How does income affect consumption at different quantiles: quantitle regressions based on CHFS(2019)

Shuaizhou Lou & Jichao Lin

May 7, 2022

Abstract

Consumption plays a vital role in promoting economic cycling and boosting a country's development. In the process urbanization, optimizing consumption structure in the rural area is the key step for rural revitalization in modern China. We use the data of China Household Financial Survey (CHFS) in 2019 to explore how income increase affects consumption at different quantile. We find that high consumption family spends more on both survival and non-survival consumption in response to an income increase compared to the low consumption families. We further find that the percentage increment is on the contrary lower for the high consumption cohort due to an already large consumption base.

Keywords: consumption, rural family, quantile regression

Contents

1 Introduction

In most empirical analysis, we focus on the influence of explanatory variables x on the explained variables$y$, and the idea is to analyze the parameter results from the perspective of the mean, namely mean regression. However, mean regression is often affected by extreme values, making parameter estimation very unstable (especially when performing grouped regression). On the other hand, OLS based linear regression model can only allow us to analyze the average effect of x on y. When the conditional distribution $y|x$ is not symmetrical, the conditional expectation $E(y|x)$can hardly reflect the whole condition distribution. If we can estimate some important conditional distributions of $y|x$, we can have a more comprehensive understanding of conditional distribution $y|x$.

2 Comparison between OLS and QR standard

Suppose we have a linear regression model:

$$y_i = \beta_0 + \beta_1 X + \alpha Z + \epsilon_i$$

For OLS estimating method, $\beta_1$should satisfy that,

$$\min_{\hat{\beta}_1} MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

But for Quantile Regression, $\beta_1$should satisfy that,

$$\min_{\hat{\beta}_1} Q_Y(\tau) = \tau \sum_{y_i \geq \hat{y}_i} |y_i - \hat{y}_i| + (1-\tau) \sum_{y_i \leq \hat{y}_i} |y_i - \hat{y}_i|$$

OLS based model can also be regarded as "mean regression" and what it describes is the centralized tendency of conditional distribution. However if the distribution is asymmetric, it cannot well reflect the whole conditional distribution. If the important quantile information can be obtained, such as 1/4 quantile, median, and 3/4 quantile, we can have a more comprehensive understanding.

From a practical point of view, quantile information is also very important. For example, when evaluating the impact of an intervention on the audience, we hope to understand not only the "average" impact of the intervention, but also the "heterogeneous" impact of the intervention on the population at different positions of the characteristic distribution (the end or top of the distribution).

3 Research question

Since reform and opening up, China's economy has grown rapidly. Per capita income of Chinese residents as well as per capita consumption, showed an increasing trend from 2003 to 2019. The income and consumption of both urban and rural residents are rising dramatically. As one of the "troika" that drives China's economy, the share of urban household consumption in annual income,however,dropped from 90% to 70%, after the 2008 global financial crisis and 2011 European debt crisis.

The relationship of income and consumption are in debate centuries ago. Keynes believed that there are many factors that cause consumption changes,such as income,political system,prices, interest rates and so on.But among them the decisive role is income. Residents consumption is absolutely correlated with their income, and the increase of income will lead to the increase of consumption. Friedman divided income into two parts, one part of it is long-term income that can be expected and the other is temporary income.In order to maximize their own benefits, they will change consumption according to the expected long-term income.Aguiar et al. [9] believe that income gap can lead to consumption inequality. Frank et al. believed that income inequality would promote household consumption. Fisher et al. studied the consumption expenditure data from 1987 to 2011 in the United States and found that people's overall consumption level was greatly affected by the widening income gap. We try to constraint our observations to families in rural area, and explore how income increase affect their consumptions at different level.

4 Model specification

The analysis of the influence of income on the consumption of rural households mainly adopts the traditional regression model, which is essentially to analyse the influence of income on the expectation of various consumption conditions, while the conditional expectation is only an indicator reflecting the centralized trend of conditional distribution. The traditional regression model does not take into account the influence of income on the whole conditionally distributed consumption, that is, in different range of consumption, the influence of income on consumption may have structural changes, which reduces the accuracy of the depiction of the influence of income on consumption. In addition, the square sum of residuals in the least square model is easily affected by extreme values, and the regression results are inevitably biased.

