentheses in columns (5) and (6) represent the category in the table.

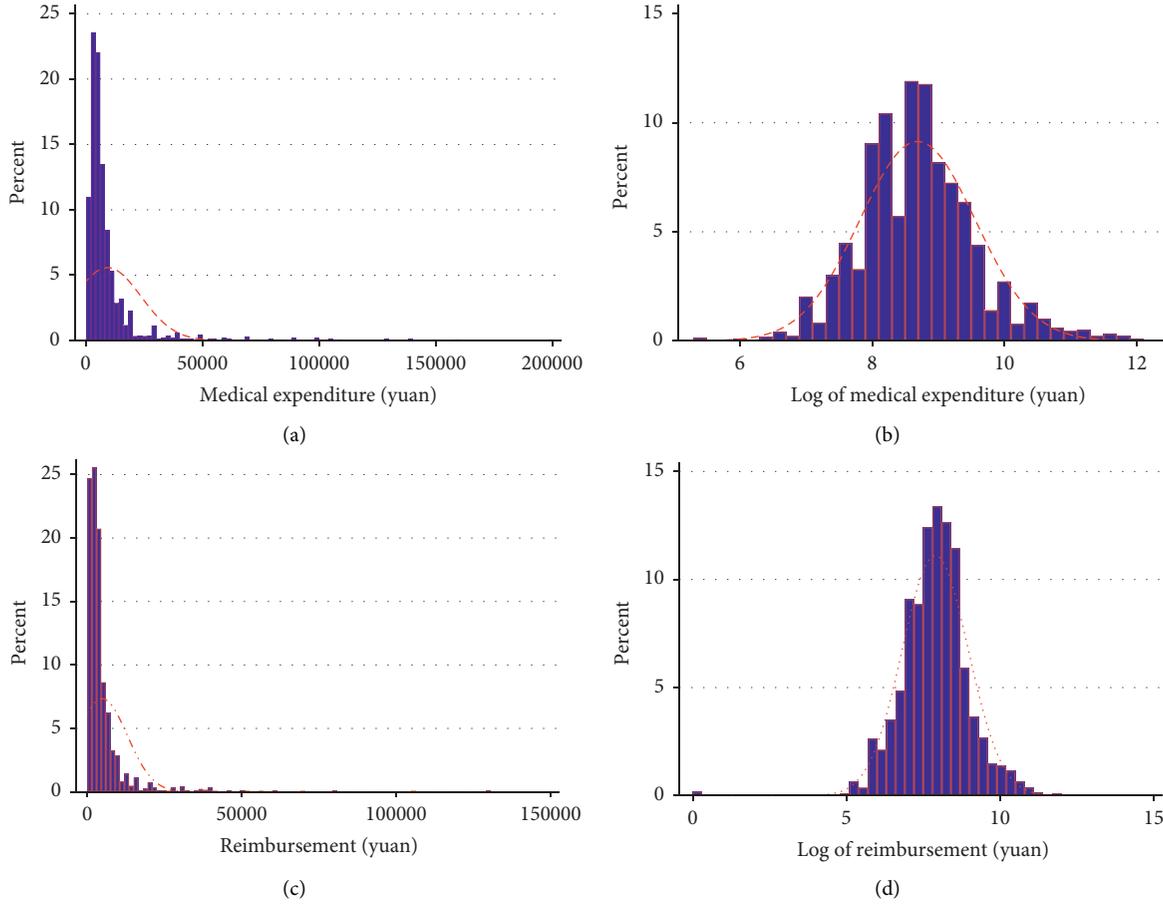


FIGURE 1: Histogram of medical expenditure and reimbursement amount. (a) Total medical expenditure of hospitalization. (b) Log of total medical expenditure of hospitalization. (c) Reimbursement amount of basic medical insurance. (d) Log of reimbursement amount of basic medical insurance.

$$\begin{aligned}
 & \text{Profit}(\text{Health\_Dummy}_i = 1 \mid \text{Income}, X) \\
 &= \sum_{j=1}^5 \text{Income}_{ij} \vartheta_j + W_i \beta + Z_i \gamma + \delta_i, \\
 & \text{Profit}(\text{Health\_Suitable}_i = 1 \mid \text{Income}, X) \\
 &= \sum_{j=1}^5 \text{Income}_{ij} \vartheta_j + W_i \beta + Z_i \gamma + \zeta_i,
 \end{aligned} \tag{3}$$

where  $\delta_i$  and  $\zeta_i$  are vectors of random errors.

