education sciences

MDPI

Article

Research on the Contribution Mechanism of Vocational Human Capital Characteristics to Income

Xiaowei Hao [1], Xi Yang [2], Kunzhao Kou [3,*], Yu Zhang [3,*] and Congbin Guo [4]

[1] Stanford Center on China's Economy & Institutions, Stanford University, Stanford, CA 94305, USA
[2] School of Economics, Peking University, Beijing 100871, China
[3] Graduate School of Education, Beijing Foreign Studies University, Beijing 100089, China
[4] Graduate School of Education, Peking University, Beijing 100871, China
* Correspondence: koukunzhao@bfsu.edu.cn (K.K.); 8812zhangyu@bfsu.edu.cn (Y.Z.)

Abstract: Vocational education is an important way to accumulate human capital. Human capital is the core element of economic growth and has huge positive externalities. Building a scientific and effective human capital development system is an important driving force to improve workers' living standards and promote innovative development. Based on statistical techniques such as Lasso dimensionality reduction, stepwise regression, and partial least squares, as well as on the 2012–2016 China Labor Force Dynamics Survey (CLDS), this paper studies the impact of human capital on workers' wage income and capital income, and establishes an income-determining equation that can be used for interpretation and forecasting. The empirical results show that education, professional skills, health, and communication ability are important components of human capital and are significantly positively correlated with income. China should build a good and effective human capital development system to increase workers' income and narrow the income gap among residents.

Keywords: human capital; vocational education; career development system; income

check for updates

Citation: Hao, X.; Yang, X.; Kou, K.; Zhang, Y.; Guo, C. Research on the Contribution Mechanism of Vocational Human Capital Characteristics to Income. *Educ. Sci.* **2023**, *13*, 246. https://doi.org/10.3390/educsci13030246

Academic Editors: Xinqiao Liu, Haisheng Pan, Geng Wang and Wenjuan Gao

Received: 27 December 2022
Revised: 5 February 2023
Accepted: 16 February 2023
Published: 25 February 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Human capital plays a crucial role in driving innovation and economic growth. An effective evaluation system for human capital is crucial for fostering innovation in society. The importance of human capital training lies in improving individual income, reducing income inequality, boosting industrial innovation, and enhancing national competitiveness [1–3].

Vocational education has become a subject of increasing concern globally, due to the need to enhance the employability of vocational education graduates. The world is facing ecological, social, and economic changes, and a highly skilled workforce is crucial for sustainable development. In the knowledge economy, new types of human capital may be needed in low- and middle-income countries, and vocational education can provide necessary professional qualifications that are not available through general secondary education. Investing in vocational education is a means of boosting economic competitiveness and alleviating poverty [4,5].

This article aims to address the critical issue of human capital evaluation and the research gap regarding vocational education. The research framework consists of relevant literature review; data and sample description; statistical analysis using Lasso dimensionality reduction, stepwise regression, and partial least squares; regression analysis; and policy recommendations. The article is organized as follows: Section 2 reviews the relevant literature, Section 3 describes methodology and data, Section 4 analyze the results of regression analysis, and Section 5 provides conclusions and policy recommendations for talent development.

2. Literature Review

From the perspective of individuals, unique competitiveness is a key factor in developing a specialized talent. Research by Zhu (2013) has pointed out that the development of talents is related to high-quality, outstanding creativity, and outstanding workability [6]. Specialized talents can be defined according to four dimensions: behavioral ability, knowledge reserve, skill mastery, and competence. Therefore, the essence of specialized talents are also their human capital accumulation. Talents are a core component in the competitiveness of individual income sources and an important driving force for economic growth. The 2013 Human Development Report and previous reports have pointed out that economic growth itself cannot automatically translate into human development and progress [7]. Only through poverty alleviation policies involving education, nutrition, health, and work skills and a large amount of investment aimed at improving people's abilities can a government expand the number of opportunities to obtain decent work and ensure the continuous progress of humans. Both lecturers and students need to prepare for the future workforce. Various studies have highlighted the importance of digital competencies, such as digital literacy and innovative tools; however, the shift of innovative technology applications in industry to existing higher education programs often occurs rather slowly. Industry and higher education institutions need to collaborate and communicate extensively to prepare students for future jobs, and lecturers play a crucial role in this transition [8]. Promoting the accumulation of human social capital and improving the talent development system through various methods are also important goals toward income increases and economic growth.

From the macro perspective, technical talents can promote innovation and a country's development. From the individual perspective, they can rely on their human capital to obtain income from the market. This article intends to explore the core factors that contribute to income from the level of individual income to develop corresponding suggestions for talent development.

Regarding the impact of human capital characteristics on workers' income, Chinese and foreign scholars have conducted a series of studies. The most representative ones are from Schultz [9], Mincer [10] and Becker [11]. Following their research, they believe there is a significant positive correlation between people's level of education and income. In terms of education investment, the higher the level of education, the smaller the income gap between people, and the lower the level of education, the larger the income gap. Kartini and Weil recognize the importance of nutrition and health to human capital and believe that health human capital increases the income level of farmers and can prevent farmers from falling into a poverty trap [12,13]. Gao and Yang (2006) believe that vocational education and on-the-job training will contribute more to farmers' income than basic education [14]. Different characteristics of human capital produce corresponding contributions to the incomes of different groups, so measuring the contribution rate of different characteristics of human capital to income can provide corresponding guidance for the talent development system.

The application of machine learning methods in research on human capital and its contribution to income is relatively new. Most existing studies rely heavily on databases and survey data, and prior information is used to guide the research and model development. This approach, however, becomes less effective in the era of big data, when it becomes increasingly difficult to obtain reliable prior information. In such a scenario, data mining methods can provide a valuable solution to help researchers understand the problem and advance scientific knowledge [15,16]. In this context, we plan to introduce machine learning methods to research on the effect of human capital on income. Our research will focus on the microdata of the Chinese labor force dynamics survey, which provides a rich source of information on human capital and income. In order to extract useful insights from this data, we will use Lasso for variable selection and dimensionality reduction, as well as partial least squares and stepwise regression to construct an income-determining equation. This equation can then be used for interpretation and forecasting purposes. One of the

advantages of using machine learning methods is that they can help researchers identify important variables that are not included in the prior information. For example, Lasso can help reduce the number of variables by identifying the most important ones, while partial least squares and stepwise regression can help integrate microdata and construct an equation that is able to explain the relationship between human capital and income. This can provide valuable insights into the effect of human capital on income, and help researchers understand the key drivers of income growth in the Chinese labor force. Overall, the integration of machine learning methods into research on human capital and its effect on income provides a new and innovative approach that can help researchers overcome the challenges of big data and advance scientific knowledge. By using the microdata of the Chinese labor force dynamics survey, we hope to make significant contributions to this field of research and provide valuable insights for policymakers and researchers.

