Access to the Internet and the Chinese household consumption structure: Based on IV regression and PSM analysis

Liping Wang¹, Hao Deng²

¹School of Economics, Shandong University of Finance and Economics, Jinan, China
²School of Economics, Shandong University of Finance and Economics, Jinan, China

Corresponding author and e-mail: Liping Wang, 007wangliping@163.com

Abstract. Based on the data of China's Household Finance Survey (CHFS) in 2017 consisting of 38230 samples across China, this paper uses the instrument variable (IV) regression method and the propensity score matching (PSM) technique to study the influence of access to the Internet on the household consumption structure. The conclusions are as follows: first, access to the Internet is conducive to the upgrading of the consumption structure of the residents. Increase of the income, assets and education level could upgrade the consumption structure. Second, people who do not use the Internet will upgrade the consumption structure if they use the Internet, but the upgrading degree of the consumption structure is limited by the income level. Third, from the regional perspective, the Internet can promote the upgrading of consumption structure in the eastern and western regions of China, but it has no significant effect on the consumption structure in the central and northeast regions.

1. Introduction
The development of the Internet affect all aspects of people’s lives and bring all kinds of convenience. The rise of the “Internet+” is the embodiment of the deep integration of the Internet and industry. The wide application of the Internet has brought opportunities for social development, and has changed the traditional economic development mode. If the Internet can promote the optimization of consumption structure, it will inevitably lead to a new round of economic growth. Therefore, it is of great significance to study the effect of the Internet on the residents’ consumption structure.

The consumption structure has been a hot topic for scholars in recent years. Han et al. (2007) found that the price of housing affects the consumption structure. Wang et al. (2015) found that the commercial health insurance is conducive to promoting the upgrading of consumption structure. Li et al. (2013) and Yang et al. (2016) studied the influence of the scale of family on the consumption structure [1-3]. There is also literature about the relationship between the Internet and consumption structure (Tang, 2018; Zhang et al., 2019). However, most of them do not consider the problem of self-selection in the use of Internet, which may lead to deviation bias of regression results. Therefore, in order to avoid this problem, this paper uses the instrument variable (IV) regression method and the propensity score matching (PSM) technique to study the relationship between access to the Internet and consumption structure [4].

2. Model and data

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd
2.1. Model

In order to study the influence of the use of Internet on the consumption structure, the following model is constructed:

\[ cs_i = \alpha + \beta_1 D_i + \beta_2 X_i + \varepsilon_i \]  \hspace{1cm} (1)

\( cs_i \) is the consumption structure of household \( i \) and is represented by the proportion of monthly enjoyment consumption and development consumption to total household consumption. If this value increases, it indicates that the consumption structure is upgraded. \( D_i \) represents access to the internet. \( X_i \) are the control variables representing the household characteristics.

When evaluating the effects of Internet use, the group using the Internet is the treatment group and the group without the Internet is the control group. If the consumption structure difference between the treatment group and the control group is directly compared, the result may have errors, because whether to use the Internet involves the problem of self-selection. In order to solve this problem, the propensity score matching method is used to estimate the impact of the use of the Internet on the consumption structure. The dummy variable \( D_i = \{0,1\} \) represents whether an individual uses the Internet. If the household uses the Internet, then \( D_i = 1 \), \( c_i = c_{i1} \). Otherwise, \( D_i = 0 \), \( c_i = c_{i0} \). Thus, \( (c_{i1} - c_{i0}) \) is the effect of using the Internet for individual \( i \). Because individual \( i \) can only have one state, and can not get \( c_{i1} \) and \( c_{i0} \) at the same time, this results in the missing data. \( c_i \) could be expressed as:

\[ c_i = D_i c_{i1} + (1 - D_i) c_{i0} = (c_{i1} - c_{i0}) D_i + c_{i0} \] \hspace{1cm} (2)

where \( (c_{i1} - c_{i0}) \) represents the treatment effect of individual \( i \). Because the treatment effect is a random variable, we consider its expected value, that is, the average treatment effect (ATE):

\[ ATE = E(c_{i1} - c_{i0}) \] \hspace{1cm} (3)

The expected treatment effect of taking one random sample from those using the Internet is the average treatment effect (ATT):

\[ ATT = E(c_{i1} - c_{i0}) | D_i = 1 \] \hspace{1cm} (4)

The expected treatment effect of randomly taking one sample from those not using the Internet is the average treatment effect of the control group (ATU):

\[ ATU = E(c_{i1} - c_{i0}) | D_i = 0 \] \hspace{1cm} (5)

If the consumption of Internet users and non-users is compared directly, the bias is:

\[ E(c_{i1} | D_i = 1) - E(c_{i0} | D_i = 0) = E(c_{i1} | D_i = 1) - E(c_{i0} | D_i = 1) + E(c_{i0} | D_i = 1) - E(c_{i0} | D_i = 0) \] \hspace{1cm} (6)

In general, consumers with good financial conditions are more likely to use the Internet and consume more [5-6], so the result bias is positive, overestimating the impact of using the Internet on consumption. Therefore, it is necessary to match the data with propensity score. The outcome variable is the consumption structure. The treatment variable is access to the internet. The covariates include the income, assets, age, education, marital status, health, type of work unit and insurance.