**2.2.2. The Principle of Random Forest Model Measuring Importance.** The random forest model can be applied to both classification and regression tasks. In the two-step model, the hospitalization decision-making behavior belongs to the two-class model, and the hospitalization reimbursement behavior belongs to the regression model, so the random forest model is suitable for this situation. It uses the bagging algorithm to randomly sample the sample information to generate multiple training sets and then uses a decision tree as the base classifier for each training set and uses the majority voting results of multiple trees as the final prediction value [40]. According to the best variable selected

in the decision tree as the classification node, the importance of the variables is sorted. The random forest model can be applied to both classification and regression tasks. In the two-step model, the hospitalization decision-making behavior belongs to the two-class model, and the hospitalization reimbursement behavior belongs to the regression model, so the random forest model is suitable for this situation. It uses the bagging algorithm to randomly sample the sample information to generate multiple training sets and then uses a decision tree as the base classifier for each training set and uses the majority voting results of multiple trees as the final prediction value. In the random forest model, it is difficult to quantify the direction and extent of the influence of the change of the explanatory variable on the dependent variable  $Y$  with specific numerical values, but the importance of the variables can be ranked according to the best variable selected in the decision tree as the classification node. The specific principle is that every time bagging is used to resample to build a decision tree, some samples will not be selected. These data are out-of-bag data (denoted as obb); then, these samples can be used for cross-validation. The so-called out-of-bag data mean that each time a decision tree is established, one piece of data is obtained by repeated sampling for training the decision tree. At this time, about 1/

3 of the data is not used and does not participate in the establishment of the decision tree. This part of data can be used to evaluate the performance of the decision tree and calculate the prediction error rate of the model, which is called out-of-bag data error. According to the cross-validation error as the scoring basis,

- (1) for each decision tree, obb is used to calculate out-of-bag data error, which is recorded as error-obb1,
- (2) then, noise interference is added to feature  $j$  in the out-of-bag data randomly, and the out-of-bag data error is calculated again, which is recorded as error-obb2,
- (3) suppose that there are  $n$  decision trees; then, the importance of feature  $j$  is

$$\text{Out - of - Bag\_Score} = \frac{\sum_{i=1}^N (\text{error\_obb}_{2,i} - \text{error\_obb}_{1,i})}{N} \quad (4)$$

In the second step, if the random noise leads to a significant increase in the error of out-of-bag data, it shows that this feature has a greater impact on the prediction results of the established random forest model, which means that the variable is of high importance.

**2.3. Descriptive Analysis.** Table 2 reports the hospitalized medical expenditure of the migrants participating in the basic medical insurance (greater than 0), the hospitalization reimbursement rate, and the incidence of diseases requiring hospitalization. The results showed that the average hospitalization expenses for those enrolled in medical insurance are 9295.942 yuan, and the actual average reimbursement ratio is 55.2%. In the quintile income level sample, it can be found that the insured population has the highest hospitalization expenses in the lower-income group, the higher-income group is the second, and the highest-income group is the lowest. However, the ratio of reimbursement for hospitalized medical expenditure is the opposite. The high-income groups had a higher reimbursement ratio than the low-income groups on average. The highest-income group had the highest reimbursement ratio, reaching 58.251%, and the low-income group had the lowest reimbursement ratio, only 52.697%. The fairness differences in the floating population with basic medical insurance from Table 2 remained. It appeared that the lower-income group consumes the highest hospitalization expenses but enjoyed the lowest medical reimbursement ratio. Nevertheless, the highest-income group consumed the lowest hospitalization expenses and owned the highest medical reimbursement ratio. According to the above analysis, the difference between the total hospitalization cost and the medical reimbursement ratio might be due to the medical insurance system. It might be caused by different groups having different health rights. This study used the possibility of being hospitalized to represent differences in access to health rights. It was found that people with lower incomes had relatively less access to

health rights, which further illustrated the differences in health access to the benefits of different income groups.

### 3. Results

**3.1. Factors of the Basic Medical Insurance Compensation Level and Hospitalization Accessibility of the Insured.** Table 3 shows the influencing factors of the basic medical compensation level and hospitalization accessibility of the insured floating population, using the Heckman two-step model. The Wald test indicated that the medical insurance reimbursement model had sample selection bias at the 1% level. Therefore, the study selected the hukou as the excluded variable to distinguish the selection and outcome models. Among control variables, increased age, being male, and being employed were significantly associated with lower reimbursement. Additionally, higher levels of education were significantly associated with higher medical insurance compensation. Controlling for other factors that were unchanged at the level of 1%, there were significant differences among the medical compensation levels of different income groups. The medical insurance compensation received by any other age group was significantly lower than that received by the highest-income group. Column 4 shows that the highest-income group was significantly more likely to be hospitalized than any other income group.