Based on the spatial and temporal differences in China's economic development, this article uses stepwise regression, Lasso, and partial least squares to explore the differences in the effects of human capital characteristics on income and then explore the structural characteristics of the talent development system that the government needs to build. Our conclusions, based on the empirical research, show that the determinants of wage income are an efficient and practical skill education system, the worker's barrier-free communication skills, higher education level, and their own industry choices. High-tech industries have a higher return on income compared to traditional industries. Moreover, the determinants of capital income are more affected by education level, health status, and regional differences [17].

Therefore, the goals that should be achieved are to improve the current talent development system, strengthen public education investment, improve vocational skills training, develop high-tech industries, and promote population mobility and integration. The development of vocational education and high-tech industries is important for increasing wage income. Both have a greater impetus and help improve the human capital of the whole of society [18]. Some developing countries are currently pushing for vocational education and training as a way of building human capital and strengthening economic growth [19,20]. However, in reality, many of these countries' education systems, particularly those for vocational education, do not know which abilities to focus on nurturing. This could result in negative impacts on building human capital if expanding vocational education replaces academic education [21,22].

Compared with previous studies, this article is based on micro-survey data, with the help of statistical methods such as Lasso dimensionality reduction, stepwise regression, and partial least squares to construct an interpretable and predictable income-determining equation. Moreover, the previous literature seldom analyzes the impact of human capital characteristics on capital income, but, according to Piketty, capital income is an important component of income [23,24]. In developing countries, the rate of return on capital income is much higher than labor income, so this article also analyzes wage income and capital income, in an effort to find the difference between the determinants of the two. We hope to answer the following questions: What are the most significant human capital characteristics that have a significant impact on income? To what extent do education level, work experience, and professional skills contribute to income? What are the key differences between wage income and capital income, and what are the major factors that influence each?

3. Methodology

Too many independent variables will bring a dimension disaster; when the number of independent variables is greater than the sample size, the coefficient β cannot be estimated by the least square method. In addition, linear models also risk overfitting; at this time, if you use brand new data to verify the model, the effect is usually weak. Therefore, this article first uses the Lasso algorithm for dimensionality reduction [25].

3.1. The Principle of Lasso Regression

In linear regression, the common expressions are as follows:

$$\begin{gathered}\text{Assumption function}: \ h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n \\ \text{Loss function}: \ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)^2 \\ \text{Objective function}: \ \min J(\theta_0, \theta_1, \ldots, \theta_n)\end{gathered}$$

Y represents the dependent variable, X represents the independent variables, and θ represents the coefficient estimates for different variables in the above linear regression equation. Here, the θ_i represents the parameters (also called weights) parameterizing the space of linear functions mapping from X to Y. When there is no risk of confusion, we will drop the θ subscript in $h_\theta(x)$.

The lasso algorithm solves the problem of overfitting by introducing a regularization term (L1 norm, or ridge regression) in the loss function. The updated regression loss function is as follows:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)^2 + \lambda \sum_{j=1}^{n} |\theta_j|$$

Lasso regression will cause the estimation of some regression coefficients to be 0; thus, by removing variables with regression coefficients of 0, lasso regression plays a role in variable selection. And λ is called the regularization parameter.

3.2. The Principle of Stepwise Regression

The basic idea behind stepwise regression is to introduce variables into the model one by one. After each independent variable is introduced, an F test is performed, and the selected independent variables are tested one by one. The originally introduced independent variable changes due to the introduction of the subsequent independent variable. When it is no longer significant, it is deleted. One must ensure that only significant variables are included in the regression equation before introducing each new variable. This article adopts the backward stepwise regression method.

3.3. The Principle of Partial Least Squares

The partial least squares method is a mathematical optimization technique that finds the best function match for a data set by minimizing the sum of squares of errors. This article explicitly uses the following PLSR algorithm to achieve this.

The PLSR (Partial Least Squares Regression) algorithm is a regression modeling method concerning multi-dependent variable Y to multi-independent variable X. In the process of establishing regression, the algorithm considers extracting the principal components of Y and X as much as possible. It also considers maximizing the correlation between the principal components extracted from X and Y.

3.4. Data and Sample Description

This study's dependent variable was residents' income, including wage and capital income. Wage income refers to all the labor remuneration obtained by employees through various channels, including the wages of their main occupations and other labor incomes obtained by engaging in secondary occupations, other part-time jobs, and sporadic labor. Capital income refers to after-tax operating income, including agricultural operating income (self-sufficient agricultural production needs to be converted into income according to market value) and business operating income, such as from shops and factories.

Among them, compared with wage income, the rate of return on capital income is higher; at the same time, for individual residents, capital income is more related to their actual standard of living, so it makes a greater contribution to income. We select different

independent variables for these two types of income to construct income-determining equations according to the previous literature.

Among them, the independent variables of wage income were divided into two categories. The residents' characteristics included age (18 to 65 years old), professional level, education level, whether they understand foreign languages, health status, etc. The other category was the resident's work background, including whether he has work experience, whether he is engaged in agricultural production activities, the type of employer, the type of work industry, social class, whether he has registered permanent residence, and his local adaptation situation. The independent variables of capital income were excluded from the independent variables of wage income, and relevant information, such as the industry and the type of employer, was removed. Table 1 reports the categories and specific definitions of variables:

Table 1. Variable classification and definition.

Category		Meaning	Variables	Total Number
Independent variables	Quantitative variables	Knowledge of foreign languages	X1	103
		Whether they have a professional qualification certificate	X2	
		Number of professional qualification certificates	X3	
		Number of local friends	X4	
		Local dialect level	X5	
		Social class	X6	
		Whether they are engaged in agricultural production	X7	
		Whether they have relocated the household registration	X8	
	Qualitative variables	Location	X9-X37	
		Education level	X38-X46	
		Political outlook	X47-X49	
		Health level	X50-X54	
		Father's education level	X55-X64	
		Mother's education level	X65-X72	
		Industry	X73-X89	
		Employer type	X90-X103	
Dependent variable	Quantitative variable	Wage income	Y1	2
		Capital income	Y2	

Table 2 reports the descriptive statistics of the variables:

Table 2. Statistical description of variables.