Based on the above considerations, this paper adopts the quantile regression model proposed by Koenker & Bassett. Quantile regression model is an extension of mean regression, which can estimate the influence of explanatory variables on the explained variables at each loci. Quantile regression has relatively loose assumptions. When data are distributed with spikes or thick tails, it is not susceptible to extreme values and can avoid deviations. Quantile regression is more robust and can depict data more accurately. The quantile regression model is established as follows

$$Q_q(y_i|income_i, X_i) = \beta_{0q} + \beta_{1q} income_i + \beta_{2q} X_i + \epsilon_i$$

where $y_i$is the households' annual consumptions. We take into account two different kinds of consumptions: survival consumption and non-survival consumption. Survival consumption includes food consumption, clothing consumption, resident consumption and equipment consumption. And non-survival consumption consists of medical consumption, travel consumption and education consumption.

$income_i$ represents the annual total income of rural households. X_i controls for other factors. It includes household head characteristics like age, sex, marriage status, education level and health status of household head, which can affect the consumption of rural households. It also takes other household features like family size, hukou, kid numbers, financial assets and debts. Financial assets are selected as the factor index of household assets, including cash, deposits, stocks and funds. Debts can reflect the uncertainty of the future and play a role in risk sharing.

5 Data

The choose CHFS 2019 as our dataset. After the cleaning of data, we keep the rural household head information and in total we have 9877 observations. The descriptive statistics are shown in the following table. What needs to point out is that both survival and non-survival consumptions have very large standard deviation, and range from less than 2000 to over 1 million. About 85 percent of household heads are male and about 90 percent are married. The average education year is 7 and since we control our observation in rural area, over 90 percent household have agricultural hukou. More information of variables is demonstrated figure1.

Variable	N	Mean	SD	Min	Max
age	10446	54.96	9.687	21	70
male	10446	0.859	0.348	0	1
married	10446	0.885	0.319	0	1
eduyear	10446	7.451	3.342	0	19
CPCmember	10446	0.122	0.327	0	1
HighSchoolAbv	10446	0.138	0.345	0	1
health	10446	0.735	0.441	0	1
job	10446	0.854	0.353	0	1
familysize	10446	3.468	1.680	1	12
hukou agri	10446	0.928	0.259	0	1
hhkid	10446	0.556	0.497	0	1
total income	10446	51624	122630	0	7.863e+06
ln income	10446	9.777	2.343	0	15.88
surv con	10446	31318	38672	1177	1.448e+06
nonsurv con	10446	22991	46663	0	2.328e+06
ln fina assets	10446	8.668	2.752	0	15.37
ln total debt	10446	3.943	5.109	0	15.71

Figure 1: Descriptive Statistics

To have a clearer understand of the distribution of the survival and non-survival consumptions, we plot out the density of both variables. It can be shown that both consumptions are very skewed and have thick right tails, which indicates that the outcome would be very biased if we simply adopt linear regression.