It can be seen from the two aspects of medical insurance compensation level and hospitalization accessibility that have been discussed that people with the highest-income levels were most likely to be hospitalized, which means that they have more opportunities to use medical insurance and more likelihood of receiving medical insurance compensation. However, even if they were hospitalized, the medical insurance compensation for the highest-income group was more than that of other income groups. Therefore, the results suggested a problem of unfair medical insurance benefits, which is especially beneficial for people with incomes above 20%.

**3.2. The Importance of Factors Affecting Medical Behavior and Hospitalization Reimbursement for the Floating Population.** Random forest algorithm is used to rank the importance of the characteristics that affect the insured's hospitalization medical consumption, reimbursement proportion, and hospitalization possibility. The importance of seven factors, marriage status, education level, age, gender, income, working status, and registered residence difference, is shown in Figure 2. This result is consistent with the conclusion that the statistical model is basically consistent and significant in Table 3. According to Figure 2, marriage has the most important impact on the insured, which may be due to the fact that about 77% of the data represent the status of married, and marriage has a certain protective effect on the health of the floating population, thus affecting people's medical behavior. Education level has a greater impact on the proportion of reimbursement, while age and gender have a relatively small impact on the degree of reimbursement. Income level and working condition are relatively important

TABLE 2: The insured sample hospitalization and health information.

Variable	Income level	Mean	SD	Skewness	Kurtosis	Sample size
Hospitalized medical expenditure	Lowest income	8859.077	13580.2	5.050	34.989	364
	Lower income	11205.2	20874.27	5.752	42.506	278
	Middle income	9289.12	13531.02	4.898	32.498	416
	Higher income	9721.834	15747.9	4.987	34.228	415
	Highest income	8543.306	10214.15	5.792	53.587	725
	Total	9295.942	14206.34	5.901	50.378	2198
Percentage of hospitalized expenditure reimbursement (%)	Lowest income	53.548	90.250	14.897	255.297	364
	Lower income	52.697	26.271	89.468	5.772	278
	Middle income	54.019	29.373	2.779	23.849	416
	Higher income	54.178	40.143	3.928	28.345	415
	Highest income	58.251	38.366	5.130	51.884	725
	Total	55.2	48.880	16.795	496.718	2198
Incidence of hospitalized diseases (%)	Lowest income	1.008	9.990	9.807	97.194	36,103
	Lower income	1.062	10.249	9.550	92.205	26,186
	Middle income	1.275	11.219	8.686	76.450	32,630
	Higher income	1.457	11.982	8.103	66.656	28,486
	Highest income	2.355	15.163	6.285	40.496	30,792
	Total	1.425	11.854	8.196	68.168	154,197

The unit of expenditure indicator is yuan.

TABLE 3: Factors of the medical compensation level and being hospitalized accessibility.

Variable	Heckman model		Possibility of hospitalization
	First step (1)	Second step (2)	
<i>Male</i>			
Male = 1	-0.4014*** (0.0177)	-0.3588*** (0.0226)	-0.4048*** (0.0178)
Age/10	-0.6179 (0.0602)	-0.5530*** (0.0700)	-0.6160*** (0.0603)
Age square/100	8.7709*** (0.9432)	7.9400*** (1.0908)	8.7391*** (0.9442)
<i>Education</i>			
Junior high school	-0.0509 (0.0277)	-0.0400 (0.0310)	-0.0541 (0.0277)
Senior high school	0.0880** (0.0319)	0.1135** (0.0355)	0.0805* (0.0321)
Specialist college	0.3169*** (0.0371)	0.3434*** (0.0414)	0.3124*** (0.0384)
College or above	0.4411*** (0.0427)	0.5347*** (0.0468)	0.4433*** (0.0453)
<i>Marriage status</i>			
First married or remarried	0.8634*** (0.0402)	0.7929*** (0.0494)	0.8623*** (0.0402)
Divorced or widowed	0.6423*** (0.0766)	0.5958*** (0.0873)	0.6500*** (0.0763)
<i>Working status</i>			
Employee	-0.0039 (0.0645)	0.0240 (0.0713)	-0.0156 (0.0642)
Employed	-0.0896 (0.0709)	-0.1655* (0.0785)	-0.0996 (0.0706)
Self-employed	-0.0213 (0.0657)	-0.0761 (0.0726)	-0.0242 (0.0654)
<i>Income</i>			
Lowest income	-0.0662* (0.0285)	-0.1106*** (0.0317)	-0.0759** (0.0286)
Lower income	-0.1372*** (0.0294)	-0.1402*** (0.0330)	-0.1402*** (0.0294)
Middle income	-0.1378*** (0.0261)	-0.1347*** (0.0291)	-0.1402*** (0.0261)
Higher income	-0.1319*** (0.0262)	-0.1604*** (0.0292)	-0.1385*** (0.0263)
<i>Hukou</i>			
Nonrural	-0.0319 * (0.0141)	—	-0.0208 (0.0264)
Constant	-1.7123*** (0.1105)	-1.5294*** (0.1335)	-1.6919*** (0.1104)
Year dummies	Yes	Yes	Yes
City dummies	Yes	Yes	Yes
Wald test (rho = 0) $\chi^2$ (1):		2143.94***	—
Observations	151999	2198	154197