Variable Category	Meaning	Average Value	Maximum Value	Minimum Value
Quantitative variables	Knowledge of foreign languages	0.152	1	0
	Whether they have a professional qualification certificate	0.216	1	0
	Number of professional qualification certificates	0.789	18	0
	Number of local friends	0.075	1	0
	Local dialect level	0.066	1	0
	Social class	3.052	9	0
	Whether they are engaged in agricultural production	0.629	1	0
	Whether to relocate the household registration	0.370	1	0
	Wage income	10.291	12.900	3.689
	Capital income	10.365	13.304	3.689
Qualitative variables	Location	Shanghai, Yunnan, Region: Inner Mongolia, Beijing, Jilin, Sichuan, Tianjin, Ningxia, Anhui, Shandong, Shanxi, Guangdong, Guangxi, Xinjiang, Jiangsu, Jiangxi, Hebei, Henan, Zhejiang, Hubei, Hunan, Gansu, Fujian, Guizhou, Liaoning, Chongqing, Shaanxi, Qinghai, Heilongjiang		
	Education level	no school, elementary school, junior high school, technical secondary school, high school, junior college, undergraduate, master's, doctorate		
	Political outlook	the masses, democratic parties, members of the Communist Party of China		
	Health level	(not filled in), average, healthy, relatively unhealthy, very unhealthy, very healthy		
	Father's education level	(not filled in), never attended school, elementary school, junior high school, technical secondary school, high school, college, undergraduate, master's, doctorate, other		
	Mother's education level	(not filled in), never attended school, elementary school, junior high school, technical secondary school, high school, college, undergraduate, master's, doctorate, other		
	Industry	(not filled in), unclear, not applicable, transportation, storage, post and telecommunications, other industries, manufacturing, farming: agriculture, forestry, animal husbandry, sideline and fishery production (such as farming, breeding chickens, ducks, aquatic products, etc.), hygiene, sports and social welfare industry, state agencies, party and government agencies and social organizations, geological survey industry, water conservancy management industry, construction industry, real estate industry, wholesale and retail trade, catering industry, refusal to answer, education, culture and art, radio, film and television industry, electricity, gas and water production and supply industry, social service industry, scientific research and comprehensive technical service industry, extractive industry, finance and insurance industry		
	Employer type	(not filled in), unclear, not applicable, individual industrial and commercial, public institution, party and government organs, people's organization, army, others, farming: agriculture, forestry, animal husbandry, sideline fishery production (such as farming, breeding chickens, ducks, aquatic products, etc.), state-owned/collective institutions, state-owned enterprises, foreign investment, joint ventures, refusal to answer, autonomous organizations such as village neighborhood committees, private non-enterprises, social organizations and other social organizations, private and private enterprises, freelance workers (freelancers, casual workers, vendors, etc.), nanny without dispatch unit (self-operated driver, hand craftsman, etc.), collective enterprise		

Note: Except for the number of professional qualification certificates, social class, wage income, and capital income, other variables are all dummy variables.

The sample used in this study was obtained from the panel data of the China Labor force Dynamics Survey (CLDS) of Sun Yat-sen University. The China Labor Force Dynamics Survey (CLDS) established a comprehensive database with the labor force as the object of the investigation through a biennial follow-up survey of urban and rural villages in China, including labor force individuals and families. The three-level tracking and cross-sectional data of the community can provide high-quality basic data for empirically oriented theoretical research and policy research. The CLDS completed the 2012 national baseline and

follow-up surveys in 2014 and 2016. The CLDS targets the working-age population, aged 15–64, focusing on the status of and changes in labor force education, employment, labor rights, occupational mobility, occupational protection and health, occupational satisfaction, and happiness. At the same time, it investigates the political, economic, and social development of the communities where the labor force is located and the demographic structure of labor force households, family property and income, household consumption, family donations, rural household production, and land. In 2012, the CLDS completed the first national official survey involving 303 villages, 10,612 households, and 16,253 labor force individuals. In 2014, the CLDS completed the first follow-up survey, adding 101 community samples to the surveys of 2012, and completed 14,226 household and 23,594 individual questionnaires. In 2016, the third national CLDS survey involved replacing a quarter of the surveys from 2014 with community samples, and a total of 401 community questionnaires, 14,226 household questionnaires, and 21,086 individual questionnaires were completed. The success rate of all households registered in this survey round was 71.37%, and the completion rate of individual questionnaires within a family was 76.86%. In the follow-up survey samples, the success rate of family tracking was 71.99%, and the success rate of individual tracking was 58.40% [26].

The steps for selecting samples were as follows: First, to ensure the continuity of the research subjects, we selected individuals who had participated in the survey for all three years; secondly, since the old questionnaire ID was updated to the new ID in 2014, this article used the data from 2014 to conduct the individual matching. In the specific extraction process, the indicator name changed, and this article made corresponding adjustments and finally obtained a balanced sample of 4,091 individuals and 12,273 observations. Next, this study took the natural logarithms of wage income and capital income. Using the quantitative independent variables, in order to eliminate the influence of data dimensions and speed up the regression speed, a linear normalization operation was adopted:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

We split qualitative variables into multiple dummy variables to facilitate subsequent regression. In addition, we deleted all observations with zero income, missing values, and incomes from the highest 1% of the year to avoid interference from extreme values effecting the forecast results.

4. Results

Next, we established a regression model on the reduced dimensionality data to study the determinants of residents' wage income and capital income. In order to ensure the accuracy and robustness of the regression results, this study divided the CLDS data into two parts: a training set and a test set. Data from 2012 and 2014 were used as the training set, and data from 2016 formed the test set. In this section, we used stepwise regression and partial least squares to analyze the training set.

4.1. Dimensionality Reduction Result

This study used the Lasso algorithm with the R to perform a preliminary dimensionality reduction on the data. Moreover, it uses LARS (Least Angle Regression) minimum angle regression to find the step that satisfies the optimal regression coefficient on the solution path, which minimizes the mean square error MSE count to filter the variables. The number of independent variables after dimensionality reduction is shown in the following table (Table 3):

Table 3. Dimensionality reduction results.

Y	Lasso Dimensionality Reduction	
	Number of Original Independent Variables	Number of Remaining Independent Variables
Wage income	103	49
Capital income	72	33

After dimensionality reduction, the original 103 independent variables of wage income were effectively reduced to 49, and the original 72 independent variables of capital income were reduced to 33. The specific dimensionality reduction results are shown in the following table (Table 4):

Table 4. Specific description of dimensionality reduction results.

