Individual-Level Factors Affecting the Residents' Feeling of the Degree of Social Justice: Evidence from Chinese Survey Data

Jieteng CHEN Zonglin LI

Abstract

We pay attention to the citizens' feeling about whether this society is justice in China, which is influenced by individual and regional features. The goal of this paper is to quantitatively study and find the potential factors have impacts on social justice. PCA is used to draw information of the individual attributes and perform data reduction. Then we adopt ordered probit model to analyze the effect of extracted principal components on peoples' feeling about the degree of social justice. We find that both better living environment and health status have positive effects on residents' feeling about the level of society justice. And people's attitudes toward politics can indicate their feelings of social justice. Finally, increment of the social status and optimistic expectation for the future social status indicate better feeling of the degree of social justice.

Keywords: Social Justice PCA Ordered Probit

1 Introduction

The rapid development of China's Economy has brought about a great improvement on people's living standards. While the material needs are being satisfied, the mental state and feelings of the people also receive more attention. The residents' feeling of the degree of social justice, which reflects people's feeling about whether this society is justice and fair, is the variable we interest in.

Feeling of justice toward society is an individual emotion, so it is reasonable to think this feeling is influenced by variables of both individual and regional attributes.

Previous studies mainly tried to analyze qualitatively social justice and its influential factors through inducing different theories. We aim to empirically study the influencal factors of individual sense of social justice quantitatively. The data used to empirical exploration is China General Social Survey (CGSS). Because of the high dimensionality of the potential factors in our data, we conduct the principal component analysis (PCA) to extract the features of individuals, which is helpful to obtain information in the original independent variables while decreasing the number of them.

Since the personal society justice is measured by qualitative and ordered discrete choice, we first decide to apply the ordered probit or logit model to analyze the influence of the extracted principal components and regional dummies on individual feelings of the social justice. Considering we may still omit some variables, we finally choose the ordered probit model, which allows for the correlation of error among alternatives.

We find that both better living environment and health status have positive effects on residents' feeling about the level of society justice. And people's attitudes toward politics can indicate their feelings of social justice. Finally, increment of the social status and optimistic expectation for the future social status indicate better feeling of the degree of social justice.

The structure of this paper is as follows: After the introduction and literature review, the third section describes our data used for empirical analysis. The forth and fifth sections show our PCA and ordered probit models and results. The estimation results of the probit model and heterogeneity analysis are provided in section six and seven. In the last section, we discuss the conclusions.

2 Literature Review

It's generally believed that with the widening of the gap between the rich and the poor and the increase in inequity, the sense of inequity in society has risen significantly. The

increment of the feeling of the society injustice will lead to the accumulation of social contradictions and even cause mass incidents, which has seriously threatened China's political and social stability (Yang and Gao, 2016). Therefore, it's crucial to find the variables or factors that significantly influence the citizens' feelings towards the level of equality in the society.

Many existing literatures conduct the analysis to find these factors from some **theories**, such as self-interest theory, which believes that people's feeling of the degree of social justice is determined by how much he gained in the distribution of the society's treasure, and local comparison theory, which believes that the citizens' feeling of the degree of social justice is determined by not only himself's current social status, but also the status change compared with his acquaintance (Wang and Feng, 2013). Meanwhile, there are some study using ordered logistic model to capture the influential factors of people's feeling of the degree of social equality (Wang, 2016; Ma and Jia, 2017). However, we know that The multinomial **ordered logistic** regression model has the independent of irrelevant alternatives property (IIA), which means the probability of one observation belong to class i relative to class j only depends on the decision boundary between class i and class j, and this boundary is not affected by the existence and the properties of other classes. Moreover, Kunimitsu (2015) uses probit to study the regional factors affecting the satisfaction of residents.

Taking into account the methods above, we use PCA and ordered probit to study the individual-level factors affecting the residents' feeling of the degree of social justice. Since the 40 individual-level variables we choose as the independent variables are mostly categorical variables, the estimation of the coefficients may be not stable and the correlations among them are relatively large, so we choose to conduct PCA first. Meanwhile, compared with the probit model, we choose to use the ordered probit model, since the residents' feeling of the degree of social justice has order. What's more, compared with the ordered logit model, we use the ordered probit model which assumes

that the distribution of the error can be correlated between two different categories.

3 Data and Descriptive Statistics

The data used in this article is from the 2012 Chinese General Social Survey project (CGSS). It is aimed to monitor the changing relationship between social structure and quality of life in both urban and rural China.

Table 1 shows the descriptive statistics of the variables (exclude the 31 regional dummies).