Standard errors are in parentheses; \*  $p < 0.05$ , \*\*  $p < 0.01$ , and \*\*\*  $p < 0.001$ .

to reimbursement proportion and hospitalization possibility of floating population. Part of the reason is that higher education level often leads to better work and higher income, better medical security, and being more conducive to its full

use of medical resources [1, 13]. The characteristics of household registration are the least important, perhaps because China is gradually breaking the dual-structure system of urban and rural areas and is constantly promoting

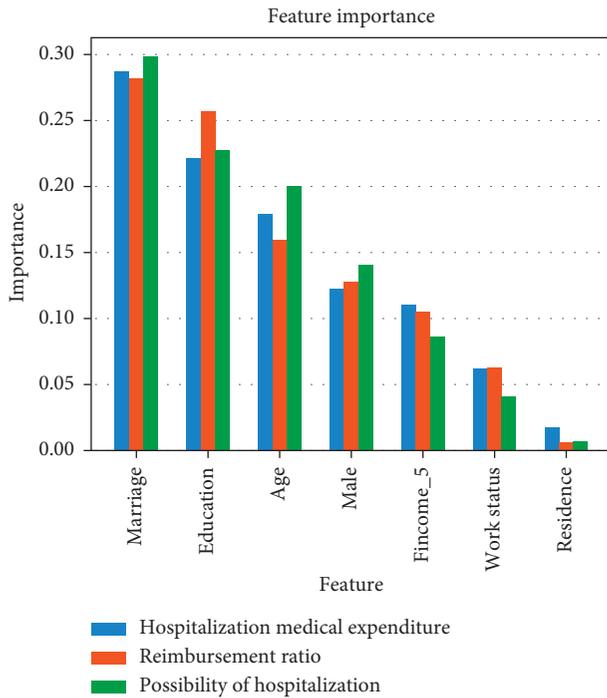


FIGURE 2: Ranking of the importance of influencing factors.

the medical reform policies such as settlement and reimbursement in different places.

**3.3. Possible Influencing Channels.** Table 4 contains the analysis of the potential channels of unfair benefit from the basic medical insurance system for the floating population from the medical service utilization (i.e., the total hospitalization medical expenses) and the possibility of a basic medical insurance application. As is evident from Table 4, the total cost of hospitalization in the lowest-income group was significantly lower than that of other income groups; that is, many of them might have chosen not to get hospitalized. The insured groups of different income levels have significant differences in the utilization of medical services compared with the highest-income group. Since the insurance amount of medical insurance is reimbursed according to a certain proportion of the total medical expenses, the difference in the population that avails medical service utilization is likely to cause the difference in the medical compensation level. As seen from the last column in Table 4, the possibilities of applying medical insurance for the lower-income, middle-income, and high-income groups are lower by about 14% compared to the highest-income group, and the lowest-income group is lower by about 7.67% compared to the highest-income group.

**3.4. The Role of Commercial Medical Insurance.** Among the migrants with basic medical insurance, many also participated in commercial medical insurance. Commercial medical insurance can effectively alleviate the medical consumption burden of the floating population as a supplement to basic medical insurance. The next step is to

examine the supplementary equity of the multilevel medical security system by adding commercial medical insurance factors to the Heckman two-part model (participation in the commercial medical insurance is recorded as 1, and non-participation is recorded as 0). The second column in Table 5 shows that the reimbursement ratios of other income groups were significantly different from those of the highest-income group. Meanwhile, with the increase in income, the reimbursement ratio gradually decreases. However, the reimbursement ratio for the lowest-income group was not the lowest, and the lowest-income group spends the least on hospitalization, as displayed in the last column in Table 5. Comparing Tables 3 and 5, it is shown that, after adding the commercial insurance restrictions, the coefficient of medical reimbursement ratio of other groups became smaller than that of the highest-income group, indicating that commercial medical insurance had reduced the difference in medical reimbursement ratios between different groups to a certain extent. The last column in Table 5 presented that the total hospitalization medical expenses of all other groups are significantly different compared to the highest-income group. Compared with the full sample in Table 4, under the restrictions of commercial insurance, group differences were reduced; however, the lowest-income group still had the least expenditure on hospitalization.