Dependent Variables	Extract Variables after Dimensionality Reduction
Wage income	X1 (whether they understand a foreign language), X2 (whether they have a professional qualification certificate), X4 (number of local friends), X5 (level of local dialect), X7 (whether engaged in agricultural production), X9 (whether they live in Shanghai), X12 (whether they live in Beijing), X13 (whether they live in Jilin), X15 (whether they live in Tianjin), X17 (whether they live in Anhui), X18 (whether they live in Shandong), X22 (whether they live in Xinjiang), X23 (whether they live in Jiangsu), X24 (whether they live in Jiangxi), X25 (whether they live in Hebei), X26 (whether they live in Henan), X27 (whether they live in Zhejiang), X29 (whether they live in Hunan), X31 (whether they live in Fujian), X34 (whether they live in Chongqing), X38 (whether education level is technical secondary school?), X40 (whether education level is a doctorate?), X41 (whether education level is junior college), X42 (whether education level is primary school), X43 (whether education level is never attended school), X44 (whether education level is undergraduate), X50 (whether health level is average), X52 (whether health level is relatively unhealthy), X54 (whether health level is very healthy), X56 (whether father's education level is other), X57 (whether father's education level is junior high school), X59 (whether father's education level is a junior college), X61 (whether father's education level is never attended school), X62 (whether father's education level is a bachelor's degree), X76 (whether the industry is farming: agriculture, forestry, animal husbandry, sideline fishery production (such as farming, breeding chickens, ducks, aquatic products, etc.)), X80 (whether the industry is a construction industry), X83 (whether the industry is refusal to answer), X85 (whether the industry is the production and supply of electricity, gas and water), X86 (whether the industry is a social service industry), X87 (whether the industry is scientific research and comprehensive technical service industry), X89 (whether the industry is financial and insurance), X91 (whether employer is a public institution), X94 (whether employer is agriculture, forestry, animal husbandry, sideline and fishery production), X95 (whether employer is a state-owned/collective institution), X96 (whether employer is a state-owned enterprise), X100 (whether employer is a private non-enterprise, a social organization, etc.), X101 (whether employer is a private or private enterprise), X102 (whether employer type is freelance workers (freelancers, casual workers, vendors, babysitters without dispatch units, self-operated drivers, manual craftsmen, etc.)), X103 (whether employer is a collective enterprise)
Capital income	X1 (whether they understand a foreign language), X2 (whether they have a professional qualification certificate), X5 (level of local dialect), X7 (whether engaged in agricultural production), X10 (whether they live in Yunnan), X14 (whether they live in), X15 (whether they live in Tianjin), X16 (whether they live in Ningxia), X17 (whether they live in Anhui), X18 (whether they live in Shandong), X19 (whether they live in is it Shanxi), X20 (whether they live in Guangdong), X21 (whether they live in Guangxi), X23 (whether they live in Jiangsu), X24 (whether they live in Jiangxi), X27 (whether they live in Zhejiang), X30 (whether they live in Gansu), X31 (whether they live in Fujian), X32 (whether they live in Guizhou), X35 (whether they live in Shaanxi), X36 (whether they live in Qinghai), X42 (whether education level is elementary school), X43 (whether education level is undergraduate), X44 (whether education level is undergraduate), X45 (whether education level is master's), X46 (whether education level is high school), X49 (whether political outlook is the masses), X50 (whether health level is average), X52 (whether health level is relatively unhealthy), X53 (whether health level is very unhealthy), X54 (whether health level is very healthy), X58 (whether father's education level is PhD), X59 (whether father's education level is junior high school), X61 (whether father's education level is no school), X64 (whether father's education level is high school), X67 (whether mother's education level is junior high school), X68 (whether mother's education level is college), X72 (whether mother's education level is high school)

4.2. Implementation Process and Regression Results

This study used the step function in the R software (Auckland, New Zealand) to implement the stepwise regression method and used the PLSR library to implement the partial least square method. The compilation environment is R version 3.5.0 (RStudio Version 1.1.419). The regression equation was established as follows:

$$y_g = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_i x_i + \varepsilon_g$$

Among the variables, y_g is wage income, x_i is the independent variable of wage income, and i is the number of independent variables.

$$y_z = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \cdots + \beta_m z_m + \varepsilon_z$$

Among these variables, y_z is capital income, z_m is the independent variable of capital income, and m is the number of independent variables.

4.2.1. Stepwise Regression Method Results

The stepwise regression method can select a variable adjustment scheme corresponding to the smallest AIC value by observing the change in AIC after a certain variable is deleted. The specific results are detailed below.

It can be seen in Table 5 that after the application of the stepwise regression method, the number of independent variables of wage income dropped from 49 to 29, while the number of independent variables of capital income dropped from 33 to 25; the effect of dimensionality reduction was better. The regression results are as follows (Table 5):

Table 5. Stepwise regression method variable selection.

	Stepwise Regression	
	Number of Original Independent Variables	**Number of Current Independent Variables**
Wage income	49	29
Capital income	33	25

The overall *p*-value of the wage income model is less than 0.01, and the fitting effect is good. From the perspective of human capital, X2 (whether they have a professional qualification certificate), X5 (local dialect level), X42 (whether education level is elementary school), X43 (the coefficient of the variable X44 (whether the education level is an undergraduate)) are the most significant. The coefficients of the independent variables of X2, X5, and X44 are positive, while the coefficients of X42 and X43 are negative. This shows that, given other factors remain unchanged, the number of professional qualification certificates and a high level of local dialect education have a greater degree of positive impact on wage income. In comparison, few academic qualifications has a greater degree of negative impact on wage income.

Individuals who have a large number of professional qualifications and proficiency in various vocational skills, along with a high level of education, tend to have higher wage incomes. This phenomenon reflects the growing demand for high-quality talent in the job market. Additionally, the regional characteristics of China's multi-dialect system may play a role in determining the incomes of individuals. People who are proficient in their local dialects tend to be more favored in the job market, which makes it easier for them to obtain higher wage incomes. In conclusion, holding professional qualifications, having strong communication skills, and having a good level of education are important factors in determining an individual's wage income in China.

The overall *p*-value of the capital income model is less than 0.01, and the fitting effect is good; among them, the coefficients of the variables X7 (whether engaged in agricultural production), X42 (whether education level is elementary school), and X43 (whether education level is never attended school) are the most significant. Moreover, these are all negative. This shows that low education levels and engaging in agriculture have a greater degree of negative impact on capital income. Those with low education levels and agricultural practitioners have lower incomes.

4.2.2. Partial Least Squares Regression Results

The partial least squares method can manually select variables by observing the contribution rate of principal components to each variable. However, from a data point of

view, due to the small difference in the contribution rates of the components of the partial least squares method, this paper did not further extract variables. The regression results are detailed below.

The overall *p*-value of the wage income model is less than 0.01, and the fitting effect is good; among them, the coefficients of the variables X2 (whether they have a professional qualification certificate), X42 (whether education level is elementary school), and X44 (whether education level is undergraduate) are the most significant. Among these, the coefficients of X2 and X44 are positive, and the coefficients of X42 are negative. This shows that holding many professional qualification certificates and having a high level of education can help individuals obtain a higher wage income.