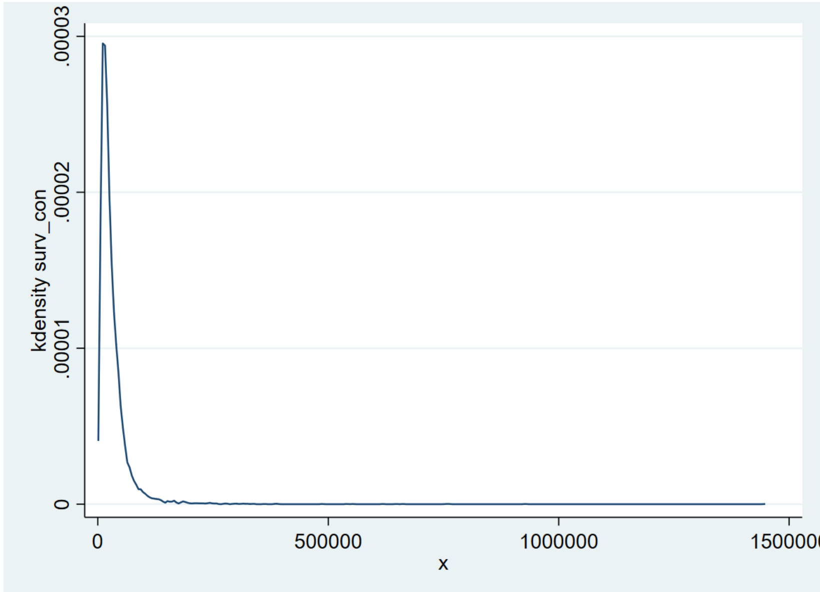


Figure 2: Cumulative desity function of survival consumption

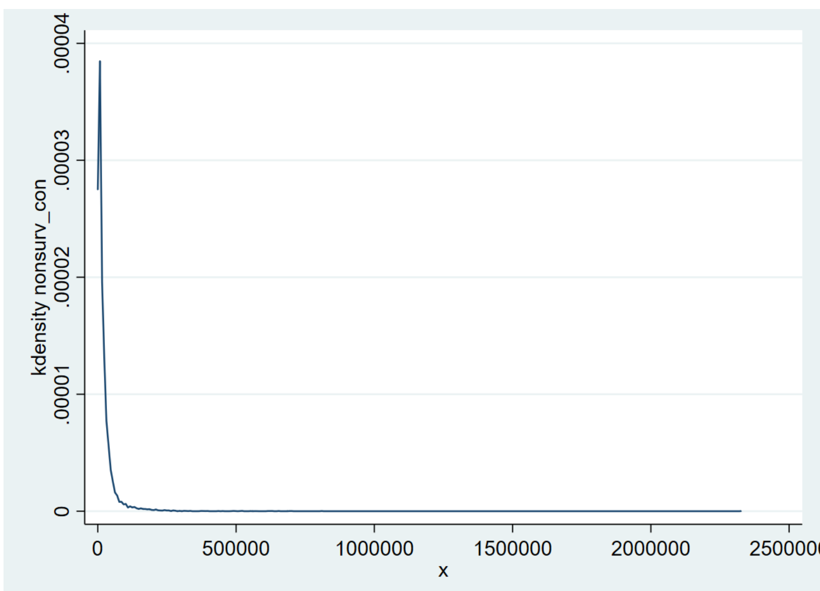
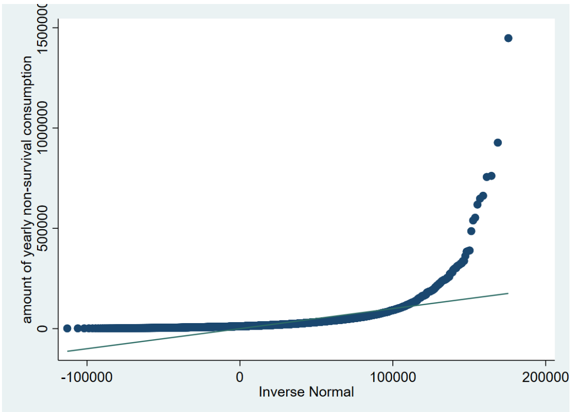
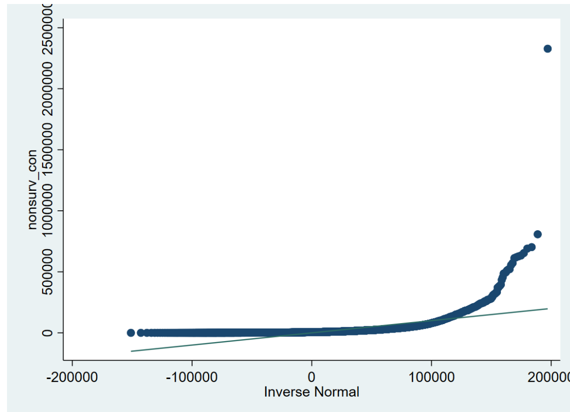


Figure 3: Cumulative desity function of non-survival consumption

Further, we plot out the Q-Q graph, the dots deviate from the straight line, which means that the consumption are not normally distributed.



(a) origin waveframe



(b) wavefram frame

Figure 4: Q-Q graph for survival and non-survival consumptions

6 Empirical Results

We first regress on the survival consumption with total income. And the results are shown in the following table. Comparing to first column, which is the results of linear regression, for low consumption family, one yuan increase in total income leads to only 0.038 yuan in survival consumption. And this income propensity to consumption increases with the quantile rise to 0.5 and 0.75, with 0.073 and 0.125 respectively. This indicates that for high consumption family, they have a tendency to spend more in survival consumption when they get their income raised. All of the above results are significant at 99 percent level.