Variable	Variable Meaning	Observations	Mean	Std.Dev.	Min	Max
a35	Level of agreement about "I believe the society justice"	11753	3.059	1.073	1	5
a2	Gender (Male = 1)	11765	1.488	0.500	1	2
s5a	Living location	11765	3.034	1.783	1	6
s5b	Neighborhood condition	11765	4.780	2.328	1	8
s6	Living type (Family = 1)	11603	1.013	0.114	1	2
a4	Ethnic	11756	1.432	1.527	1	8
a501	Religion	11763	0.853	0.354	0	1
a7a	Education level	11763	4.852	3.018	1	14
a8a	Individual yearly income	10633	0.000	1.000	-0.163	70.930
a10	Political status	11736	3.599	0.978	1	4
a15	Physical health condition	11761	3.542	1.086	1	5
a17	Mental health condition	11748	3.841	1.010	1	5
a18	Type of Hukou	11765	1.733	1.094	1	8
a29	Major source of information	11732	4.106	0.846	1	7
a43ab	Change of the social status in the last ten years	10304	1.164	1.199	0	12
a43c	Expected social status after ten years	11546	5.243	2.115	1	10
a44	Have voted?	11754	1.555	0.569	1	3
a45	Union member (Yes = 1)	11723	2.666	0.667	1	3
a46	Level of agreement about "Gov should not interfere when people criticize the Gov in public places"	11688	2.831	1.132	1	5

Variable	Variable Meaning	Observations	Mean	Std.Dev.	Min	Max
a58	Farming experience	11762	2.835	1.747	1	6
a611	Enjoy medical insurance	11759	1.116	0.339	1	3
a62	Yearly income for the family	10392	0.000	1.000	-0.192	33.211
a64	Economic status of the family	11718	2.624	0.731	1	5
a69	Marital status	11765	3.269	1.401	1	7
a89b	Father's highest education level	11661	3.253	2.936	1	14
a90b	Mather's highest education level	11705	2.392	2.436	1	14
n3c	Social status compared with others	5625	1.923	0.319	1	3
n1501	Level of agreement about "I won't affect Gov's decisions"	5808	2.977	1.572	1	7
n1503	Level of agreement about "I am interested in Politics"	5807	4.306	1.763	1	7
n1504	Level of agreement about "I want to make contributions to the society"	5792	2.652	1.189	1	7
n1603	Level of trust on neighbors	5810	1.952	0.589	1	4
n1604	Level of trust on co-workers	5534	2.088	0.581	1	4
n1605	Level of trust on strangers	5813	3.293	0.601	1	4
n1606	Level of trust on doctors	5802	2.008	0.654	1	4
n1608	Level of trust on company executives	5416	2.472	0.668	1	4
n1609	Level of trust on journalists	5578	2.321	0.718	1	4
n1611	Level of trust on teachers	5778	1.796	0.614	1	4
n1612	Level of trust on local officials	5735	2.425	0.791	1	4
n1613	Level of trust on central officials	5697	1.998	0.775	1	4
n1614	Level of trust on policemen	5735	2.077	0.706	1	4
n1616	Level of trust on judges	5625	2.087	0.692	1	4

Table 1: Statistical Description I

As shown in **Table** 1, we find that the number of observations is 11765, while the missing-value problem mostly exists in questionnaire N, which is caused by random sampling, that is, the respondents who are asked to answers questionnaire N is randomly selected. Therefore, there does not exist sample selection bias due to the missing observations in questionnaire N. Moreover, we can see that most of our variables are

categorical variables including the dependent variable "$a35$", so the final regression model we choose should not be the simple linear regression and we may need to perform data reduction.

After deleting the missing values, we get 3598 obseravations in total, and the distribution of the dependent variable a35 is displayed in **Tabel** 2:

a35 (the degree of social justice)	Freq.	Percent	Cum.
Totally unfair	246	6.84	6.84
Relatively unfair	934	25.96	32.80
I cannot make specific comments	737	20.48	53.28
Relatively fair	1529	42.50	95.78
Totally fair	152	4.22	100.00
Total	3598	100.00	

Table 2: Statistical Description of the Dependent Variable

We can conclude that the most respondents believe the society is relatively fair (42.5%), while there still exist 32% of the respondents believe the society is not fair at some level.

4 Principal Component Analysis (PCA)

Since the 40 individual-level variables we have are mostly categorical variables, it may be problematic if we directly conduct the probit regression on them. Moreover, some of the variables we have may share similar information (variation), the correlations among them are relatively large, therefore we cannot directly perform regression.

Thus, we first use 40 individual-level variables to conduct principal component analysis, through which we could get principal components to further perform probit regression together with the regional dummies.

4.1 Introduction of PCA

Principal component analysis is concerned with explaining the *variance- covariance structure* of a set of variables through a few linear combinations of the original variables. Its general objectives are data reduction and interpretation.