The overall *p*-value of the capital income model is less than 0.01, and the fitting effect is good; the coefficients of the variables X42 (whether education level is elementary school), X43 (whether education level is no school), X50 (whether health level is average), X52 (level of health) are the most significant, and all are negative. This shows that low education and poor health have a relatively high degree of negative impact on the capital incomes of residents. This shows that society needs more healthy workers with high education who can obtain high capital incomes.

Combining the results of the two regression methods, we find that having a large number of professional qualification certificates, a high level of local dialect, and a high level of education helps residents obtain higher wage incomes. Meanwhile, residents with poor health and a low education level have lower capital incomes.

The regression results show that society has a higher demand for workers with high education levels, high professional skills, and good health, reflecting the importance of human capital. There is a certain gap in the rate of return between capital income and wage income, which needs to be adjusted.

4.3. Comparative Analysis of the Prediction Capabilities of the Two Regression Methods

In order to better observe the prediction effects of the regression models, we made a prediction for the 2016 data and compared it with the test set; that is, the 2016 CLDS data. The prediction results and comparison chart are detailed below.

4.3.1. Stepwise Regression Method to Predict Results

According to Figures 1–4, the stepwise regression method and the partial least squares method have a good degree of prediction for wage income and capital income. In addition, Table 6 reports the amount of error of the two regression methods:

Table 6. Error comparison table of stepwise regression method and partial least square method.

Method	Dependent Variable	Data Set	MAD	MSE	MAPE
Stepwise regression method	Wage income	Training set	0.275988	0.583369	0.062372
		Test set	0.288995	0.770744	0.076404
	Capital income	Training set	0.373707	0.973191	0.084687
		Test set	0.401561	0.963245	0.089789
Partial least square method	Wage income	Training set	0.267251	0.574956	0.061624
		Test set	0.28743	0.769891	0.076593
	Capital income	Training set	0.387758	1.020318	0.086448
		Test set	0.390108	0.924637	0.088307

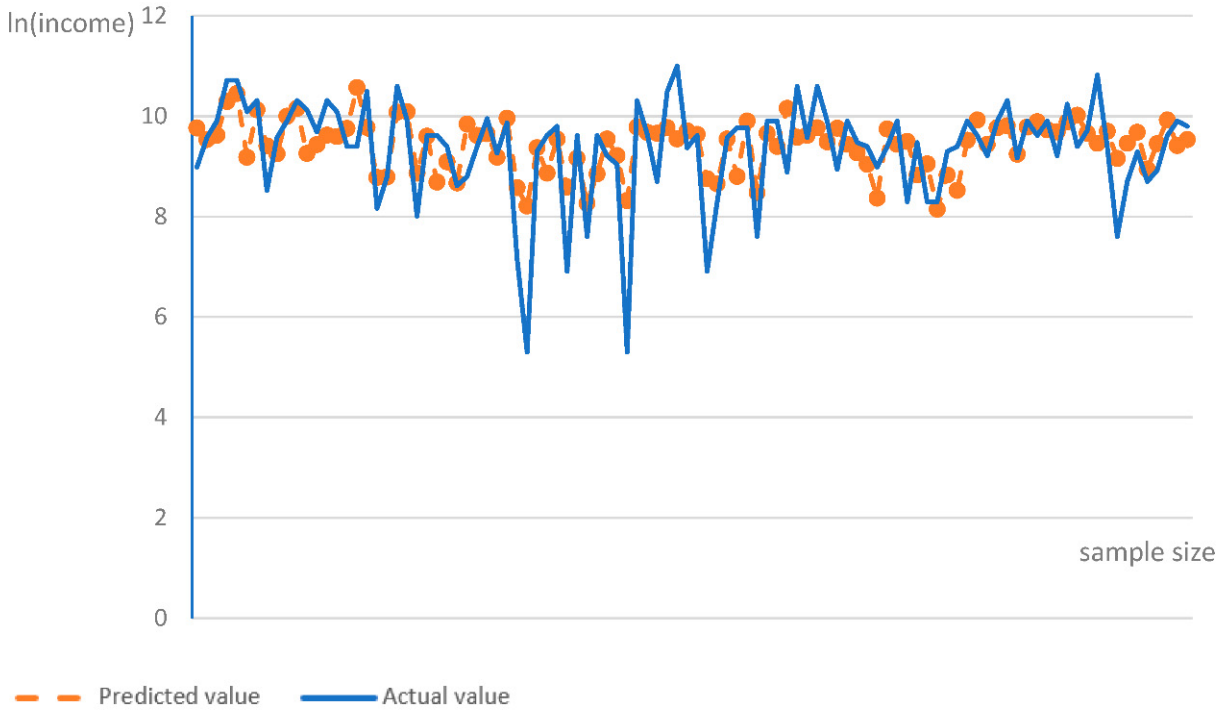


Figure 1. Wage income forecast via stepwise regression vs. actual value comparison chart.