## 4. Discussion

The basic medical insurance system is a beneficial project that covers all citizens and requires fairness in system design. However, there is some unfairness endemic to developing basic medical insurance systems in countries worldwide, forcing the continuous optimization and the reform of the basic medical insurance system to achieve the basic medical insurance system's original intention. The previous literature has defined the fairness of medical services [41, 42], but few studies discussed the definition of the fairness of medical insurance. The fairness of medical insurance is different from the fairness of medical services. The fairness of medical insurance emphasizes the accessibility of services (e.g., the possibility of hospitalization) and underlines cost-sharing accessibility (e.g., the medical insurance applicability).

This study reveals the fairness of the benefits of basic medical insurance for the floating population with different income levels, mainly from the perspective of equalization. The results showed that different income groups have significant differences in the benefits of basic medical insurance among the floating population, and the highest-income group benefited the most. Although the low-income and the high-income groups paid the same medical insurance premiums, they did not receive the same medical compensation. The highest-income group obtained more compensation than their premium contributions, indicating unequal benefits of the basic medical insurance system among migrants and the existence of the phenomenon that the "poor" subsidized the "rich." However, we found that the higher-income migrants benefited the least in medical insurance compensation, as listed in Table 3, which is different from the earlier studies [26–28, 43]. The previous results

TABLE 4: Influencing channels.

Variable	Total hospitalization medical expenses		Possibility of medical insurance application
	Heckman model		Probit model
	First step (1)	Second step (2)	
Income level (highest-income group as a reference)			
Lowest income	-0.0735** (0.0285)	-0.3207*** (0.0682)	-0.0767** (0.0286)
Lower income	-0.1373*** (0.0294)	-0.2523*** (0.0725)	-0.1399*** (0.0294)
Middle income	-0.1393*** (0.0261)	-0.2574*** (0.0628)	-0.1398*** (0.0261)
Higher income	-0.1363*** (0.0262)	-0.2591*** (0.0630)	-0.1398*** (0.0263)
Wald test (rho = 0) chi <sup>2</sup> (1)	6.21***		
Observations	151999	2198	154197

Standard errors are in parentheses; \*  $p < 0.05$ , \*\*  $p < 0.01$ , and \*\*\*  $p < 0.001$ .

TABLE 5: Influence of commercial medical insurance.

Variable	Medical reimbursement		Total hospitalization medical expenses	
	Heckman model		Heckman model	
	First step (1)	Second step (2)	First step (1)	Second step (2)
Lowest income	-0.0502 (0.0288)	-0.0923** (0.0321)	-0.0576* (0.0288)	-0.3062*** (0.0687)
Lower income	-0.1257*** (0.0298)	-0.1270*** (0.0333)	-0.1258*** (0.0297)	-0.2243** (0.0729)
Middle income	-0.1329*** (0.0265)	-0.1302*** (0.0295)	-0.1344*** (0.0265)	-0.2538*** (0.0637)
Higher income	-0.1236*** (0.0265)	-0.1501*** (0.0295)	-0.1275*** (0.0265)	-0.2544*** (0.0635)
Wald test (rho = 0) chi <sup>2</sup> (1)	2135.82***		7.81**	
Observations	149022	2165	149022	2165

Standard errors are in parentheses; \*  $p < 0.05$ , \*\*  $p < 0.01$ , and \*\*\*  $p < 0.001$ .

concluded that the lowest-income people benefited the least in medical subsidies. The possible reason was that the hospitalization expenses of the lowest-income group were 32.07% less than those of the highest-income group. In comparison, hospitalization expenses of the higher-income group are only 25.74% less than those of the highest-income group, as shown in Table 4. Nevertheless, the medical insurance applicability of the lowest-income group is 7.67% lower than that of the highest-income group, and that of the higher-income group is 13.98% lower than that of the highest-income group. These indicated that, compared with the higher-income group, the lowest-income group is more likely to spend less on hospitalization expenses but is more willing to be reimbursed by medical insurance. High-income people had more hospitalization expenses. However, they were often unwilling to choose reimbursement of medical insurance. The reason might be that the current medical insurance reimbursement for the floating population has not been fully integrated with the inflow place. Medical treatment settlement in different places is still a problem [44–46]. The cumbersome medical reimbursement procedures and the high cost make the higher-income group abandon the reimbursement and choose to pay. However, the lowest-income people were more willing to use medical reimbursement to relieve their financial burden even if the procedures were complicated.