4.2 Results of PCA

As explained above, PCA is used to convert variables in the data set into several linearly uncorrelated variables called **principal components**. First, we need to conduct **parallel analysis** to make sure the number of principal components we need.

4.2.1 Parallel Analysis

Parallel analysis is a statistical method used to determine the number of principal components to keep.

The scree plot is a kind of visualization of the parallel analysi. It is a line plot of the eigenvalues of the principal components in PCA. According to this plot, we can find the "elbow" and the components to the left of this "elbow" should be remained for further analysis, replacing the original independent variables.

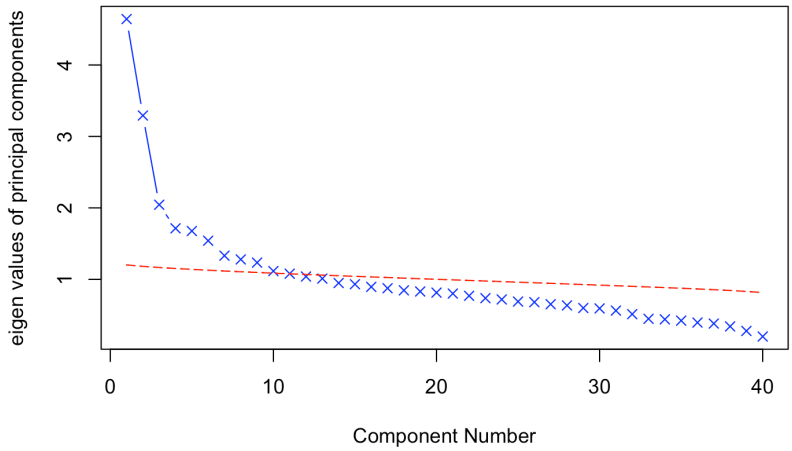

Figure 1: the Scree Plot

After conduct the parallel analysis, the above **Figure** 1 suggests that the number of components is 10. Below are eigenvalues and the proportion of total and cumulative variance explained by these ten PCs shown in Table 3:

	RC1	**RC2**	**RC3**	**RC4**	**RC5**	**RC6**	**RC7**	**RC8**	**RC9**	**RC10**
Eigenvalues	3.79	2.82	1.93	1.87	1.73	1.73	1.73	1.70	1.31	1.28
Proportion Variance	0.09	0.07	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03
Cumulative Variance	0.09	0.17	0.21	0.26	0.30	0.35	0.39	0.43	0.46	0.50
Proportion Explained	0.19	0.14	0.10	0.09	0.09	0.09	0.09	0.09	0.07	0.06
Cumulative Proportion Explained	0.19	0.33	0.43	0.52	0.61	0.70	0.78	0.87	0.94	1.00

Table 3: Eigenvalues and the proportion of total and cumulative variance explained by PCs

The first principal component accounted for the largest proportion (19%) of the total variation in the original 40 variables, followed by the second principal component accounted for the largest proportion (14%) of the total variation in the original 40 variables. Similarly, 10% of the total variation explained by the third component, 9% of the total variation explained by the forth to the eighth component. The last two components explained 7% and 6% percent of the variation in the original independent variables.

4.2.2 Principal Components and Scores

After obtaining the optimal number of principal components, we can further get the exact principal components and scores to conduct the further probit regression.

The loading coefficients are listed in the table below:

Variable	RC1	RC2	RC3	RC4	RC5	RC6	RC7	RC8	RC9	RC10
a2	-0.060	-0.124	0.073	-0.216	0.489	-0.053	0.059	0.192	0.158	0.343
s5a	-0.076	**0.868**	-0.105	-0.077	-0.024	-0.025	-0.061	-0.039	0.045	0.095
s5b	-0.048	**0.816**	-0.043	-0.032	-0.034	0.004	-0.069	-0.011	-0.043	0.140
s6	0.033	-0.014	0.002	0.003	0.004	-0.019	0.008	-0.041	0.033	-0.374
a4	-0.012	0.119	0.066	0.072	-0.035	-0.018	-0.023	0.018	**0.636**	-0.046
a501	-0.029	0.054	0.055	0.086	-0.058	-0.003	-0.017	-0.117	**-0.638**	0.121
a7a	0.140	-0.483	0.448	0.261	-0.250	0.083	0.036	0.103	-0.181	-0.100
a8a	0.012	-0.035	-0.002	0.039	0.000	**0.920**	0.011	0.008	-0.007	-0.014
a10	-0.030	0.235	-0.059	-0.073	**0.576**	-0.004	-0.026	-0.005	0.136	-0.144
a15	0.040	-0.087	0.081	**0.751**	0.030	0.008	-0.009	0.125	-0.008	-0.173
a17	-0.016	-0.143	-0.040	**0.766**	0.005	0.003	-0.072	-0.003	0.035	0.137
a18	0.111	-0.657	0.035	0.028	-0.056	0.064	0.002	-0.071	-0.015	0.118
a29	-0.016	0.147	0.345	-0.012	0.160	0.027	0.036	0.098	-0.097	-0.208
a43ab	-0.044	0.161	-0.067	-0.047	-0.030	-0.031	0.097	**0.747**	0.080	-0.029
a43c	-0.077	-0.024	0.105	0.183	-0.042	0.027	0.031	**0.806**	0.060	0.016
a44	0.068	-0.290	0.182	0.108	0.295	0.007	0.076	0.036	-0.066	-0.244
a45	-0.043	0.522	-0.043	-0.023	0.322	-0.006	0.007	0.086	0.165	-0.152
a46	0.208	-0.073	-0.010	0.059	0.137	-0.012	-0.001	0.055	-0.354	-0.189
a58	-0.060	-0.022	-0.148	-0.296	0.092	-0.054	-0.007	-0.072	0.159	**0.520**
a611	-0.069	0.143	-0.171	0.064	0.068	-0.050	0.132	-0.274	0.256	0.074
a62	0.023	-0.068	0.077	-0.003	0.000	**0.917**	-0.030	0.037	-0.017	-0.008
a64	0.048	-0.073	0.060	0.385	-0.157	0.106	-0.063	0.455	-0.033	0.266
a69	-0.017	-0.078	-0.281	-0.368	0.149	-0.006	-0.102	-0.041	0.054	0.315
a89b	0.067	-0.186	**0.792**	0.043	-0.040	0.027	0.009	0.004	0.069	0.072
a90b	0.062	-0.149	**0.825**	0.055	-0.021	0.014	-0.003	0.004	0.036	0.044
n3c	0.014	0.054	0.103	0.206	-0.039	-0.023	0.020	0.007	-0.161	0.388
n1501	-0.041	-0.022	0.074	0.169	-0.387	-0.052	0.004	0.166	0.197	0.165
n1503	0.083	-0.012	0.013	0.107	**0.630**	0.003	0.067	-0.057	-0.033	0.064
n1504	0.157	0.097	-0.002	0.012	0.447	-0.024	0.110	-0.159	-0.154	0.251
n1603	0.224	-0.131	0.116	0.000	0.013	0.012	**0.729**	0.052	-0.163	-0.051
n1604	0.244	0.082	0.020	-0.086	0.063	-0.010	**0.753**	-0.010	-0.051	0.028
n1605	0.016	-0.091	-0.050	0.012	0.073	-0.023	0.584	0.044	0.151	-0.032

Variable	RC1	RC2	RC3	RC4	RC5	RC6	RC7	RC8	RC9	RC10
n1606	0.577	-0.060	0.008	0.027	-0.076	-0.018	0.117	-0.028	0.019	-0.046
n1608	0.498	0.053	-0.057	-0.029	-0.040	0.041	0.334	-0.088	0.136	0.112
n1609	0.539	-0.089	0.112	0.121	0.047	0.022	0.256	-0.053	-0.019	0.088
n1611	0.618	-0.118	0.049	0.015	0.021	0.002	0.048	-0.030	-0.046	0.088
n1612	**0.701**	-0.008	0.020	-0.065	0.071	0.009	0.056	0.051	-0.077	-0.177
n1613	**0.725**	-0.074	0.097	0.025	0.120	0.003	-0.022	0.053	-0.123	-0.046
n1614	**0.788**	-0.023	-0.007	-0.002	0.030	-0.008	0.001	-0.021	-0.026	-0.066
n1616	**0.792**	-0.015	-0.023	-0.007	0.022	0.027	0.006	-0.015	0.016	-0.028

Table 4: the Loadings of the Principal Components

From **Table** 4, we can get the variables that make more contribution to the corresponding principal components, which have been highlighted in bold.

Combining with the meanings of the variables, we can get the meanings of these ten components. The first component "$RC1$" has higher loadings on variables "$n1612$", "$n1613$", "$n1614$" and "$n1616$", which are questions about the respondents' attitude towards officials and policemen, thus we can inteprated this component as ***the level of trust towards government officials***. The second component "$RC2$" has higher loadings on variables "$s5a$" and "$s5b$", which are questions about the respondents' ***living environment***. The variables "$a89b$" and "$a90b$", which are questions about the respondents' ***parents' education level*** are the higher loadings on the third component "$RC3$". The forth component "$RC4$" has higher loadings on variables "$a15$" and "$a17$", which are questions about the respondents' ***mental and physical health condition***. Similarly, the variables "$a10$" and "$n1503$" have the higher loadings on the fifth component "$RC5$", which means "$RC5$" can be interprated as the respondents' ***degree of political participation***. Moreover, the sixth component "$RC6$" has higher loadings on variables "$a8a$" and "$a62$", which are questions about the respondents' yearly income, thus we can inteprated this component as ***the income level of the respondent***. The seventh component "$RC7$" means the ***level of trust towards the acquaintance***, since the vairable "$n1603$"