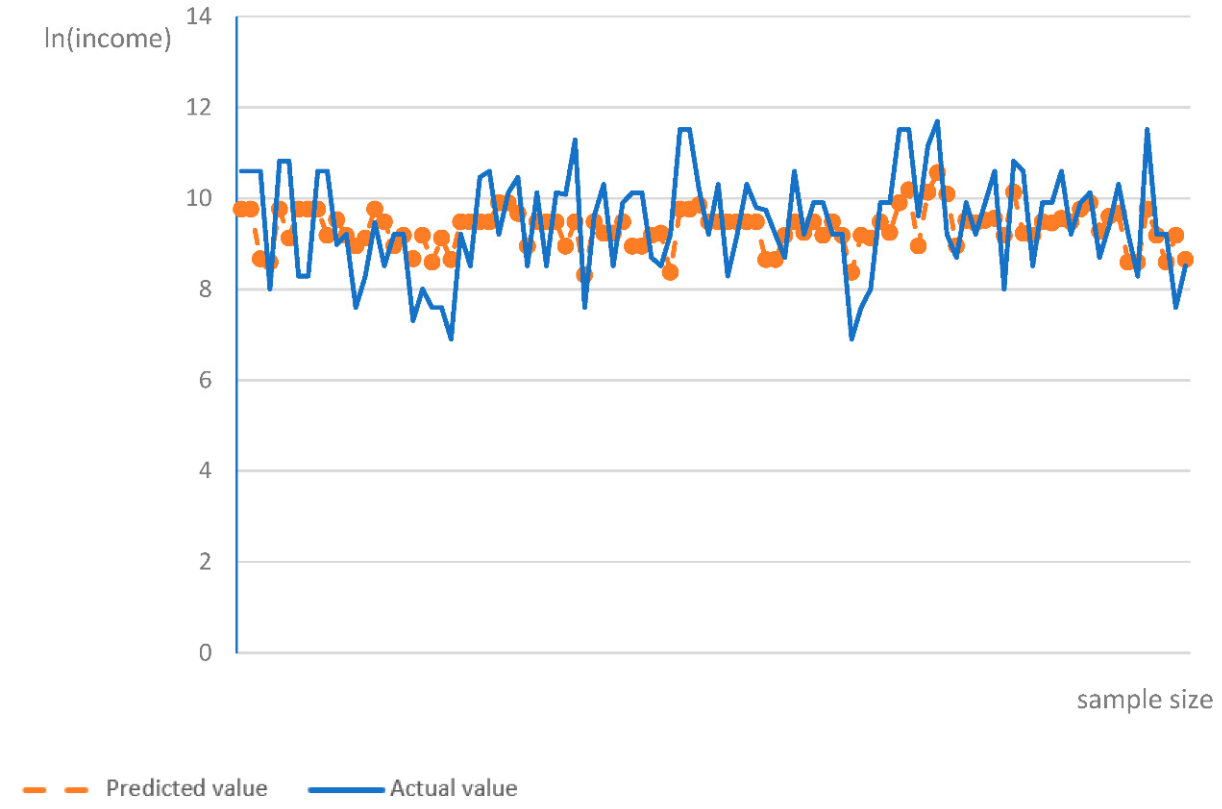


Figure 2. Capital income forecast via stepwise regression vs. actual value comparison chart.

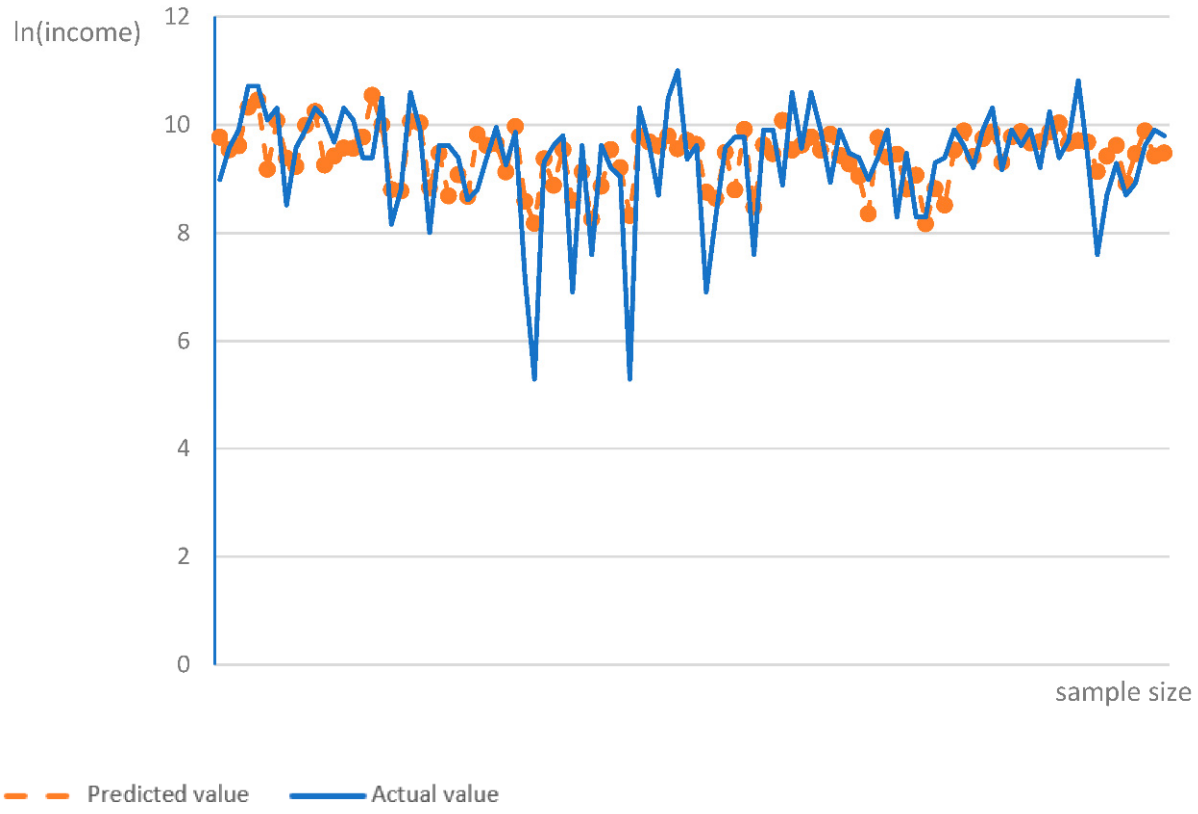


Figure 3. Wage income forecast by partial least square method vs. actual value comparison chart.

