Research on food redistribution model based on principal component analysis and factor analysis

Yuhan Zhu¹, *, Yihuan Huang², Yuhang Yan³
¹Capital University of Economics and Business, Beijing, 100000, China
²School of Economics and Management, Beijing University of Chemical Technology, Beijing, 100029, China
³School of Artificial Intelligence, Beijing Technology and Business University, Beijing, 102488, China

*Corresponding author: yuhanzhu2020@cueb.edu.cn

Abstract. We have divided rebuilding food systems into two phases, one of which is to maximize productivity in a way that maximizes sustainability and economic benefits. Another is to minimize hunger and food insecurity, with regard to equity. Firstly, we use Euclidean Metric and Complete Connection Method for cluster analysis according to the production efficiency and production structure of various foods. Then the countries are classified into countries with different production structures and productivity. Secondly, we quantified sustainability indicators to make standard deviations and risks to food production. Using models, principal component analysis, and factor analysis, we figure out how to distribute food among countries to maximize food productivity worldwide. Thirdly, the model is used to analyze the coefficient of food insecurity and the degree of hunger in each country, so as to obtain the utility level of food distribution in each country. The level of food utility in developed countries is relatively flat, while that in underdeveloped countries is steeper. So, the utility of food distribution in less developed countries is higher than that in developed countries. Therefore, when we measure the total utility of food, we multiply the utility difference of the insecurity factor by the quantity of food mixture and add the original food utility level to maximize the food utility level.

Keywords: food redistribution, principal component analysis, factor analysis.

1. Background

Equity and efficiency have always been two indicators of economics. The efficiency of food production is bound to be reduced in order to equalize the distribution of goods and reduce hunger. If we do it purely on the basis of economic efficiency, you're going to have mass starvation. The food system is crucial to the survival of the world's population. The global food system is in a state of instability due to factors such as efficiency still falling short of market demands and environmental damage to the food system.

In 2019, 9.7 percent of the world's population faced severe food insecurity. This means that the amount of food consumed is reduced to the point where starvation is possible. A further 16 percent of the world's population suffers from moderate levels of food insecurity, which means they do not have...
regular access to nutrition and adequate food. Combined moderate and severe food insecurity is estimated to total 25.9 percent of the world's population, or about 2 billion people. Even in food-rich countries, such as food-exporting countries, people still suffer from hunger. This fact shows that the current food system does not meet the needs of a large segment of society. The circulation and security of food has always been based on price fluctuations and profitability. In a perfect market with no trade barriers, and with food flowing completely freely between countries, the food system must be distributed in a way that most closely follows the laws of the market and is most efficient with the intervention of Adam Smith's invisible hand. However, as the basic guarantee of life and the basic condition of national stability, the market efficiency should not be taken as the only evaluation direction, but also the factor of fairness should be considered. Further improvement of the food system to meet the needs of the poorest is therefore crucial.

The current food system has a big impact on the environment.[1] For example, food production emits nearly a quarter of the world's greenhouse gases, uses nearly two-thirds of the world's fresh water, reduces forests and destroys biodiversity. Besides, the use of chemical fertilizers and pesticides causes nitrogen and phosphorus pollution.[2] Using models and data from the global food system,[3] the team predicts that by 2050, in the absence of technological change and other mitigation measures, the stress on the environment caused by the food system will increase by 50 to 92 percent on each indicator.

Therefore, re-optimizing the food system to improve equity, sustainability, and efficiency is of maximum utility to people's survival and the smooth operation of society.

2. Efficiency regression model
The 10 sets of data from http://www.fao.org/, represent efficiency indicators for Brazil, China, Europe, India, the Republic of Korea, South Africa, Thailand, Turkey, Ukraine and the United States of America. [4] In order to make the categories of models more accurate. This paper uses efficiency data to aggregate countries. To some extent, differences in national population and land size do not affect country categories. On the other hand, the amount of food used in different countries will make us clearer.

First, the data are normalized to obtain variable correlations as shown in Figure 1. As can be seen from the scatter plot, Y is positively correlated with x1, x2, x3, x5 and x6, x1 is highly correlated with x2, x1 is highly correlated with x3, and x1 is highly correlated with x5, x1 is negatively correlated with x4, x2 is negatively correlated with x4, x3 is negatively correlated with x4, x5 is negatively correlated with x4, and Y is negatively correlated with x4.

