

Research on the Economic Growth for Under-Developed Counties -- Based on Two-way Fixed-effects Model

Zhengze Li^{1,*}, Zibo Huang², Yongcheng Zhang³, Zili Huang⁴

¹ Chinese University of Hong Kong, Shenzhen, Shenzhen, 518172, China

² China University of Political Science and Law, Beijing, 102249, China

³ Chinese University of Hong Kong, Shenzhen, Shenzhen, 518172, China

⁴ Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA

*Email: lzzcupl@126.com

Abstract. The county economy, especially the problem of poor counties, is currently a hot issue in China. At present, the research on the key factors for county economy is scarce, partly because it is difficult to obtain accurate and sufficient data. Because power consumption and industrial level are often highly correlated, and the power consumption is real-time data and hard to falsify, this paper used spatial, electricity consumption and economic data of 66 Chinese counties, which contains 29 poor counties, from 2009 to 2016 to find out the key factors for under-developed counties' economic growth by applying fixed effect regression model and machine learning models. The result shows that for poor counties, though the 1st industry is still the fundamental industry for county-level economy, the development of 3rd industry has significant positive impact on local economy. However, the development of 2nd industry, including recruiting large companies, and the input of electricity resources cannot well drive the local economies, which may suffer great loss due to the elimination policy of overcapacity in recent years. And the machine learning results support the above conclusions and suggest an obvious geographical cluster of poor counties, and the location, development of 1st and 3rd industry and net income of rural residents can explain most difference between the poor and non-poverty counties. These conclusions can be helpful for the government to lead the poor counties to get rid of poverty, and the cluster of poor counties should be focused.

1. Introduction

With the progress of poverty alleviation, the issues of economic growth for under-developed county have attracted more and more attentions. As the foundation of China's economy, the economy of under-developed counties has significant impacts on the economic growth of China, and the successful development of them will greatly affect whether China can achieve the leap and sustainable growth of economy.

There are some methods and ideas to find factors that affect economic growth of county, and the factors that the researchers mainly focused on are the political, social and some other macro indicators. For example, Liu examines the impacts of fiscal decentralization on the county long-run economic development, and the result indicates that the Chinese county development is frustrated by the fiscal decentralization system [1]. Wu used the difference-in-difference method to analyse the casual relationship between the China's province-managing-county reform (PMC) and the economic growth and finds that the PMC has a positive effect on economic growth of county [2]. Yu employs the quasi-experimental design to argue that the peripheral counties along the upgraded railway lines



experienced reductions in GDP and GDP per capita due to the concurrent drop of fixed asset investment [3]. Zhang *et al.* proves that county economic growth and the county financial development has a positive u-curve relationship based on the panel data of 1732 counties in China from 2013 to 2016 [4].

From the researches above, it can be seen the current researches mainly focus on exploiting the empirical models to investigate the effect of one or more economic factors on the county economy growth. Nevertheless, there are few researches paying attention on the effect of industrial structure on county economy, partly because it is difficult to obtain accurate and sufficient data. On the basis of neoclassical economic growth theory, the industrial structure plays an important role in the economic growth. Through the research of county industrial structure, it will be conducive to finding out the reasons for the differences of economic growth between poor counties and non-poor counties.

Due to the close correlation between modern industrial production and energy, electricity consumption will be helpful for us to find out the key factor for under-developed counties' economy and their industrial structure. Moreover, the power consumption is real-time data and hard to falsify, which makes it undoubtedly a great indicator to reflect economy, and more direct and objective than other economic statistics. In practice, there are also many methods to apply electricity consumption to measure the economy, like the Li Keqiang Index, which is proved to be useful and practical [5]. Huang *et al.* focus on the electricity consumption of county-level tourism industry and establishes a forecasting model [6]. Shiu and Lam exploit the error-correction model to examine the casual relationship between real GDP for China and electricity consumption during 1971-2000 [7]. Wolde-Rufael also tests the long-run and casual relationship between GDP per capita and electricity consumption per capital for 17 African countries from 1971 to 2001 [8].