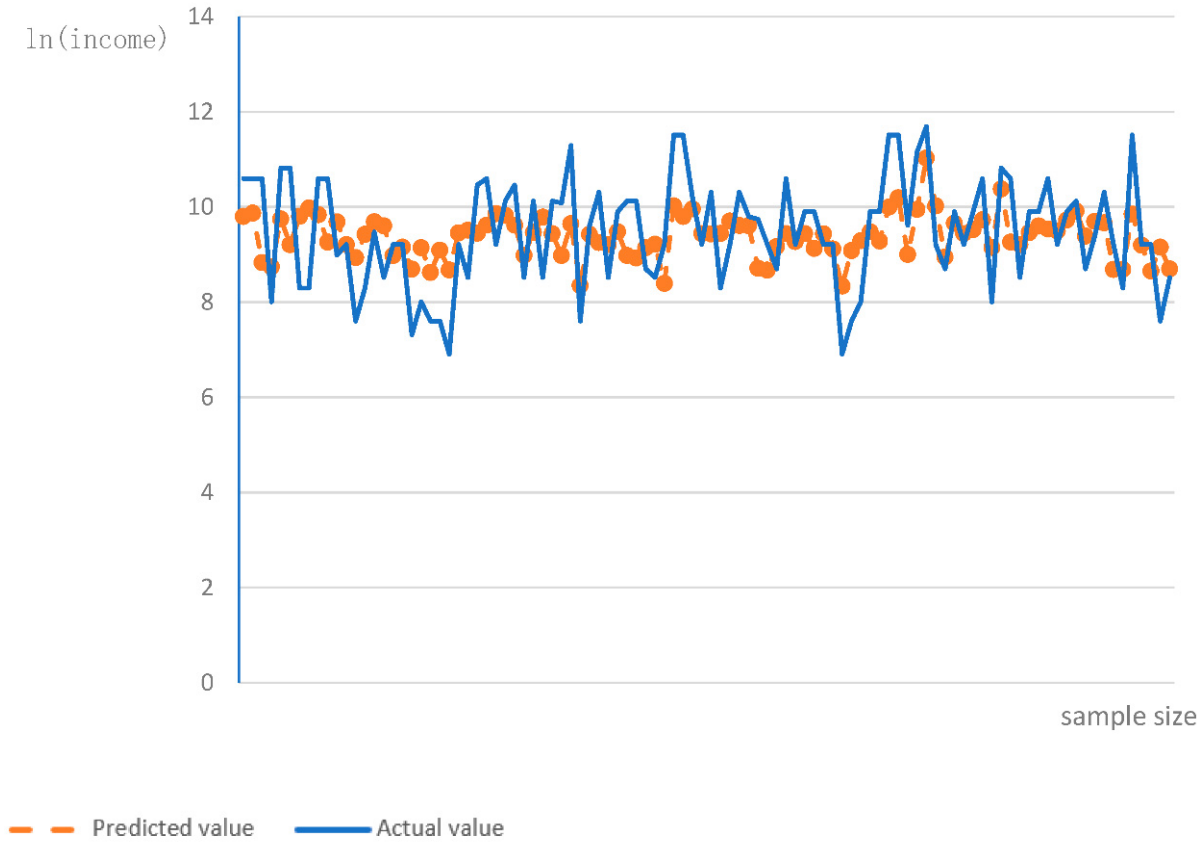


Figure 4. Capital income forecast via partial least squares method vs. actual value comparison chart.

4.3.2. Partial Least Squares Prediction Result

MAD (Median Absolute Deviation) refers to the median of the absolute value of the residual median deviation. In statistics, MAD is a measure of statistical dispersion, and it is a robust measure of the variability of a univariate sample of quantitative data.

$$MAD = median(|X_i - median(X)|)$$

where X_i is the ith value in the dataset and $median(X)$ is the median value in the dataset.

MSE (Mean Squared Error) refers to the mean value of the sum of squared errors of the predicted data and the corresponding points of the original data.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2$$

where n is the sample size of the dataset, y_i is the ith actual value in the dataset, and $\hat{y}$ is the prediction value in the dataset.

MAPE (Mean Absolute Percentage Error) refers to the average of the absolute value of the relative percentage error.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)}{y_i}$$

It can be seen from Table 5 that the error index values of the two methods are small, the values are less than 1 or close to 1, and the differences are within 0.2. Therefore, both the stepwise regression method and the partial least squares method can be used to obtain reasonable income forecasts.

5. Conclusions and Policy Recommendations

Constructing a scientific and practical talent development system is an important driving force for promoting the development of an innovative society. Over the past 40 years, China's economic growth has created a miracle that has attracted worldwide attention. However, with the gradual disappearance of the demographic dividend, a more scientific talent development system has been built to increase the human capital stock affecting the entire society and increase residents' income, as well as narrow the gap between regions and industries. This is imperative. This also has reference value for other developing countries.

This article explains the essential human capital characteristics affecting wage and capital income. Research on Chinese labor force dynamics survey data through statistical methods such as Lasso dimensionality reduction, stepwise regression, and the partial least squares method found that society has a high level of education and high education. In today's fast-paced and highly competitive job market, there is a growing demand for individuals who possess a unique combination of professional skills, good health, and strong communication abilities. These qualities are seen as critical in many industries, and they are valued by employers as they help increase productivity and performance. As a result, individuals who possess these human capital characteristics are more likely to find high-paying job opportunities and achieve greater financial success. Moreover, the development of these human capital traits can have a lasting impact on an individual's career and earning potential. A strong foundation of professional skills can help individuals stay ahead in their field and remain competitive, while good health and strong communication abilities can help them maintain positive relationships with coworkers and clients. These traits are highly sought after by employers, and individuals who possess them are more likely to earn higher salaries, receive promotions, and advance in their careers. Therefore, it is crucial for individuals to invest in their human capital and cultivate these important traits. This can be achieved through education, training, and personal development, and it can help individuals achieve greater financial success and stability in their careers. By

focusing on developing their human capital, individuals can take control of their future and maximize their earning potential in the job market. This verifies the research conducted by Mincer [27]. At the same time, in China, there is a significant difference between the rate of return on capital income and the rate of return on wage income. The rate of return on capital income is significantly higher than that of wage income; differences in industries and regions also significantly affect wage and capital income. Compared with traditional industries, high-tech industries have higher wage incomes. These facts show that we need an excellent talent development system to increase residents' incomes while narrowing the income gap between regions and industries to balance the rate of return on wage income and capital income as much as possible [28–30].

Therefore, the government must pay attention to the role of human capital and build a good talent development system. The Chinese government is also attaching more and more importance to vocational education and human capital development and put forward the document *Opinions on Deepening the Reform of Modern Vocational Education System Construction* in 2022.

Therefore, we put forward the following suggestions. First, the government should increase public education investment in the whole of society and improve vocational skills training to develop the core competitiveness of residents at work and the accumulation of human capital across the whole of society. It should also increase residents' human capital by subsidizing vocational education and enlarging public education expenditures to match talent development within the social scope with social needs. Second, we must focus on developing high-tech industries and introducing more workers from traditional industries into high-tech industries. The technology industry, through inter-industry mobility, can increase residents' wage income and narrow the income gap between industries. Moreover, through the support of high-tech industries, more workers will be guided to move among them, by which means residents' income and national economic competitiveness can be increased. Third, the government should also strengthen the promotion of common language, reduce transaction costs caused by communication and communication barriers, and reduce communication barriers caused by dialects as much as possible. Because of this, it is difficult for individuals to utilize their human capital, hindering talent development fully. Fourth, the government should establish a sound worker flow system, promote inter-regional population mobility, introduce more high-quality workers from outside, improve overall economic competitiveness in competition and communication between workers, and reduce the inter-regional income gap through this system.

The aim of this article was to evaluate the impact of income factors using statistical methods such as Lasso dimensionality reduction, stepwise regression, and partial least squares. The 2016 data was used as a test set and the results showed a good fit with the actual data, suggesting that the introduction of this method could enhance the empirical methods in this field. Furthermore, the article also compares the impact of capital income based on prior research and provides relevant policy recommendations for the talent development system, thereby enriching the research outcomes in this field.

In addition, this article introduces the use of machine learning techniques based on micro-data to verify and strengthen the impact of traditional human capital on income, and offers insightful policy recommendations for China's current situation. However, further research is still needed to optimize the talent development system and to understand the impact of the rapid development of the internet on the structure of human capital in the new era.