and "$n1604$" are questions about the respondents' attitude towards neighbors and co-workers. The variables "$a43ab$" and "$a43c$" have the higher loadings on the eighth component "$RC8$", which means "$RC8$" can be interprated as the respondents' ***change and expectation of their social status***. The ninth component "$RC9$" has higher loadings on variables "$a4$" and "$a501$", which are questions about the respondents' ***ethnic and religion***. The variable "$a58$" has higher loading on the tenth component "$RC10$", which means "$RC10$" can be interprated as whether the respondents have ***farming experience***.

5 Ordered Probit

5.1 Why use Probit

Our target variable is an ordinal statistic, so traditional regression methods toward continuous dependent variable are not suitable. We adapt probit regression model to deal with this problem.

Probit models allow for correlated errors among different observations. Considering our model can not get all the explanatory variables and some relevant variables still may be omitted, or even some explanatory variables are unobservable, these omitted variables may not follow indenpendent distribution. In other words, the error terms containing these omitted variables can be correlated over alternatives. The assumption that the error terms are independent of each other is called *Independence of Irrelevant Alternatives*, which is hold and required on logistic model. However, we can't assert the error terms are independent and has *IIA* property. Fortunately, probit models tolerate the correlated errors. So we decide to use porbit models here.

5.2 Ordered Probit Model

The *Ordered Probit Model* is a kind of generalization of the probit regression model, which has a wide application in classification situation.

Considering the target variable, (Social justice), has a natural ordering. (1 means respondents think this society is very unfair and 5 represents respondents believe the society is very fair), the ordered probit model can provide an appropriate fit of our data. It makes no assumption of the interval distance between different categories while holding the order of dependent variable.

After getting the scores for each component, we conduct the ordered probit regression with the province-level regional dummies. the ordered probit model is following.

$$y^{\star} = \mathbf{RC^T}\beta + \lambda_{\mathbf{i}} + \epsilon$$

RC is the vector of the ten components, β is the vector of regression coefficients we want to estimate, λ_i is the province fixed effect that needs to be controlled, and ϵ is the error term. Most importantly, $y^{\star}$ is unobservable, we can only know when it crosses thresholds:

$$y_i = \begin{cases} 1 & \text{if } y_i^{\star} \leq \alpha_1 \\ 2 & \text{if } \alpha_1 < y_i^{\star} \leq \alpha_2 \\ 3 & \text{if } \alpha_2 < y_i^{\star} \leq \alpha_3 \\ 4 & \text{if } \alpha_3 < y_i^{\star} \leq \alpha_4 \\ 5 & \text{if } y_i^{\star} > \alpha_4 \end{cases}$$

α's are unknown thresholds parameters that should be estimated along with β and λ.

The probability that observation i will locate in alternative j is:

$$P_{ij} = P(y_i = j) = P(\alpha_{j-1} < y_i^{\star} \leq \alpha_j) = F(\alpha_j - y_i^{\star}) - F(\alpha_{j-1} - y_i^{\star})$$

For ordered probit model, F is the standard normal CDF(cumulative distribution function). The exact estimation process is realized by maximum likelihood, which is completed in *STATA* software.

6 Empirical Results

6.1 Estimation of the ordered probit model of individual society justice

Table 5 shows the estimation result of the ordered probit model. Although the pseudo R-squared value of the discrete outcome data was a little low, other statistics such log likelihood test Chi2 and t-statistic shows that the explanatory power of the ordered probit model is still acceptable.

Variable	**Coeff.**	**Std.Err** .	**t-statistic**	Remarks
RC1	-.293	.019	-15.13***	No of observations: 3596
RC2	.121	.0212	5.74***	Pseudo R-squared: 0.066
RC3	-.071	.018	-3.84 ***	LR Chi2(38):640.80
RC4	.015	.018	0.83 **	Prob >Chi2 =0.0000
RC5	-.091	.018	-4.97 ***	Log likelihood:-4555.477
RC6	-.155	.018	-8.44***	
RC7	-.017	.017	-1.00 **	
RC8	.067	.018	3.67 ***	
RC9	.106	.021	5.07 ***	
RC10	.139	.019	7.44***	
α_1	-1.769	.084		
α_2	-.619	.079		
α_3	-.035	.079		
α_4	1.787	.086		
Significance level: *10% **5% ***1%				
α shows the threshold values of classifications				

Table 5: Estimation of the ordered probit model

Most principal components, such as the level of trust to government officials and the satisfication of living environment and mental and physical health condition, have significant effects on the personal society justice level.