Though researches on large observation like a nation or province using the electricity consumption are relatively sufficient, there are few studies with large sample pool and small observation on counties in the field of economic structure and growth. Therefore, this paper uses big data on electricity consumption to analyse China's current targeted poverty alleviation and problems in the sustainable development of the county economy, and on this basis, provides us a new perspective to discuss the poverty issues.

2. Data Description

The raw dataset comes from the State Grid Corporation of China. The dataset is well sorted into panel data with 30 electricity consumption variables of many subdivided industries variables, which includes annual electricity consumption data of 70 subdivisions of 66 counties in Hebei, Henan and Hunan provinces from 2009 to 2016. Moreover, the corresponding economic, spatial and policy data of these counties are collected from the official statistics bureau of each region.

In this paper, the growth rate of GDP and 2nd industry added value will be chosen as the dependent variables to measure the local economic outcome, and the longitude and latitude of the county will be used to measure the spatial influence.

3. Methods

The fixed effects regression model is a type of panel data analysis method that varies with individuals but not with time in panel data. It helps to control the unobserved heterogeneity [9].

From the perspective of time and individuals, the explanatory variable of the individual fixed-effects regression model can represent its marginal influence on the explained variable, and the effects of all other variables that are not included in the regression model or unobservable but affect the explained variable can be controlled. In the end, it appears as a fixed effect that varies with the individual but not with time. Besides, the economic data always contains a strong time tendency, and the same holds for fixed-time effect. With the addition of fixed-time effect, the model comes to a two-way fixed-effects model as shown in Figure 1.

Based on the above analysis, the empirical model used in this study is a balanced panel data individual fixed effects regression model. The regression method uses a multiple stepwise regression method. The regression model is constructed as follows:

$$y_{it} = (1 + Poverty_i) \left(\sum_{j=1}^J \beta_j Economy_{ji(t-1)} + \sum_{k=1}^K \beta_k Electricity_{kit} \right) + \lambda_i + \mu_t + u + \varepsilon_{it} \quad (1)$$

The multiple regression equation is as follows:

$$E(y_{it}) = (1 + Poverty_i)(\sum_{j=1}^J \beta_j Economy_{ji(t-1)} + \sum_{k=1}^K \beta_k Electricity_{kit}) + \lambda_i + \mu_t + u \quad (2)$$

Where i is the i^{th} county, t is the t^{th} period, j/k is the j^{th}/k^{th} explanatory variable, y_{it} is the explained variable, $Poverty_i$, $Economy_{ji(t-1)}$ and $Electricity_{kit}$ are the explanatory variables, λ_i is the county fixed effect, μ_t is the period fixed effect, β_j/β_k is the partial regression coefficient, and u is the regression constant, ϵ_{it} is random error.

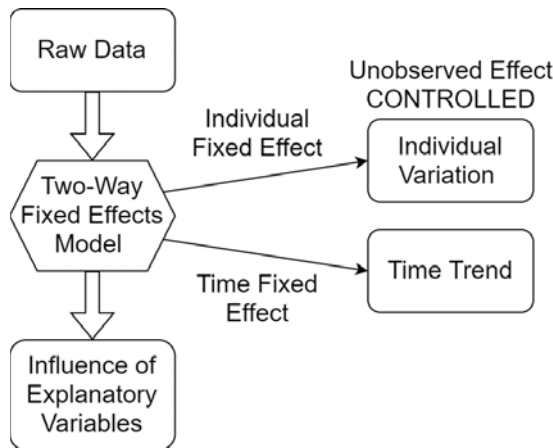


Figure 1. Principle of Two-way Fixed-effects Model.

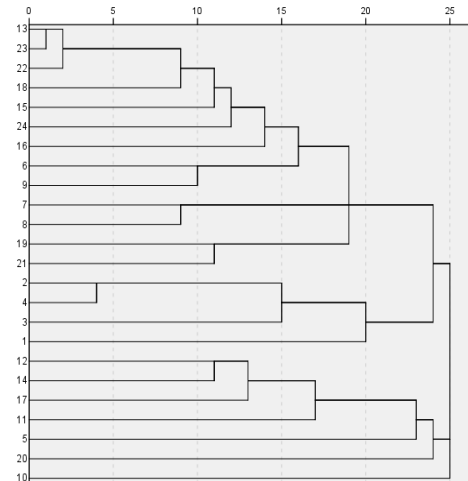


Figure 2. Pedigree diagram of the clustering.