Author Contributions: Conceptualization, X.H and X.Y.; methodology, X.Y.; software, X.Y.; formal analysis, X.H and X.Y.; writing, X.H., K.K. and Y.Z.; supervision, C.G.; funding acquisition, K.K. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Major Project in 2020 under the 13th Five Year Plan of National Education Science: Research on the Characteristics of Vocational Education Type and the Construction of the Dual track and Dual channel System with General Education: VJA200003;

The Research Project for youth of Beijing Office of Education Sciences Planning: Research on the construction of quality assurance system of vocational undergraduate education: BDCA22103.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sakalas, A.; Liepe, Z. Human capital system evaluation in the context of the European Union countries. *Inžinerinė Ekon.* **2013**, *24*, 226–233. [CrossRef]
2. Bradley, S.; Taylor, J. Human capital formation and local economic performance. *Reg. Stud.* **1996**, *30*, 1–14. [CrossRef]
3. Zula, K.J.; Chermack, T.J. Integrative literature review: Human capital planning: A review of literature and implications for human resource development. *Hum. Resour. Dev. Rev.* **2007**, *6*, 245–262. [CrossRef]
4. Tilak, J.B.G. Vocational Education and Training in Asia. In *International Handbook of Educational Research in the Asia-Pacific Region: Part One*; Keeves, J.P., Watanabe, R., Maclean, R., Renshaw, P.D., Power, C.N., Baker, R., Gopinathan, S., Kam, H.W., Cheng, Y.C., Tuijnman, A.C., Eds.; Springer: Dordrecht, The Netherlands, 2003; pp. 673–686. [CrossRef]
5. Wallenborn, M. Vocational Education and Training and Human Capital Development: Current practice and future options. *Eur. J. Educ.* **2010**, *45*, 181–198. [CrossRef]
6. Zhu, X. *Research on Incentive Mechanism for Innovative Talents*; China Economic Publishing House: Beijing, China, 2013.
7. Malik, K. Human Development Report 2013. The Rise of the South: Human Progress in a Diverse World. In The Rise of the South: Human Progress in a Diverse World (15 March 2013); UNDP-HDRO Human Development Reports. 2013. Available online: https://www.undp.org/egypt/publications/human-development-report-2013-rise-south-human-progress-diverse-world (accessed on 20 December 2022).
8. Ter Beek, M.; Wopereis, I.; Schildkamp, K. Don’t Wait, Innovate! Preparing Students and Lecturers in Higher Education for the Future Labor Market. *Educ. Sci.* **2022**, *12*, 620.
9. Schultz, T.W. Investment in human capital. *Am. Econ. Rev.* **1961**, *51*, 1–17.
10. Mincer, J. Investment in human capital and personal income distribution. *J. Political Econ.* **1958**, *66*, 281–302. [CrossRef]
11. Becker, G.S. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*; University of Chicago press: Chicago, IL, USA, 2009.
12. Shastry, G.K.; Weil, D.N. How much of cross-country income variation is explained by health? *J. Eur. Econ. Assoc.* **2003**, *1*, 387–396. [CrossRef]
13. Weil, D.N. Accounting for the effect of health on economic growth. *Q. J. Econ.* **2007**, *122*, 1265–1306. [CrossRef]
14. Gao, M.; Yao, Y. The Micro Foundation of Farmer Income Gap: Physical Capital or Human Capital. *Econ. Res.* **2006**, *41*, 10.
15. Schochet, P.Z. A Lasso-OLS hybrid approach to covariate selection and average treatment effect estimation for clustered RCTs using design-based methods. *arXiv* **2020**, arXiv:2005.02502.
16. Siemens, G.; Baker, R.S.d. Learning analytics and educational data mining: Towards communication and collaboration. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, New York, NY, USA, April 29–2 May 2012; pp. 252–254.
17. Tripney, J.; Hombrados, J.; Newman, M.; Hovish, K.; Brown, C.; Steinka-Fry, K.; Wilkey, E. Technical and vocational education and training (TVET) interventions to improve the Employability and employment of young people in Low-and Middle-Income countries: A systematic review. *Campbell Syst. Rev.* **2013**, *9*, 1–171. [CrossRef]
18. Xie, Z.; Liu, Y. Deepening Industry-education Integration and Promoting Revolution of Vocational Education——Strategic thinking on development of new technology application personnel in higher vocational colleges. *China High. Educ. Res.* **2018**, *6*.
19. Hanushek, E.A.; Schwerdt, G.; Woessmann, L.; Zhang, L. General education, vocational education, and labor-market outcomes over the lifecycle. *J. Hum. Resour.* **2017**, *52*, 48–87. [CrossRef]
20. Meer, J. Evidence on the returns to secondary vocational education. *Econ. Educ. Rev.* **2007**, *26*, 559–573. [CrossRef]
21. Loyalka, P.; Huang, X.; Zhang, L.; Wei, J.; Yi, H.; Song, Y.; Shi, Y.; Chu, J. The impact of vocational schooling on human capital development in developing countries: Evidence from China. *World Bank Econ. Rev.* **2016**, *30*, 143–170.
22. Aizenman, J.; Jinjarak, Y.; Ngo, N.; Noy, I. Vocational education, manufacturing, and income distribution: International evidence and case studies. *Open Econ. Rev.* **2018**, *29*, 641–664. [CrossRef]
23. Piketty, T. *Capital in the 21st Century*; President and Fellows, Harvard College: Cambridge, MA, USA, 2013.
24. Piketty, T.; Saez, E. Income and wage inequality in the United States, 1913–2002. In *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries*; Oxford University Press: Oxford, UL, USA, 2007; Volume 141.
25. Linoff, G.S.; Berry, M.J. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. John Wiley & Sons: Hoboken, NJ, USA, 2011.

26. Cai, H. *China Labor-Force Dynamics Survey: 2017 Report*; Social Sciences Academic Press: Beijing, China, 2017.
27. Mincer, J.A. Schooling, Experience, and Earnings. In *Education, Income, and Human Behavior*; NBER: Cambridge, MA, USA, 1974.
28. Milanovic, B. *Increasing Capital Income Share and Its Effect on Personal Income Inequality*; LIS Working Paper Series; Harvard University Press: Cambridge, MA, USA, 2016.
29. Chao, F.; Bin, H. Could the Investment of Education Human Capital Reduce the Wage Gap of Rural Residents? *Educ. Econ.* **2017**, *9*.
30. Pereira, P.T.; Martins, P.S. Does Education Reduce Wage Inequality? Quantile Regressions Evidence from Fifteen European Countries. Discussion Papers. 2000. Available online: https://docs.iza.org/dp120.pdf (accessed on 20 December 2022).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.