It is clear that "RC4", mental and physical health condition, is significantly positive.

It is reasonable to introduce that a good health status can help improve people's justice toward the society. Furthermore, we believe a good and satisfactory living surroundings also has a positive impact on the feeling of justice toward the society. The estimated coefficient of "RC2" is significantly positive, which confirms our conclusion to some extent.

What's more, people's attitude toward political things also can indicate their feelings of social justice. According to the estimated results, "RC1" and "RC5" both have the significantly negative coefficients. We attribute these results to the reasons that the distrust of government officials can lead to a reduction in social justice of people and low political participation always indicates unfair attitudes towards society.

Another interesting finding is that "RC8", change and expectation of the social status, also has significantly positive effect on the level of justice toward society. We think this result confirms with the reason that people who can change the social class in which they live and have good expectations for the future will more believe that this society is fair.

Even though some variables are omitted or unobservable, our selected principal components can reveal the personal society justice well.

6.2 Comparison between Probit and Logit

The logistic regression is similar with probit model, except that the assumptions of the distribution of the error are different. Logit model is also widely used in categorical outcomes classification problems, so we adapt ordered to test the robustness of our model and results.

The logistic regression is similar with probit model, except that the assumptions of the distribution of the error are different. As we have mentioned in Sec 5, equation 3 shows the F for probit is the standard normal CDF, but for logit, F is the logistic CDF $F(z) = e^z/(1 + e^z)$. The multinomial logistic regression model has the independent of

irrelevant alternatives property(IIA), which means the probability of one observation belong to class i relative to class j only depends on the decision boundary between class i and class j, and this decision boundary is not affected by the existence and the properties of other classes.

Variable	**Coeff.**	**Std.Err.**	**t-statistic**	Remarks
RC1	-.517	.035	-14.84***	No of observations: 3596
RC2	.225	.037	6.13***	Pseudo R-squared: 0.066
RC3	-.129	.032	-4.03***	LR Chi2(38):642.71
RC4	.033	.0327	1.01**	Prob >Chi2 =0.0000
RC5	-.159	.032	-4.99***	Log likelihood:-4554.5237
RC6	-.279	.032	-8.58***	
RC7	-.031	.027	-1.13	
RC8	.127	.032	3.90***	
RC9	.178	.037	4.85***	
RC10	.228	.033	7.00***	
α_1	-3.126	.151		
α_2	-1.052	.138		
α_3	-.086	.137		
α_4	3.197	.157		
Significance level: *10% **5% ***1%				
α shows the threshold values of classifications				

Table 6: Estimation of the ordered logit model

In our model, we have assumed that there exists some unobservable or unattainable variables that can affects people's social justice and IIA may not holds. In other words, we suppose that the errors are not independent of each other, the error for one alternative provides information about the error for another alternative, which violates the IIA assumption. That is also the reason why we adapt probit models to derive our main conclusions.

Although our assumption can not satisfy the IIA of multinomial logistic regression perfectly. We want to test whether our conclusions still holds under the correlated erroe

assumption. Logit model is also widely used in categorical outcomes classification problems, so we adapt ordered logistic regression to test the robustness of our model and results.

Based on the logistic regression results, we find that our conclusions still hold. "RC2" living environment and "RC4" health condition can still have positive effects on personal social justice. "RC1" and "RC5", which presents people's attitudes toward political issues and government officials, can reveal their feelings of society fairness. In summary , logistic regression results also support our conclusions.

7 Heterogeneity Analysis

7.1 Estimation of the ordered probit model on different genders

In this subsection, we divide the dataset into two parts according to the respondents' gender. After building our model on two sub dataset, we explore the heterogeneity among the whole population and draw some conclusion.

Variable	**Coeff.**	**Std.Err** .	**t-statistic**	Remarks
RC1	-.341	.028	-12.25***	No of observations: 1882
RC2	.084	.030	2.83 ***	Pseudo R-squared: 0.078
RC3	-.047	.027	-1.76*	LR Chi2(38):393.74
RC4	-.012	.029	-0.40	Prob >Chi2 =0.000
RC5	-.080	.031	-2.61***	Log likelihood:-2316.104
RC6	-.120	.027	-4.45***	
RC7	-.020	.018	-1.11	
RC8	.092	.026	3.49 ***	
RC9	.152	.031	4.89***	
RC10	.151	.029	5.17***	
α_1	-1.903	.125		
α_2	-.752	.119		
α_3	-.204	.118		

Variable	Coeff.	Std.Err .	t-statistic	Remarks
α_4	1.785	.129		
Significance level: *10% **5% ***1%				
α shows the threshold values of classifications				

Table 7: Estimation of the ordered probit model on Male

Among the male group as **Table** 7, we find that most of our conclusion still hold. "RC2", which parents the living environment, still indicates a better society justice. However, "RC4", mental and physical health condition seems to have on impact on the male's attitude of society justice.