4. Data Pre-processing

4.1. Data Cleaning

The data set was first cleaned, including sorting data format, eliminating null values and outliers, organizing variable labels and so on.

4.2. Variables Clustering

In the panel data analysis of this paper, too many explanatory variables can easily lead to the problems of multicollinearity and over-fitting, so it is necessary to analyse and refine the independent variables and reduce the dimension. This study uses the method of systematic clustering, and uses Pearson correlation as the distance between variables. The 24 electrical sub-industry variables are clustered, and the pedigree diagram is shown in the Figure 2.

Table 1. Result of the clustering of electricity consumption variables.

Clusters		Percentage of Power Consumption
Cluster1	Agriculture	4.19%
Cluster2	Service industry	9.95%
Cluster3	Industry and infrastructure	84.99%
Cluster4	Transportation industry	0.79%
Cluster5	Technical industry	0.08%

Based on the results, these 24 variables are grouped into five clusters, and they are named by their main components in Table 1. It can be seen that Industry and infrastructure accounts for 85% of total industrial electricity consumption. The Pearson correlation between each cluster has been largely reduced compared to the raw data.

4.3. Dummy Variables

Besides, the Longitude and latitude of the counties are used to measure the location of each county, and they are also clustered as 2 clusters. And a dummy variables *North* are created to represent the 2 location clusters. And notice that since 2014, the Chinese government has started to eliminate of backward and overcapacity industrial sector, which must have a large impact on the industrial outcome. Hence, a dummy variables *Policy* are created to represent the policy effect since 2014.

$$North = \begin{cases} 1, & \text{if } latitude_i \geq 30 \\ 0, & \text{if } latitude_i < 30 \end{cases} \quad Policy = \begin{cases} 1, & \text{if } year_t \geq 2014 \\ 0, & \text{if } year_t < 2014 \end{cases} \quad (3)$$

5. Modelling and Analysis

5.1. OLS Linear Regression

This paper first tried the OLS linear regression analysis for variables *G_GDP* and *G_GDP_2ND*. The variable of electricity doesn't lag by one period because they are not significant. The results are sorted in the Table 2. It shows that the two models are significant with the R-squared of 0.69 and 0.35.

Table 2. Summary of OLS Linear Regression Analysis for Model 1 and Model 2 (N = 462).

Variable	Model 1: G_GDP			Model 2: G_GDP_2ND		
	<i>B</i>	<i>SE B</i>	β	<i>B</i>	<i>SE B</i>	β
NORTH	-0.03	0.01	0.00***	-0.04	0.01	0.00***
POLICY	-0.04	0.01	0.00***	-0.09	0.01	-0.01***
G_GDP_1ST(-1)	0.19	0.06	0.08***	0.14	0.08	0.09*
G_GDP_2ND(-1)	0.07	0.05	0.02	0.16	0.06	0.08
G_GDP_3RD(-1)	-0.08	0.01	0.00***	0.01	0.01	0.00
G_NUM_BIG_COMPANY(-1)	0.05	0.03	0.01	0.24	0.04	0.08***
G_RURAL_INCOME(-1)	-0.05	0.07	-0.02	-0.19	0.09	-0.13**
G_TOTAL_ELEC_CON	-0.06	0.03	-0.01**	0.04	0.04	0.01
G_SERVICE_CON	0.02	0.00	0.00***	0.00	0.00	0.00
G_INDUSTRY_CON	0.05	0.02	0.01*	-0.01	0.03	0.00
G_TRANSPORT_CON	0.00	0.00	0.00	0.00	0.00	0.00
POVERTY * X						
POVERTY * G_GDP_1ST(-1)	-0.15	0.09	-0.09*	-0.16	0.12	-0.15
POVERTY * G_GDP_2ND(-1)	-0.05	0.06	-0.02	-0.15	0.08	-0.10*
POVERTY * G_GDP_3RD(-1)	0.10	0.06	0.04*	0.06	0.08	0.04
POVERTY * G_NUM_BIG_COMPANY(-1)	-0.03	0.04	-0.01	-0.23	0.06	-0.10***
POVERTY * G_RURAL_INCOME(-1)	-0.04	0.07	-0.02	0.10	0.09	0.07
POVERTY * G_TOTAL_ELEC_CON	0.13	0.07	0.06**	0.06	0.09	0.04
POVERTY * G_SERVICE_CON	-0.02	0.00	0.00***	0.00	0.00	0.00
POVERTY * G_INDUSTRY_CON	-0.03	0.04	-0.01	0.06	0.06	0.02
POVERTY * G_TRANSPORT_CON	0.00	0.00	0.00	0.00	0.01	0.00
C	0.14	0.01	0.01***	0.16	0.02	0.02***
R ²		0.69			0.35	
F		48.67***			11.92***	