Other findings are that the elder the household head, the less the family will spend on the consumption with raise in income. Education year health condition and family size all have significantly positive effects on consumption. Financial assets and debts have positive influences on consumption. But the results can be biased since these two variables can be corelated with each other.

For non-survival consumption shown in the following table, the results are pretty similar. The higher the consumption is, the larger propensity of the increase in income to consumption. Again, all these results are significant at 99 level of significance.

	Regression on survival consumption			
	OLS	q25	q50	q75
total_income	0.050***	0.038***	0.073***	0.125***
	(0.003)	(0.001)	(0.002)	(0.003)
age	-199.155***	-94.338***	-214.434***	-296.318***
	(43.997)	(17.278)	(24.494)	(39.522)
male	-304.680	-889.758**	-1610.153***	-1910.969*
	(1085.775)	(426.405)	(604.486)	(975.340)
married	934.205	227.267	1132.640*	1414.870
	(1217.061)	(477.964)	(677.577)	(1093.273)
eduyear	324.433***	187.627***	205.542***	168.578
	(117.624)	(46.193)	(65.485)	(105.661)
CPCmember	-794.711	1042.132**	650.625	1192.783
	(1113.259)	(437.199)	(619.787)	(1000.029)
health	3537.445***	1582.189***	2369.758***	3499.716***
	(849.496)	(333.614)	(472.942)	(763.094)
job	-3570.887***	-637.085	-1365.733**	-3157.781***
	(1052.253)	(413.240)	(585.823)	(945.228)
familysize	3519.331***	1694.558***	2655.910***	4131.999***
	(304.916)	(119.746)	(169.756)	(273.903)
hukou_agri	-9242.493***	-3650.143***	-6270.542***	-6345.118***
	(1404.373)	(551.525)	(781.860)	(1261.534)
hhkid	44.505	-244.768	-721.783	-1078.891
	(1057.663)	(415.365)	(588.835)	(950.088)
ln_fina_assets	1423.384***	521.535***	637.122***	793.580***
	(141.448)	(55.549)	(78.749)	(127.061)
ln_total_debt	592.227***	99.069***	179.236***	371.036***
	(73.197)	(28.746)	(40.751)	(65.752)
constant	18916.609***	10056.865***	21621.524***	30096.802***
	(3707.980)	(1456.196)	(2064.352)	(3330.840)
N	9877	9877	9877	9877
R_2	0.107	0.087	0.121	0.146

Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 5: regression on survival consumption

	Regression on non-survival consumption			
	OLS	q25	q50	q75
total_income	0.040***	0.025***	0.041***	0.074***
	(0.003)	(0.001)	(0.001)	(0.003)
age	-113.431**	-28.003**	-68.856***	-158.292***
	(47.102)	(12.626)	(19.645)	(38.271)
male	-1460.267	-646.092**	-1354.368***	-1798.288*
	(1162.412)	(311.585)	(484.819)	(944.481)
married	2104.535	-155.380	-100.498	483.170
	(1302.964)	(349.260)	(543.441)	(1058.682)
eduyear	608.459***	129.576***	254.868***	409.358***
	(125.926)	(33.755)	(52.521)	(102.318)
CPCmember	6184.507***	786.787**	2009.929***	4453.991***
	(1191.836)	(319.472)	(497.091)	(968.388)
health	-1914.619**	-1172.525***	-1564.575***	-2475.183***
	(909.456)	(243.780)	(379.316)	(738.949)
job	-6048.650***	-594.173**	-1871.465***	-3753.894***
	(1126.523)	(301.965)	(469.851)	(915.320)
familysize	4028.851***	1764.352***	3027.961***	4629.862***
	(326.438)	(87.502)	(136.151)	(265.236)
hukou_agri	-3277.161**	-470.301	-1050.944*	-3404.687***
	(1503.498)	(403.013)	(627.079)	(1221.619)
hhkid	3692.820***	407.780	1762.926***	3688.979***
	(1132.315)	(303.517)	(472.267)	(920.027)
ln_fina_assets	1032.204***	221.427***	306.882***	439.924***
	(151.432)	(40.591)	(63.159)	(123.041)
ln_total_debt	740.653***	182.001***	298.075***	563.242***
	(78.363)	(21.005)	(32.684)	(63.671)
constant	2617.775	-289.536	2757.217*	11621.321***
	(3969.698)	(1064.079)	(1655.683)	(3225.452)
N	9877	9877	9877	9877
R_2	0.114	0.087	0.127	0.147

Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 6: regression on survival consumption

7 Further comparison between linear regression and quantile regression

To further check the performance of quantile regression, we choose six variables and plot out their coefficients and confidence interval at quantile 0.05 to 0.9 with steps 0.05. Also, we plot the coefficients of the linear regression in the same graph as comparison.

As shown in the graph below, the coefficients of some of the variables are greatly deviate from the horizontal lines, which represents the coefficients of OLS. One specific findings is that the deviation enlarges as the quantile get larger. This indicates that for both high survival consumption and non-survival consumption, the effects of income increase are quite different from other quantile, and this feature can not be capture using simple OLS regression.

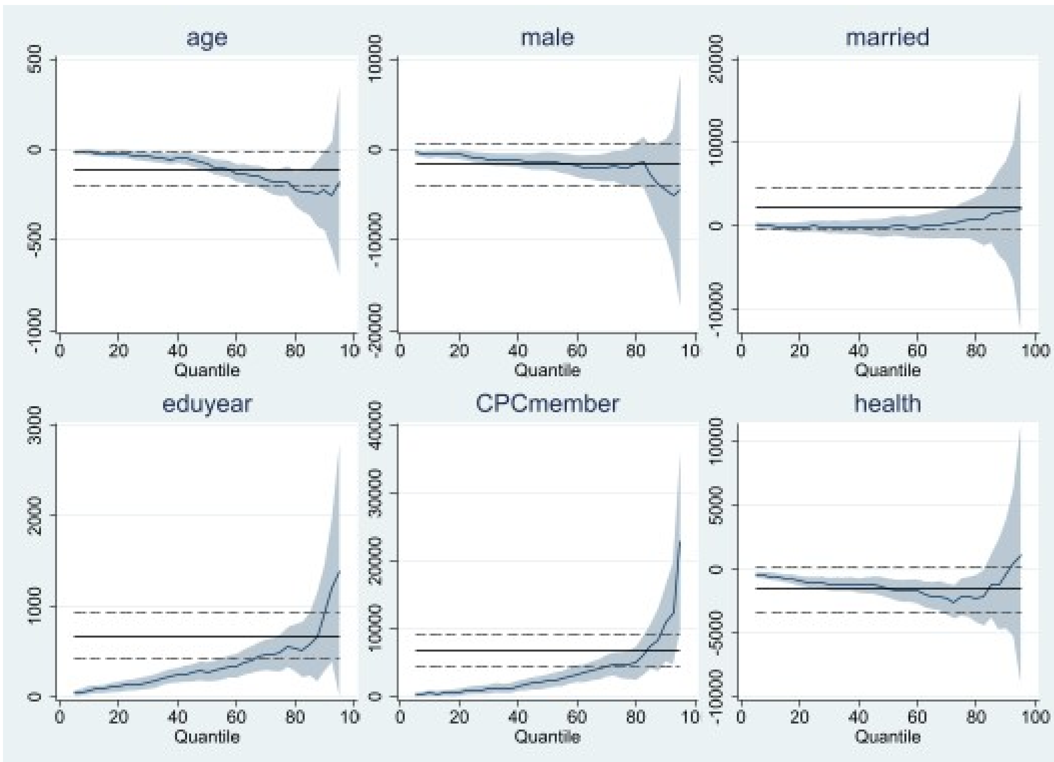


Figure 7: Coefficients of variables for survival consumption

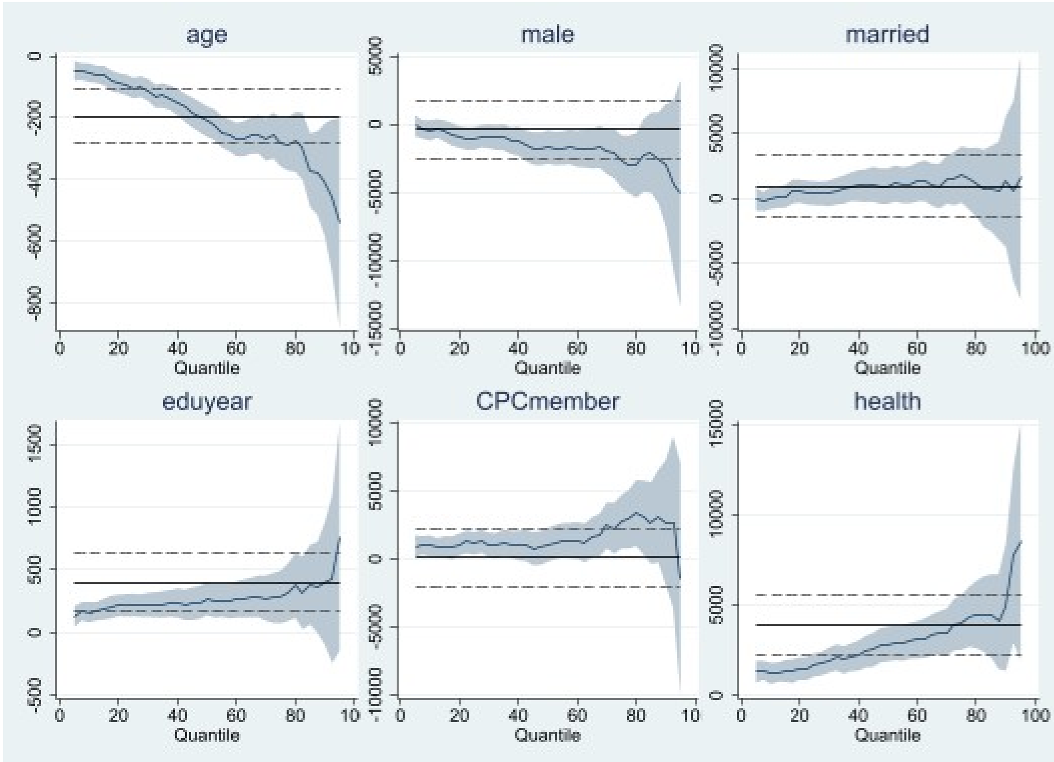


Figure 8: Coefficients of variables for non-survival consumption

8 Further exploration

We further explore the question by taking the logarithm of both consumptions and income due to the concern of the widely spread of the two variables. The economic indication of the coefficient of log-income can be explained to be percentage increment in consumption due to one percent increase in income. Similarly, we compare OLS and quantile regression and the results are reported in the following two tables.

It's worth noting that contrary to basic model where logarithm was not taken, for the low consumption family, 10 percent increase in the income increased their survival consumption by 0.33 percent, whereas for the high consumption cohort, the amount decreases to 0.22 percent. It's therefore can be briefly concluded that the higher consumption the family has, the higher absolute