Variable	Coeff.	Std.Err.	t-statistic	Remarks
RC1	-.252	.028	-9.13***	No of observations: 1714
RC2	.164	.031	5.22 ***	Pseudo R-squared: 0.065
RC3	-.109	.027	-3.99***	LR Chi2(38):307.47
RC4	.025	.027	0.94	Prob >Chi2 =0.0000
RC5	-.107	.033	-3.22***	Log likelihood:-2202.379
RC6	-.198	.026	-7.65***	
RC7	.049	.141	0.35	
RC8	.049	.026	1.72*	
RC9	.064	.030	2.10**	
RC10	.131	.030	4.41***	
α_1	-1.763	.120		
α_2	-.588	.113		
α_3	.048	.113		
α_4	1.745	.121		
Significance level: *10% **5% ***1%				
α shows the threshold values of classifications				

Table 8: Estimation of the ordered probit model on Female

As for female as shown in **Table** 8, we find that the quality of living surroundings can still reveal people's social justice. And their social justice is affected by their trust on government officials and the their political participation. However, it seems that we

can't assert that the change and expectation of social class of female can improve their society justice postively.

7.2 Estimation of the ordered probit model of different religion

Based on whether the respondent believes in religion, we treat them as two different groups. Similar to our method in above subsection, what we intend to do is to search the differences amd heterogeneity between different religional groups of people

Table 9 below is the estimation results of the model on people with religion belief. Most of the result is in conformity with our conclusion, except for the "RC4", mental and physical health status.

Variable	**Coeff.**	**Std.Err.**	**t-statistic**	Remarks
RC1	-.286	.021	-13.71***	No of observations: 3132
RC2	.131	.023	5.68***	Pseudo R-squared: 0.066
RC3	-.070	.020	-3.46***	LR Chi2(38):551.27
RC4	.014	.021	0.70	Prob >Chi2 =0.0000
RC5	-.085	.020	-4.36***	Log likelihood:-3931.577
RC6	-.169	.020	-8.45***	
RC7	-.022	.019	-1.25	
RC8	.072	.020	3.57***	
RC9	.141	.029	4.85***	
RC10	.135	.020	6.65***	
α_1	-1.780	.089		
α_2	-.615	.084		
α_3	-.023	.084		
α_4	1.848	.092		
Significance level: *10% **5% ***1%				
α shows the threshold values of classifications				

Table 9: Estimation of the ordered probit model on people with religion

Table 10 show the estimated model on the people without religion belief. For someone without religion, "RC8", the change and expectation of the social status doesn't

have a significantly positive impact on people's social justice.

Variable	**Coeff.**	**Std.Err** .	**t-statistic**	Remarks
RC1	-.334	.054	-6.11 ***	No of observations: 464
RC2	.109	.060	1.83*	Pseudo R-squared: 0.098
RC3	-.122	.057	-2.15**	LR Chi2(38):129.49
RC4	-.002	.050	-0.03	Prob >Chi2 =0.0000
RC5	-.164	.057	-2.87***	Log likelihood:-597.664
RC6	-.093	.051	-1.81*	
RC7	.148	.274	0.54	
RC8	.075	.053	1.41	
RC9	.154	.075	2.04**	
RC10	.140	.053	2.63***	
α_1	-1.811	.286		
α_2	-.693	.277		
α_3	-.137	.276		
α_4	1.511	.286		
Significance level: *10% **5% ***1%				
α shows the threshold values of classifications				

Table 10: Estimation of the ordered probit model on people without religion

There are various religious beliefs in China, and different religions always have different interpretations of life meaning and social equality. For example, some religions advocate that all lives are equal, and some religions advocate the existence of the afterlife. So religion often affects people's mental status and health and attitude towards life. Thus this should be the reason for the heterogeneity here.

7.3 Estimation of the ordered probit model of different marital status

In this subsection, we split the dataset into two parts according to the respondent's marital status.