Note: Variable X_{t-1} was represented as $X(-1)$, variable named with G indicates the growth rate of it, variable named with CON indicates the electricity consumption of it. * $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$, same below.

Comparing the results of the two models, it can be found that per capital net income of rural residents has a significant negative effect on the added value of 2nd industry while the effect on GDP is not significant, and nor is the interaction of poverty and income. It may reflect the fact that the progress of

county-level industrialization and urbanization in the non-poor counties occupied part of agricultural resources, like rural labor force and land, which makes the development of 2nd industry negatively correlated with the income of rural residents.

Besides, it's notable that the growth of total electricity consumption has a significant positive effect on the growth of GDP in poor counties but a significant negative effect in non-poor counties, which indicates that electricity resources still have relatively high marginal utility for economic growth in poor counties.

It also can be seen that the *NORTH* and *POLICY* all have significant effects on these two models, and in model 2, both *POLICY* and *NORTH* have a relative smaller β , which indicates that the growth of added value of 2nd industry would suffer more negative impact by the location and overcapacity policy. It can be inferred that the growth rate of GDP would be largely influenced by county's location and time, and the simple linear regression model cannot well measure such impact.

Hence, County and Period Fixed Effects Regression was conducted to better measure the marginal impact of each variable.

5.2. Two-way Fixed Effects Model

This paper used two-way fixed effects model to fit the data. 203 observations of poor counties and 259 observations of non-poverty counties are regressed respectively.

Table 3 shows that it's suitable to establish cross-section and period fixed model for *G_GDP* and *G_GDP_2ND* based on the F test result. Note that the effects of variables *NORTH* and *POLICY* will be absorbed by the county and period fixed effects, so they are excluded in this part. The results of Model 3 and 4 are sorted in the Table 4, and the period fixed effect of Model 3 is shown in Figure 3.

Table 3. Summary of Cross-section and Period Fixed Effects Test for Model 3 and 4.

Effects Test	Model 3	Model 4
	<i>Statistic (d.f.)</i>	<i>Statistic(d.f.)</i>
Cross-Section/ Period F	2.13*** (-71372)	3.89*** (-71372)
Cross-Section/ Period Chi- square	157.36*** (71)	256.47*** (71)



Figure 3. Period Fixed Effect of Model 3.

From Table 4, it can be seen that the two models are generally significant, and reached to better R-squared of 0.77 and 0.58, which greatly improves the explanatory ability.

In year 2015, the central government of China had paid much more attention to the elimination of overcapacity and out-dated industry. Also, the period fixed effect of Model 3 shows that the growth rate of GDP does suffer a common reduction even since 2012, and reached its lowest point in 2015 (see Figure 3), which follows our common sense

According to the regression result, for both poor and non-poor county, the growth rate of GDP will be largely influenced by the growth rate of 1st and 3rd industry added value of last period, but it has nothing to do with that of 2nd industry. However, the growth of 3rd industry has significant negative impact on non-poor counties but general significant positive impact on poor counties, which means the development of 3rd industry may be a more effective path for poor counties' economy compared with 2nd industry. Moreover, in model 4, the increase of number of industrial enterprises above state designated scale has a significant coefficient of 0.08 for non-poor county but a sum-up coefficient of -0.02 for poor county. That is to say, the increase in large industrial enterprises cannot effectively bring the expected industrial growth to the local economy.