This paper reports differences in the basic medical insurance reimbursement ratio and hospitalization medical expense between other income groups and the highest-income group after adding commercial medical insurance. On

doing so, the degree of difference decreased, indicating that commercial medical insurance reduced the beneficial inequality among income groups to some extent. Therefore, while optimizing the basic medical insurance system in China, the supplementary commercial medical insurance can be appropriately adopted to improve the fairness of basic medical insurance benefits according to the differences among the income groups.

Though the Chinese health insurance reform has gained great success, more work should be done to ensure the equal benefits of health insurance. Future research should focus on how the Chinese government can decrease the premium of vulnerable groups such as the lower-income migrants or increase the reimbursement rates among those at high risk of catastrophic health expenditures. Additionally, more research on designing a medical insurance system according to the characteristics of the floating population to improve the fairness of the medical insurance system would be extremely helpful in deepening reforms of the Chinese medical insurance system.

This study has several limitations. First, the survey data were not specially designed for this study, and the data were collected before this study. Thus, the data were imperfect for accessing hospital medical insurance reimbursement amount because the reimbursement amount of basic medical insurance in the questionnaire included NRCMS, URBMI, UEBMI, employment injury insurance, and maternity insurance, which might be either underestimated or overestimated due to bias. Using various reimbursement data of different basic medical insurance to solve the fairness

of benefits of different basic medical insurance systems for the floating population is a possible direction for further research. Second, the hospitalization medical expense among the floating population was not caused by the latest inpatient service, which did not fully represent the total hospitalization medical expense. These factors might also have led to bias. Additionally, some confounding factors such as health and disease status could not be controlled because of the limitations of available data, which might be related to the medical insurance availability and the utilization of medical services for the floating population. Third, since the respondents' age was restricted to 15–59 years, this paper's findings cannot be generalized to all age groups, especially for the elderly migrant population. There might also be inequality of benefits from medical insurance among the elderly population. Machine learning methods have been applied in many fields [47, 48], but studies in the fields of medical care and healthcare are rare. This study used the random forest algorithm to solve the medical and health insurance problem, which can be regarded as its preliminary attempt. Further in-depth research is still needed.

## 5. Conclusions

This paper answers the question of the fairness of the benefits of the basic medical insurance for the floating population using the 2014 CMDS survey data and attempts to discover whether there is a phenomenon where the poor subsidize the rich. The study analyzed the fairness of the basic medical insurance benefits for the insured floating population with different income levels from the perspective of equalization by using the two-stage Heckman model. The main findings were that, regardless of the fairness of basic medical insurance compensation of hospitalization expenses, there were significant differences among the floating population with different income levels, and the highest-income group benefited the most. The phenomenon that the lower-income group subsidized the higher-income groups was also confirmed. Additionally, those who benefited the least were the higher-income group rather than the lowest-income group. Further inspection illustrated that commercial medical insurance reduced the inequality of benefits among groups to a certain extent. These findings may provide the inspiration that the Chinese government accelerates the reform of medical service fairness.

## Data Availability

The data in this paper are obtained from the China Migrants Dynamic Survey in 2014.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Acknowledgments

The authors are grateful to the National Health Commission for the data support. This study was supported by the Great

Wall Scholar Training Program for the Construction of High-Level University Teachers in Beijing.