As shown in Table 11, we find that for unmarried people, the relationship between health status and their social justice is not very obvious. What's more, the change and

Variable	**Coeff.**	**Std.Err** .	**t-statistic**	Remarks
RC1	-.269	.046	-5.87***	No of observations: 652
RC2	.183	.055	3.32***	Pseudo R-squared: 0.107
RC3	-.089	.038	-2.32**	LR Chi2(38):188.19
RC4	-.072	.042	-1.72	Prob >Chi2 =0.0000
RC5	-.181	.050	-3.60***	Log likelihood:-788.001
RC6	-.244	.045	-5.39***	
RC7	.166	.212	0.78	
RC8	.001	.043	0.01	
RC9	.118	.049	2.41**	
RC10	.089	.039	2.31**	
α_1	-1.495	.251		
α_2	-.352	.244		
α_3	.271	.244		
α_4	2.304	.264		
Significance level: *10% **5% ***1%				
α shows the threshold values of classifications				

Table 11: Estimation of the ordered probit model on unmarried people

expectation of social status seems to have no effect on people's feeling of justice toward society.

Variable	**Coeff.**	**Std.Err** .	**t-statistic**	Remarks
RC1	-.300	.021	-13.94***	No of observations: 2944
RC2	.111	.023	4.76***	Pseudo R-squared: 0.063
RC3	-.061	.023	-2.69***	LR Chi2(38):502.88
RC4	.051	.023	2.25**	Prob >Chi2 =0.0000
RC5	-.086	.020	-4.25***	Log likelihood:-3741.7227
RC6	-.136	.021	-6.61***	
RC7	-.017	.018	-0.95	
RC8	.088	.021	4.22***	
RC9	.104	.024	4.43***	
RC10	.149	.023	6.61***	

Variable	Coeff.	Std.Err .	t-statistic	Remarks
α_1	-1.826	.090		
α_2	-.664	.084		
α_3	-.083	.084		
α_4	1.717	.091		
Significance level: *10% **5% ***1%				
α shows the threshold values of classifications				

Table 12: Estimation of the ordered probit model on married people

As for the married respondents, the pattern is consistent with our main probit model and conclusions, as shown in Table 12.

8 Conclusions

We use PCA and ordered probit to study the individual-level factors affecting the residents' feeling of the degree of social justice. Since the 40 individual-level variables we choose as the independent variables are mostly categorical variables, the estimation of the coefficients may be not stable and the correlations among them are relatively large, thus we choose to conduct PCA first to extract information in the original independent variables while decreasing the number of them. Then, we choose to use the ordered probit model, since the residents' feeling of the degree of social justice has order and the probit model assumes that the distribution of the error can be correlated between two different categories, which prevent IIA compared with logit.

We find that both better living environment and health status have positive effects on residents' feeling about the level of society justice. And people's attitudes toward politics can indicate their feelings of social justice. Finally, increment of the social status and optimistic expectation for the future social status indicate better feeling of the degree of social justice.

References

[1] Stevens, J. and Sobol, M. G. (1984) Applied Multivariate Statistical Analysis, *Interfaces*, 14(5), pp. 116118.

[2] The definition of parallel analysis.

https : *//en.wikipedia.org/wiki/Parallel_analysis*

[3] Jiaming MAO. "Model Selection and Regularization"

https : *//github.com/jiamingmao/data − analysis/blob/master/Lectures*

[4] Olasege, B. S. et al. (2019) Genetic parameter estimates for body conformation traits using composite index, principal component, and factor analysis, *Journal of Dairy Science*, 102(6), pp. 52195229. doi: 10.3168/jds. 2018-15561.

[5] Yoji Kunimitsu (2015) Regional Factors Affecting the Satisfaction of Residents: Evidence from a Structural Equation Model in Japan, *Social Indicators Research*, 120(2), p. 483.

[6] 麻宝斌, 贾茹. 当代中国社会公平感代际差异及影响因素 [J]. 公共行政评论, 2017, 10(04): 95-114+194-195.

[7] 王甫勤. 新生代与传统农民工社会公平感的影响因素研究[J]. 中国人口科学, 2016(05): 110-119+128.

[8] 麻宝斌, 杜平. 结构分化、观念差异与生活经历——转型时期社会公平感的影响因素分析[J]. 江汉论坛, 2017(03): 119-124.

[9] William E. Becker, Peter E. Kennedy(1992) 'A Graphical Exposition of the Ordered Probit', *Econometric Theory*, Vol.8, No.1, pp. 127-131.

[10] Jiaming MAO. "Classification and Discrete Choice Models"
https : *//github.com/jiamingmao/data − analysis/blob/master/Lectures*

Appendix

1 The Questionnaire provided by the CGSS official & The original CGSS data

2 The dataset containing the selected potential variables & The corresponding STATA dofile to process the data

3 The Rmarkdown file to realize PCA

4 The CSV file of the principal components

5 The STATA dofile to realize ordered probit model

6 The merged data including principal component and respondents'ID with other information from Data_3.0

All the above supporting materials can be accessed by this link: github repositories