2.2. Data

This study used the data sourced from the 2017 China Household Finance Survey (CHFS) of Southwest University of Finance and Economics. Through screening the missing values, 38230 samples are finally obtained.

| Variable                  | Description                                                                 | Mean value | Standard deviation |
|---------------------------|-----------------------------------------------------------------------------|------------|--------------------|
| Consumption structure     | Share of monthly enjoyment and development consumption in total household consumption | 0.390      | 0.210              |
| Access to the Internet    | =1 if a household has access to the Internet, =0 otherwise                  | 0.470      | 0.500              |

Table 1. Descriptive statistics.
3. Empirical analysis on the influence of internet on consumption structure

3.1. Analysis based on the national level

In the instrument variable regression, we choose "whether there is a smart phone" as the tool variable. According to Table 2, the use of the Internet helps to upgrade the consumption structure. With the increase of income, the residents pursue higher quality living standards and spiritual needs, and increase the recreation consumption and development consumption. The assets help to upgrade the consumption structure. The coefficient of age is negative, statistically significant. The education would promote consumption upgrading, which may be related to the characteristics of highly educated talents. Those with longer years of education are more likely to pursue the spiritual and cultural needs and would spend more on the education and recreation. The consumption structure of those married is better than that of those unmarried. The consumption structure of employees of state-owned units is better than that of employees of non-state-owned units. People who have insurance have a better consumption structure than those who do not.

Table 2. IV regression result.

| Variable                | IV          | (1)  | (2)  | (3)  | (4)  |
|-------------------------|-------------|------|------|------|------|
| Access to the Internet  | 0.0385***   | 0.0280*** | 0.0147** | 0.0144** | 0.00340 | 0.00410 | 0.00571 | 0.00572 |
| Income                  | 0.00350***  | 0.00381*** | 0.00346*** | 0.00346*** | 0.000771 | 0.000767 | 0.000769 |
| Asset                   | 0.00226***  | 0.00358*** | 0.00344*** | 0.00344*** | 0.000720 | 0.000722 | 0.000722 |
| Age                     | -0.00100*** | -0.00101*** | 0.000103 | 0.000103 | 0.000103 | 0.000103 |
| Education               | 0.00359***  | 0.00338*** | 0.000348 | 0.000354 |
Marital status  0.0327***  0.0317***
(0.00434)  (0.00434)
Health          -0.0351***  -0.0351***
(0.00115)  (0.00115)
Type of work unit  0.0106***
(0.00319)
Insurance       0.0251***
(0.00627)
Constant        0.372***  0.320***  0.418***  0.400***
(0.00204)  (0.00860)  (0.0101)  (0.0111)
n                  38230  38230  38230  38230
R²                0.008   0.010   0.039   0.040

Note: (1)*,**,** indicate significance at 10%,5% and 1% levels, respectively. (2) Standard errors are in parentheses.

Although the instrument variable regression has already controlled the endogenous problem, the choice of using the Internet has the "self-selection" problem which should not be ignored. For this reason, this paper uses the propensity score matching method to estimate the results.

![Figure 1](image_url)

**Figure 1.** Graphical representation of propensity score matching.

Figure 1 shows that most observations are on support, indicating that the amount of loss of data for the PSM is few. Table 3 presented the data balancing test results.

**Table 3.** Data balancing test results.

| Matching method                  | Sample  | Pseudo R² | Chi²   | Mean bias (%) | B     | R  |
|---------------------------------|---------|-----------|--------|---------------|-------|----|
| k-nearest neighbor matching (k=1) | Unmatched | 0.326    | 17220  | 60.7          | 149.9* | 0.99 |
|                                 | Matched  | 0.003    | 158    | 4.5           | 13.3  | 0.88 |
| k-nearest neighbor matching (k=4) | Unmatched | 0.326    | 17220  | 60.7          | 149.9* | 0.99 |
|                                 | Matched  | 0.003    | 148    | 4.4           | 12.9  | 0.91 |
| Nearest-neighbor matching within caliper (1:4) | Unmatched | 0.326    | 17220  | 60.7          | 149.9* | 0.99 |
|                                 | Matched  | 0.003    | 148    | 4.4           | 12.9  | 0.91 |
The pseudo $R^2$, chi2, mean bias are greatly reduced after matching, and $B<25\%$, $0.5<R<2$. Thus, the sample is well matched. From Table 4, we can see that access to the Internet can promote the transformation of the consumption structure. This is consistent with the IV regression result. We only analyze the results obtained by the caliper inner neighbor matching (1:4) method.