Table 4. County and Period Fixed Effects Regression Analysis for Model 3 and 4 (N = 462, P=7).

Variable	Model 3: G_GDP			Model 4: G_GDP_2ND		
	B	SE B	β	B	SE B	β
G_GDP_1ST(-1)	0.14	0.07	0.06**	0.03	0.08	0.02
G_GDP_2ND(-1)	-0.02	0.05	-0.01	0.00	0.07	0.00
G_GDP_3RD(-1)	-0.09	0.01	-0.01***	0.00	0.01	0.00
G_NUM_BIG_COMPANY(-1)	-0.06	0.04	-0.01	0.08	0.04	0.03*
G_RURAL_INCOME(-1)	0.12	0.08	0.07	0.01	0.10	0.01
G_TOTAL_ELEC_CON	-0.06	0.03	-0.01*	0.03	0.04	0.01
G_SERVICE_CON	0.02	0.00	0.00***	0.00	0.00	0.00
G_INDUSTRY_CON	0.02	0.02	0.00	-0.04	0.03	-0.01
G_TRANSPORT_CON	0.00	0.00	0.00**	0.00	0.00	0.00**
POVERTY * X						
POVERTY * G_GDP_1ST(-1)	-0.15	0.10	-0.11	-0.10	0.12	-0.09
POVERTY * G_GDP_2ND(-1)	0.01	0.07	0.01	-0.08	0.08	-0.05
POVERTY * G_GDP_3RD(-1)	0.11	0.07	0.05*	0.10	0.08	0.06
POVERTY * G_NUM_BIG_COMPANY(-1)	0.04	0.05	0.01	-0.10	0.06	-0.04*
POVERTY * G_RURAL_INCOME(-1)	-0.14	0.09	-0.09	0.05	0.11	0.04
POVERTY * G_TOTAL_ELEC_CON	0.08	0.07	0.04	-0.03	0.08	-0.02
POVERTY * G_SERVICE_CON	-0.02	0.00	0.00***	0.00	0.00	0.00
POVERTY * G_INDUSTRY_CON	-0.03	0.04	-0.01	0.05	0.05	0.02
POVERTY * G_TRANSPORT_CON	0.01	0.00	0.00*	0.01	0.01	0.00
C	0.12	0.01	0.01***	0.12	0.01	0.01***
R ²		0.77			0.58	
F		13.83			5.81	

Note: Variable X_{t-1} was represented as X(-1), variable named with G indicates the growth rate of it, variable named with CON indicates the electricity consumption of it. * $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.

From the perspective of electricity variables, the growth of total electricity consumption has significant backward impact on the growth of GDP in non-poor counties, but it doesn't work in poor counties. Besides, the growth of total electricity consumption and industrial consumption are not significant in model 4. This abnormal phenomenon indicates that the county-level electricity consumption may have large fluctuations among years and the electricity efficiency on economic outcome is relatively low.

Moreover, the growth of electricity consumption of service and transportation industries are significant in both non-poor and poor counties in model 3. In non-poor counties, the electricity consumption on service industry will pull up the growth of local GDP, but in poor counties, its impact has statistically reduced to zero. And that of transportation is going in the opposite direction. Therefore, for poor counties, it's of significance to pay more attention on their transportation.