![Figure 1. scatter plot of variable correlation](image-url)
The difference in heteroscedastic diagnosis of that result is shown in fig. 2. The residual error changes randomly at \( E = 0 \), which is in the region with small change range, indicating that there is no heteroscedasticity in this model. Under the assumption of homogeneity, the tested \( P \) value is 0.83813. Therefore, we do not reject the null hypothesis and believe that there is no heteroscedasticity. [5] At the same time, we can also test the autocorrelation of the diagnostic sequence by DW. \( DW=3.346 \), \( P=0.05707 \), \( P=0.05707 < 0.1 \), when the significance level is 0.05, it is considered that there is no autocorrelation or positive sequence correlation between the random error terms.

In this paper, the multiple linear regression equation obtained by R language is as follows:
\[
y = 2.036E-16 + 2.589 x_1 - 1.086 x_2 - 1.336 x_5
\]

Next, principal component analysis is used to find the main factors. The sustainability standard deviation is used to quantify the risks arising from environmental effects, rather than the risks in the Lahey model to quantify the likelihood of environmental effects.

\[
\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}
\]

\[
Prob(\text{risk}) = \frac{\exp(\text{risk})}{1 + \exp(\text{risk})}
\]

Finally, probability production and coefficients are used to fit the index of R language as shown in Figure 3.
Based on the principle that the cumulative proportion is greater than or equal to 80%, two principal components $z_1$ and $z_2$ are selected to explain 88.77% of $Y$.

Regression analysis of principal components showed that:

$$\hat{y} = -9.338\times10^{-17} + 0.2438z_1 + 0.08605z_2$$

Thus, the principal component regression equation of $Y$ with respect to the four original independent variables $x_1, x_2, x_3$ and $x_4$ can be obtained as follows:

$$\hat{y} = 0.799\times10^{-17} + 0.1446x_1 + 0.1445x_2 + 0.1366x_3 + 0.03463x_4 - 0.1285933x_5$$

Check its heteroscedasticity and autocorrelation as shown in Figure 4:

As can be seen from the figure, the residual series basically fluctuates randomly around 0, without obvious heteroscedasticity and autocorrelation.
3. Equity and sustainability model

The five sets of data from http://www.fao.org/ represent efficiency indicators for Germany, China, India, Brazil and South Africa. Principal component analysis is performed on hunger, food security, consumer price index, resource pressure, forest area per second and the corresponding fair and sustainable production index to obtain variable weight indicators as shown in Figure 5.

![Figure 5. Indicator Weight Chart](image)

According to the principle that the cumulative proportion is greater than or equal to 80%, selecting two principal components \( z_1 \) and \( z_2 \) can explain 90.79% of \( Y \).

According to the load matrix, principal component 1 has a larger load value on \( x_1, x_2, x_3, \) and \( x_5 \), and principal component 2 has a larger load value on \( x_4 \).

The regression analysis of principal components gives the following regression equation:

\[
y = -5.031e-17 + 0.35931 - 0.01782z_2
\]

4. Comprehensive model of food redistribution

According to the above two models, a comprehensive model of food redistribution is obtained by means of factor extraction.

Factor analysis is a statistical technique that studies the extraction of common factors from groups of variables.\[6\] Factor extraction, the component (factor) score coefficient matrix and each individual factor score can be obtained.

The factors were extracted by regression method. Orthogonal rotation method is used to rotate the factors extracted by principal component method. Factor scores are obtained using regression methods and significant output results are analysed.

Observe the correlation coefficient between variables and factors. It can be seen that the first factor is mainly explained by variables 3, 4, 6, 7 and 8, and the second factor is mainly explained by variables 1, 2 and 5.\[7\] For example, "0.69" indicates that the degree of explanation for variable 1 by factor 1 is 69%, and also indicates that the variance contribution of factor 1 to variable 1 is 69%. "0.73" indicates that factor 2 explains 73% of variable 1, and also indicates that factor 2 contributes 73% of variance to variable 1.

Sum of squares of rows: H2. This is the contribution of two factors to the variance of an original variable. For example, the first line "1.01" indicates that factors 1 and 2 explain each variable 1 to a degree of 101%, and also indicates that factors 1 and 2 contribute 101% to the variance of variable 1.
Sum of squares of columns: the sum of squares of column 1 is $4.45 = \lambda_1$, which represents the variance contribution of factor 1 to all original variables. [8] The sum of squares of the second column is $2.79 = \lambda_2$, which represents the variance contribution of factor 2 to all original variables.