## References

- [1] L. Qin, C.-P. Chen, X. Liu, C. Wang, and Z. Jiang, "Health status and earnings of migrant workers from rural China," *China & World Economy*, vol. 23, no. 2, pp. 84–99, 2015.
- [2] C. X. Yong and J. X. Ying, "A study on medical aid for migrant workers," *Journal of Zhejiang University (Humanities and Social Sciences)*, 2006.
- [3] M. Roland, P. Ana, D. Codruta, D. Ruxandra, and P. Ana-Madalina, "Analysis of the drivers of occupational health and safety performance disclosures by Romanian companies," *Economic Computation and Economic Cybernetics Studies and Research*, vol. 54, no. 3/2020, pp. 197–214, 2020.
- [4] V. Maria Denisa, M. Erika, B. Liviu-Stelian, and E. Andreea-Oana, "Perceptions about immigrants and their integration in the European Union," *Economic Computation and Economic Cybernetics Studies and Research*, vol. 54, no. 4/2020, pp. 103–118, 2020.
- [5] M. E. Brown, A. B. Bindman, and N. Lurie, "Monitoring the consequences of uninsurance: a review of methodologies," *Medical Care Research and Review*, vol. 55, no. 2, pp. 177–210, 1998.
- [6] T. C. Buchmueller, K. Grumbach, R. Kronick, and J. G. Kahn, "The effect of health insurance on medical care utilization and implications for insurance expansion: a review of the literature," *Medical Care Research and Review*, vol. 62, no. 1, pp. 3–30, 2005.
- [7] C. Hoffman and J. Paradise, "Health insurance and access to health care in the United States," *Annals of the New York Academy of Sciences*, vol. 1136, no. 1, pp. 149–160, 2008.
- [8] W. Manning, G. A. Leibowitz, G. A. Goldberg, J. P. Newhouse, and W. H. Rogers, "A controlled trial of the effect of a prepaid group practice on the utilization of medical services," *New England Journal of Medicine*, vol. 310, no. 23, p. 1505, 1985.
- [9] C. E. Ross and J. Mirowsky, "Does medical insurance contribute to socioeconomic differentials in health?" *The Milbank Quarterly*, vol. 78, no. 2, pp. 291–321, 2000.
- [10] B. G. Saver and N. Peterfreund, "Insurance, income, and access to ambulatory care in King County, Washington," *American Journal of Public Health*, vol. 83, no. 11, pp. 1583–1588, 1993.
- [11] J. S. Weissman, R. Stern, S. L. Fielding, and A. M. Epstein, "Delayed access to health care: risk factors, reasons, and consequences," *Annals of Internal Medicine*, vol. 114, no. 4, pp. 325–331, 1991.
- [12] W. Chen, Q. Zhang, A. M. N. Renzaho, F. Zhou, H. Zhang, and L. Ling, "Social health insurance coverage and financial protection among rural-to-urban internal migrants in China: evidence from a nationally representative cross-sectional study," *BMJ Global Health*, vol. 2, no. 4, 2017.
- [13] F. Zhang, X. Shi, and Y. Zhou, "The impact of health insurance on healthcare utilization by migrant workers in China," *International Journal of Environmental Research and Public Health*, vol. 17, no. 6, 2020.
- [14] J. Z. Ayanian, B. A. Kohler, T. Abe, and A. M. Epstein, "The relation between health insurance coverage and clinical outcomes among women with breast cancer," *New England Journal of Medicine*, vol. 329, no. 5, pp. 326–331, 1993.
- [15] J. Hadley and T. Waidmann, "Health insurance and health at age 65: implications for medical care spending on new