**Table 4. PSM estimates results.**

| Matching method                        | ATT      | ATU      | ATE      |
|----------------------------------------|----------|----------|----------|
| k-nearest neighbor matching (k=1)      | 0.0306***| 0.0182***| 0.0240***|
| k-nearest neighbor matching (k=4)      | 0.0319***| 0.0186***| 0.0248***|
| Nearest-neighbor matching (1:4)        | 0.0319***| 0.0186***| 0.0248***|

Note: As Table 2.

According to the ATT of the estimated results, access to the Internet has increased the proportion of recreation and development consumption by 3.19%. The ATU and ATE values are both positive and significant at 1 per cent level. ATT>ATE>ATU, and it shows that people who do not use the Internet will upgrade their consumption structure if they use the Internet, but compared with those who use the Internet, people who do not use the Internet still have a lower share of enjoyment and development consumption than those who use the Internet originally. This is due to the income difference between those who use the Internet and those without the Internet.

3.2. **Analysis based on the regional level**

Next, we make analyses of the effect of Internet use on the consumption structure of different regions of China. The data balancing test is passed. The sample is matched well. According to Table 5, in the eastern region, access to the Internet could increase the proportion of enjoyment consumption and development consumption by 4.72%, and is conducive to the upgrading of the consumption structure. Both ATU and ATE are positive and statistically significant. The comparison of the ATT, ATU and ATE shows that people who do not use the Internet will optimize their consumption structure if they use the Internet. Estimates show that people who do not use the Internet will increase the share of enjoyment and development consumption by 3.3% if they use the Internet.

**Table 5. PSM estimates result for different regions.**

| Regions        | ATT       | ATU       | ATE       |
|----------------|-----------|-----------|-----------|
| Eastern region | 0.0472*** | 0.0330*** | 0.0405*** |
|                | (0.00815) | (0.00675) | (0.00566) |
| Central Region | 0.0116    | 0.0137    | 0.0129    |
|                | (0.00984) | (0.0101)  | (0.00789) |
| Western Region | 0.0219**  | 0.0119    | 0.0160**  |
|                | (0.00994) | (0.0106)  | (0.00789) |
| Northeast region| 0.00127  | 0.00109   | 0.00118   |
|                | (0.0120)  | (0.0110)  | (0.00902) |

Note: As Table 2.

According to the ATT of the western region, access to the Internet could upgrade the consumption structure. Both ATU and ATE are positive, but ATU is not significant, indicating that people who do not use the Internet will not have a significant impact on the consumption structure if they use the Internet. In addition, the use of the Internet has no significant impact on the consumption structure in the central and northeast regions. Maybe the reason is that the relatively low income level in the central and northeast regions inhibits the further optimization of the consumption structure.

4. **Conclusions**
This paper studies the influence of access to the Internet on the Chinese household consumption structure and draw the following conclusions: first, the use of the Internet could promote the proportion of consumption and development consumption, and is conducive to the upgrading of the consumption structure of residents. The income, assets, education, marital status, type of work unit and insurance have positive relationship with the consumption structure. Second, people who do not use the Internet will upgrade the consumption structure if they use the Internet, but the upgrading degree of the consumption structure is limited by the income level. Third, from the regional perspective, the Internet can promote the upgrading of consumption structure in the eastern and western regions, but it has no significant effect on the consumption structure in the central and northeast regions. In order to promote the upgrading of residents' consumption structure, the government should not only increase the income and education level of the residents, but also strengthen the construction of the Internet infrastructure, expand the coverage of the Internet and increase the network participation rate of residents.

Acknowledgement
This paper received financial support from the National Natural Science Foundation of China (No. 71773063), the National Social Science Fund of China (No.17ZDA038).

References
[1] Meijiao Wang and Minglai Zhu 2015 Effects of commercial health insurance to consumption and the structure of consumption: based on rational expectation and analysis of household asset structure Insurance Studies 6 pp19-31. (In Chinese)
[2] Sujin Han and Kang L 2007 Household consumption expenditure analysis based on ownership of real property Journal of Consumption Culture 10(3) pp 105-128. (In Chinese)
[3] Xiaonan Li and Yue Li 2013 Consumption structure and influence factors of rural residents in four economic regions of China The Journal of Quantitative & Technical Economics 30(9) pp 89-105. (In Chinese)
[4] Xiao Yang and Xiaotong Jin 2016 Influencing factors of developmental consumption: a comparative study of new generation migrant workers and urban youth of the same age Journal of Agrotechnical Economics 5 pp 60-70. (In Chinese)
[5] Caikun Tang 2018 Analysis on the influence of Internet on the consumption structure of rural residents Statistics & Decision 34(21) pp117-119. (In Chinese)
[6] Yongli Zhang and Lamei Xu 2019 The impact of Internet use on consumption of farmers in western poverty-stricken areas: an analysis based on household survey of 1735 farmers in Gansu province Chinese Rural Economy 2 pp 42-59. (In Chinese)