Percentage of two factors of total variance: 56% and 35%.
Component 1 represents a variance of 56%, so the variance contribution is 56%. Component 2 represents a variance of 35%, so the variance contribution is 35%.
Cumulative percentage of two factors in total variance: 56% and 90%.
Percentage of the total variance of the two factors: 61% and 39%.
The cumulative interpretation rates of the two factors are 61% and 100%.
The weight of each variable is derived from the sum of the proportions of the variance contribution rates of each principal component.

This paper uses the numbers of severe hunger and moderate hunger, the utility of severe hunger is 2 and the utility of moderate hunger is 1. Add them together and divide by the total number of countries. Another country does not need food very much, but it is more suitable for the first goal of this article to distribute food to this country and produce food more effectively, but it is not fair.

$$U_{fair} = U_{efficient} + \frac{\text{Number poor}}{\text{Number rich}}(U_{poor} - U_{rich})$$

As shown in Figure 6, the utility curve of poor countries is far less sharp than that of rich countries. Therefore, in general, we think that giving to poor countries is more effective than giving to rich countries.

Figure 6. Hunger Utility Curve

Only $U_{fair}$ have get this level the food dealer has enough motivate to transfer the food from rich country to poor country.
Figure 7. Hunger Utility Charts

But we still should consider negative externality of disturbute food. Because this way to distribute food is inefficient. So, it will make the production of food less than efficient way.

We assume the country have two kind of country one kind of country is lack of food and don’t have enough technology and labor to produce the food. We use the number of serious hungry and medium hungry, the utility of serious hungry is 2, the medium hungry is 1, make them sum,and divide the total number of the country. Another kind of country isn’t need food very much but distribute the food to this kind of country is fit to our first target more efficient to produce the food, but not fair. The utility curve of poor country is much less sharp than rich country. So, in general situation we think utility of give poor country is higher than give it to rich country. Only Ufair have get this level the food dealer has enough motivate to transfer the food from rich country to poor country. But we still should consider negative externality of disturbute food. Because this way to distribute food is inefficient. So, it will make the production of food less than efficient way.

5. Conclusion

The model is used to analyze the coefficient of food insecurity and the degree of hunger in each country, so as to obtain the utility level of food distribution in each country. The level of food utility in developed countries is relatively flat, while that in underdeveloped countries is steeper. So, the utility of food distribution in less developed countries is higher than that in developed countries. Therefore, when we measure the total utility of food, we multiply the utility difference of the insecurity factor by the quantity of food mixture and add the original food utility level to maximize the food utility level.

References
[1] Maricke M.W.J. van Leeuwen a, B C E V M, C J O, et al. The relevance of spatial scales in nutrient balances on dairy farms [J]. Agriculture, Ecosystems & Environment, 2019, 269: 125 - 139.
[2] B X F A , C J L A B, B C M , et al. Effect of typhoon rainstorm patterns on the spatio-temporal distribution of non-point source pollution in a coastal urbanized watershed [J]. Journal of Cleaner Production, 2021.
[3] Options for keeping the food system within environmental limits [J]. Nature, 2018, 562 (7728): 519 - 525.
[4] Steans J, Tepe D. Handbook on Gender in World Politics || NGOs, feminist activism and human rights [J]. 2016, 10.4337/9781783470624:187 - 194.
[5] Park K, Lee S. An Empirical Study of CDS Premium on the Korean Sovereign Bond: Some Effect of the CTD Option [J]. Emerging Markets Finance and Trade, 2017, 53: 1 - 17.
[6] Holt E W, Cook E F, Covar R A, et al. Identifying the components of asthma health status in children with mild to moderate asthma [J]. Journal of Allergy & Clinical Immunology, 2008, 121 (5): 1175 - 1180.

[7] Moravec P, Snasel V. [IEEE Sixth International Conference on Intelligent Systems Design and Applications - Jian, China (2006.10.16-2006.10.16)] Sixth International Conference on Intelligent Systems Design and Applications - Dimension Reduction Methods for Image Retrieval [J]. 2006, 2:1055 - 1060.

[8] Russell K G. A comparison of six methods of analysing row-column designs with inter-block information [J]. Statistics & Computing, 1999, 9 (3): 239 - 246.