5.3. Machine Learning Model

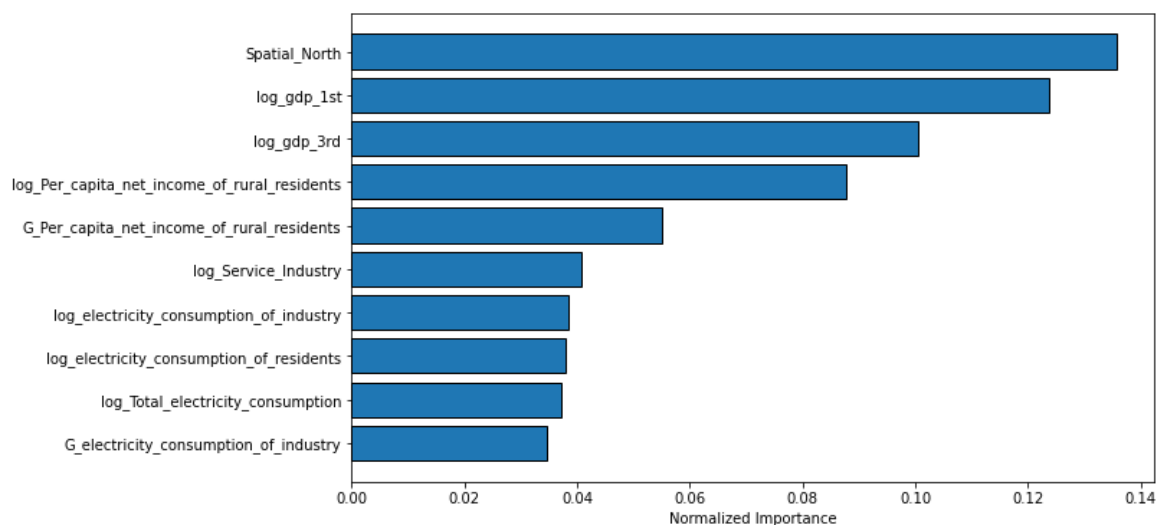
To better fit the variable *POVERTY*, this paper used 30 variables including location, electricity consumption and macroeconomic statistics (each sub-sector variable has its log-transformation and growth-rate data).

And in this case, Z-scaled transformation was applied to all variables. 75% of the data set is randomly divided into training set, five common machine learning methods including XGB Classifier, LGBM Classifier, Random Forest Classifier, Support Vector Classifier and Decision Tree Classifier are used to train the data. The results of well-tuned models are shown in Table 5.

Table 5. Results of Different Machine Learning Models.

Algorithm	Test Accuracy	Mean CV Score
XGB Classifier	0.8636	0.9318
LGBM Classifier	0.8485	0.8915
Random Forest Classifier	0.8409	0.8763
Support Vector Classifier	0.8333	0.8637
Decision Tree Classifier	0.6894	0.8358

Note: Mean CV Score means mean 5-fold CV score of well-tuned model for training set.

**Figure 4.** Feature Importance by XGBoost Classifier.

The result shows that XGB Classifier has the best performance and robustness among 5 methods, so its normalized feature importance is obtained as Figure 4 shown.

According to the feature importance, *Spatial_North* has the largest feature importance of more than 0.13, and *log_gdp_1st*, *log_gdp_3rd* and *log_Per_capita_net_income_of_rural_residents* also have relative larger feature importance of more than 0.08.

6. Conclusions

In this paper, to find out the key factors for under-developed counties' economy, two-way fixed-effects model and machine learning models were established to better fit the data. Compared to linear models, the results have been improved. The map of key factors is drawn as Figure 5 shows.

For poor counties, the increase of electricity consumption of service and transportation industry can drive the growth of local GDP, but they should also try to control the total power consumption, which is to improve the electricity efficiency. And the growth of 3rd industry has the same large positive impact on GDP, which is in line with economic laws. Besides, the growth of 1st industry is still the fundamental industry for county-level economies. It's important to note that the increase in the number of big industrial enterprises cannot effectively bring the expected industrial growth to the local economy, and it may even do the opposite work. Another interesting finding is that electricity resources, known as the blood of industry, are not necessarily correlated with county-level industrial outcome due to the low developed-level of these counties and the policy of overcapacity elimination.