money value they would spend on survival consumption but the lower percent increment of consumption in response to an income increase. And the pattern is the same for non-survival consumption. All the results are significant at a 5 percent level of significance. The intuition behind these findings is that the monetary base for the low consumption family is much smaller than those for the high consumption family. Although the higher cohort spend much more money in response to the income increase, the relative amount to their existing consumption is not so big. The policy indication is that monetary aid to the low consumption families can be more effective to boost consumption and enhance their situation.

	Regress on logarithm of survival consumption			
	OLS	q25	q50	q75
ln_income	0.025***	0.033***	0.029***	0.022***
	(0.003)	(0.004)	(0.004)	(0.004)
age	-0.008***	-0.007***	-0.010***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
male	-0.039*	-0.043	-0.086***	-0.071***
	(0.021)	(0.028)	(0.026)	(0.025)
married	0.088***	0.095***	0.118***	0.121***
	(0.024)	(0.032)	(0.029)	(0.028)
eduyear	0.013***	0.016***	0.012***	0.010***
	(0.002)	(0.003)	(0.003)	(0.003)
CPCmember	0.049**	0.069**	0.031	0.069***
	(0.022)	(0.029)	(0.026)	(0.026)
health	0.144***	0.135***	0.141***	0.151***
	(0.017)	(0.022)	(0.020)	(0.020)
job	-0.067***	-0.064**	-0.050**	-0.092***
	(0.021)	(0.028)	(0.025)	(0.024)
familysize	0.112***	0.109***	0.110***	0.117***
	(0.006)	(0.008)	(0.007)	(0.007)
hukou_agri	-0.208***	-0.219***	-0.243***	-0.214***
	(0.028)	(0.037)	(0.033)	(0.032)
hhkid	0.045**	0.045	0.039	0.063***
	(0.021)	(0.028)	(0.025)	(0.024)
ln_fina_assets	0.052***	0.055***	0.053***	0.048***
	(0.003)	(0.004)	(0.003)	(0.003)
ln_total_debt	0.014***	0.011***	0.010***	0.014***
	(0.001)	(0.002)	(0.002)	(0.002)
constant	9.270***	8.694***	9.390***	9.795***
	(0.077)	(0.101)	(0.091)	(0.089)
N	9877	9877	9877	9877
R2	0.114	0.136	0.142	0.138

Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 9: regress on log of survival consumption

	Regress on logarithm of non-survival consumption			
	OLS	q25	q50	q75
ln_income	0.020***	0.025***	0.015***	0.015**
	(0.005)	(0.006)	(0.006)	(0.006)
age	-0.010***	-0.010***	-0.011***	-0.012***
	(0.001)	(0.002)	(0.002)	(0.002)
male	-0.149***	-0.132***	-0.140***	-0.105***
	(0.033)	(0.043)	(0.039)	(0.039)
married	0.349***	0.347***	0.273***	0.292***
	(0.037)	(0.048)	(0.044)	(0.044)
eduyear	0.034***	0.035***	0.037***	0.031***
	(0.004)	(0.005)	(0.004)	(0.004)
CPCmember	0.185***	0.179***	0.200***	0.174***
	(0.034)	(0.044)	(0.040)	(0.040)
health	-0.197***	-0.255***	-0.181***	-0.185***
	(0.026)	(0.034)	(0.031)	(0.031)
job	-0.171***	-0.127***	-0.194***	-0.210***
	(0.032)	(0.042)	(0.038)	(0.038)
familysize	0.216***	0.224***	0.218***	0.190***
	(0.009)	(0.012)	(0.011)	(0.011)
hukou_agri	-0.150***	-0.141**	-0.130**	-0.202***
	(0.043)	(0.056)	(0.050)	(0.051)
hhkid	0.346***	0.432***	0.385***	0.314***
	(0.032)	(0.042)	(0.038)	(0.038)
ln_fina_assets	0.062***	0.078***	0.065***	0.058***
	(0.004)	(0.006)	(0.005)	(0.005)
ln_total_debt	0.034***	0.035***	0.032***	0.032***
	(0.002)	(0.003)	(0.003)	(0.003)
constant	7.986***	6.982***	8.057***	9.066***
	(0.118)	(0.154)	(0.139)	(0.141)
N	9865.000	9865.000	9865.000	9865.000
R2	0.186	0.202	0.192	0.165

Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 10: regress on log of non-survival consumption

9 Conclusion

OLS regression can only give the overall mean of the whole dataset, while the distribution is not normal, the estimation can be bad. The quantile regression captures the characteristics of different range of dependent variables and therefore solve the problem of OLS regression.

We investigate the effects of income increase in family's survival and non-survival consumption. Both income and consumption distribution are extremely skewed thus the traditional OLS estimation cannot have a good illustration of the consumption propensity of income. We adopt quantile regression on both the consumption and logarithm of consumption with income and log income respectively. We find that for high consumption family, they spend more on both survival and non-survival consumptions for an increase in income as compare to the low consumption families. But the percentage increase is in the contrary smaller due to the already large monetary base on the consumption. We thus suggest that an policy aiming at maximize the social welfare and boost consumption may target at the low consumption families, due to an decreasing marginal utility to consumption.

Reference

[1] Roger Koenker. Quantile regression: 40 years on. Annual Review of Economics, 9(4):155-176, August 2017.

[2] Koenker R, Hallock K F. Quantile regression[J]. Journal of economic perspectives, 2001, 15(4): 143-156.

[3] NIE rong,YANG Dan, SHEN Da-juan. An empirical study on influence of China's rural income brakets on household consumption structure-micro evidence based on CFPS data. Journal of Northeastern University, 2020, 22(4): 29.