- medicare beneficiaries,” *Health Services Research*, vol. 41, no. 2, 2010.
- [16] S. T. J. Lair, “Health insurance and health status: implications for financing health care reform,” *Inquiry*, vol. 31, no. 4, pp. 425–437, 1994.
- [17] W. H. Organization, *World Health Report 2010: Health Systems Financing—The Path to Universal Coverage*, Vol. 17, WHO, Geneva, Switzerland, 2010.
- [18] L. T. U. Politics Program, *Regional Office for the Western Pacific—Inter-Governmental Organisation*, WHO, Geneva, Switzerland, 1948.
- [19] X. Qin, J. Pan, and G. G. Liu, “Does participating in health insurance benefit the migrant workers in China? An empirical investigation,” *China Economic Review*, vol. 30, pp. 263–278, 2014.
- [20] H. Yu, “Universal health insurance coverage for 1.3 billion people: what accounts for China’s success?” *Health Policy*, vol. 119, no. 9, pp. 1145–1152, 2015.
- [21] C. Zhao, C. Wang, C. Shen, and Q. Wang, “China’s achievements and challenges in improving health insurance coverage,” *Drug Discoveries & Therapeutics*, vol. 12, no. 1, pp. 1–6, 2018.
- [22] W. Fu, S. Zhao, Y. Zhang, P. Chai, and J. Goss, “Research in health policy making in China: out-of-pocket payments in healthy China 2030,” *BMJ*, vol. 360, 2018.
- [23] W. M. Lindelow, “Can insurance increase financial risk?: The curious case of health insurance in China,” *Journal of Health Economics*, vol. 27, no. 4, pp. 990–1005, 2008.
- [24] X. Lei and W. Lin, “The new cooperative medical scheme in rural China: does more coverage mean more service and better health?” *Health Economics*, vol. 18, no. S2, pp. S25–S46, 2009.
- [25] Y. Wei and W. Xun, “Paying for outpatient care in rural China: cost escalation under China’s new co-operative medical scheme,” *Health Policy & Planning*, vol. 30, no. 2, 2014.
- [26] F. Castro-Leal, J. Dayton, L. Demery, and K. Mehra, “Public spending on health care in Africa: do the poor benefit?” *Bulletin of the World Health Organization*, vol. 78, no. 1, p. 66, 2000.
- [27] P. A. Mills, “The poor pay more: health-related inequality in Thailand,” *Social Science & Medicine*, vol. 44, no. 12, pp. 1781–1790, 1997.
- [28] T. B. Chu, T. C. Liu, C. S. Chen, Y. W. Tsai, and W. T. Chiu, “Household out-of-pocket medical expenditures and national health insurance in Taiwan: income and regional inequality,” *Bmc Health Services Research*, vol. 5, no. 1, p. 60, 2005.
- [29] M. N. Baily, “Some aspects of optimal unemployment insurance,” *Journal of Public Economics*, vol. 10, no. 3, pp. 379–402, 2006.
- [30] R. Chetty, “Sufficient statistics for welfare analysis: a bridge between structural and reduced-form methods,” *Annual Review of Economics*, vol. 1, no. 1, pp. 451–488, 2009.
- [31] P. L. Annear, M. Bigdeli, and B. Jacobs, “A functional model for monitoring equity and effectiveness in purchasing health insurance premiums for the poor: evidence from Cambodia and the Lao PDR,” *Health Policy (Amsterdam, Netherlands)*, vol. 102, no. 2–3, pp. 295–303, 2011.
- [32] Y. Li, Q. Wu, L. Xu et al., “Factors affecting catastrophic health expenditure and impoverishment from medical expenses in China: policy implications of universal health insurance,” *Bulletin of the World Health Organization*, vol. 90, no. 9, pp. 664–671, 2012.
- [33] C. Kronenberg and P. P. Barros, “Catastrophic healthcare expenditure – drivers and protection: the Portuguese case,” *Health Policy*, vol. 115, no. 1, pp. 44–51, 2013.
- [34] J. Pan, X. Lei, and G. G. Liu, “Health insurance and health status: exploring the causal effect from a policy intervention,” *Health Economics*, vol. 25, no. 11, pp. 1389–1402, 2015.
- [35] M. Buschiazzo, J. Mula, and F. Campuzano-Bolarin, “Simulation optimization for the inventory management of healthcare supplies,” *International Journal of Simulation Modelling*, vol. 19, no. 2, pp. 255–266, 2020.
- [36] H. N. Gao, D. H. Shen, and W. C. Zhang, “Identification of cutting chatter through deep learning and classification,” *International Journal of Simulation Modelling*, vol. 19, no. 4, pp. 667–677, 2020.
- [37] L. Breiman, “Random forest,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [38] Q. Zhou, G. G. Liu, and S. Krumholz, “Is Chinese national health insurance effective in the face of severe illness? a perspective from health service utilization and economic burden,” *Social Indicators Research*, vol. 132, no. 3, pp. 1–23, 2016.
- [39] S. Xie, J. Wang, J. Chen, and V. M. Ritakallio, “The effect of health on urban-settlement intention of rural-urban migrants in China,” *Health & Place*, vol. 47, pp. 1–11, 2017.
- [40] T. Hastie, “The elements of statistical learning: data mining,” *Journal of the Royal Statistical Society*, vol. 167, no. 1, p. 192, 2004.
- [41] M. Whitehead, “The concepts and principles of equity and health,” *International Journal of Health Services*, vol. 22, no. 3, pp. 429–445, 1992.
- [42] G. H. Mooney, “Equity in health care: confronting the confusion,” *Effective Health Care*, vol. 1, no. 4, pp. 179–185, 1983.
- [43] D. R. Gwatkin, A. Bhuiya, and C. G. Victora, “Making health systems more equitable,” *The Lancet*, vol. 364, no. 9441, pp. 1273–1280, 2004.
- [44] J. Liu and W. J. Jacob, “From access to quality: migrant children’s education in urban China,” *Educational Research for Policy and Practice*, vol. 12, no. 3, pp. 177–191, 2013.
- [45] M. B. Xiang, “Native place, migration and the emergence of peasant enclaves in Beijing,” *China Quarterly*, vol. 155, pp. 546–581, 1998.
- [46] Z. Yinjun, K. Bowei, L. Yawen et al., “Health insurance coverage and its impact on medical cost: observations from the floating population in China,” *PLoS One*, vol. 9, no. 11, Article ID e111555, 2014.
- [47] A. Lorenc, M. Kužnar, T. Lerher, and M. Szkoda, “Predicting the probability of cargo theft for individual cases in railway transport,” *Tehnicki Vjesnik-Technical Gazette*, vol. 27, no. 3, pp. 773–780, 2020.
- [48] D. Marić, M. Duspara, T. Šolić, and I. Samardžić, “Application of SVM models for classification of welded joints,” *Tehnicki Vjesnik-Technical Gazette*, vol. 26, no. 2, pp. 533–538, 2019.