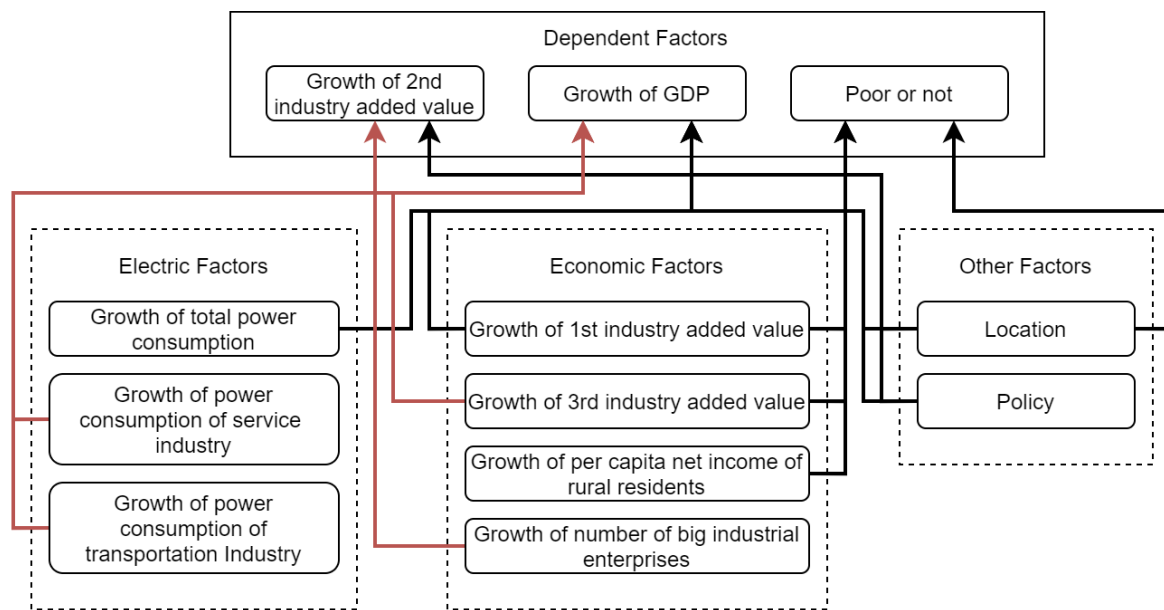


Figure 5. Map of key factors for under-developed counties' economy.

Note: Red line represents that it contains significant effect on poor counties as shown in Figure 5.

Besides, the machine learning results confirm the conclusions of regression, and suggest that the spatial factors are significant to identify a poor county, which can be inferred that there is an obvious cluster of poor counties. The feature importance of XGBoost Classifier shows that the location, development of 1st and 3rd industry and net income of rural residents can explain most difference between the poor and non-poverty counties.

These conclusions can be helpful for the government to lead the under-developed counties to better develop their economy and find the economic driver for themselves. After all, the development strategies of large economies may not be suitable for county-level economies.

References

- [1] Liu X 2016 How the Chinese Fiscal Decentralization Frustrate Its County Economic Development: A Literature Review 2016 *Modern Economy* **7** 541
- [2] Wu L 2017 The Impact of the China's Province-Managing-County Reform on Economic Growth in the Short Term and Long Term *Modern Economy* **8** 1357-65
- [3] Qin Y 2017 'No county left behind?' The distributional impact of high-speed rail upgrades in China *Journal of Economic Geography* **17** 489-520
- [4] Zhang B, Wang X and Ma Y 2019 Study on the Regional Differences of the Influence of County Finance on Economic Development -- Based on Panel Data of 1,732 Counties *International Journal of Education and Economics* **2** 4
- [5] Song X and He Z 2015 The Keqiang index: a new benchmark for China's development *Social Indicators Research* **123** 661-76
- [6] Huang Z, Li Z, Zhang Y and Guo K 2019 Forecasting on Electricity Consumption of Tourism Industry in Changli County *Int. C. Data Serv.* **2019** 77-87
- [7] Shiu A and Lam P L 2004 Electricity consumption and economic growth in China 2004 *Energy Policy* **32** 47-54
- [8] Wolde-Rufael Y 2006 Electricity consumption and economic growth: A time series experience for 17 African countries *Energy Policy* **34** 1106-14
- [9] Hill T D, Davis A P, Roos J M and French M T 2020 Limitations of fixed-effects models for panel data *Sociological Perspectives* **63** 357-69

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.