How Does the Consumers' Mental State Impact Their Households' Consumption Pattern? – Evidence from CFPS 2018

Yuxin Gong (101259039)
ECON 2708 Applied Data Analysis (2022 Winter)
Instructor: Carolina Czastkiewicz
(3458 words, 7 Pages excluding figures and table)

Introduction

The post-epidemic society has witnessed a series of economic disasters ranging from local goods shortage to global supply chain crisis, and it has also brought down numbers of companies that failed to understand consumers' behaviors of which some were directly geared by the consumer's mental state (for example, panic buy). Undoubtedly, the importance of understanding how psychological factors impact consumer behavior has been underlined by the breakout of COVID-19 (Crosta, et al., 2021).

The effect of consumers' mental state on their consuming decisions has been examined by previous studies. In general, consumers make purchase decisions under the influence of their culture, family, personality, psychological factors, etc. More specifically, studies on stress and consumer behavior have revealed that anxiety and stress are panic buying predictors (Lins, et al., 2021) and that stress influences consumer's saving and spending (Durante, et al., 2016).

Many Chinese researchers have also probed into the relationship between consumer mental state and consumer behavior in China. But so far, their studies either used outdated date (for example, from year 2013) or only focused on a particular region or group. And no study has yet analyzed the national data in post-epidemic era.

To better understand how the mental state of Chinese consumers impact their household consumption patterns and the implications for post-COVID business world, this paper uses the latest data from China Family Panel Studies in year 2018 to run OLS regressions, and interpreted the available data from year 2020 based on the regression results.

The paper holds the following **hypotheses**:

(1) A worse mental state could lead to consumers buying more non-essential goods as a way of comfort.
(2) More specifically, if a person is less satisfied with his life, he'll spend a bigger proportion of the household expenditure on non-essential goods as a coping strategy for his dissatisfaction;
(3) Similarly, if a person is less confident in his future, he'll spend a bigger proportion of the household expenditure on non-essential goods.
(4) The consumers in 2020 experienced the traumatic epidemic, and therefore are expected to have a lower level of life satisfaction and confidence in the future, which suggests more spending on non-essential goods.

However, using life satisfaction and confidence in the future as measurement of mental state, percentage of household expenditure on non-essential goods as the measurement of consumption pattern, this paper arrives at the **conclusions** as follows:

(1) Significantly, consumers' life satisfaction and confidence in the future are both in a quadratic relationship with the percentage of household expenditure on non-essential goods. Before the they increase to a turning point, the dependent variable decreases but afterwards increases.
(2) Consumers in 2020 shows significantly more confidence in the future than those in 2018, and no difference in their life satisfaction. Combined with the previous regression results and assuming that the relationship stays the same in 2020, it's expected that people would spend more out of their total spending on the non-essential goods.

Data Description

Data

Data used in this paper are sourced from the 4th wave of China Family Panel Studies (in year 2018), including the individual dataset, the child proxy dataset, the family economy dataset, and the family roster dataset. The individual dataset from the 5th wave (in year 2020) is also used. The employment of the latest national data guarantees the timeliness and universality of the results.

3,961 observations from year 2020 are used in the final regression model. The unit of observation is by person and household (the mental state of head of the house matched with the expenditure of his/her household).

Variables and Measurement

Dependent Variables

The dependent variable of this paper is the consumption pattern measured by the percentage of expenditure on non-essential goods out of the total household expenditure, which is a common way to differentiate consumption patterns.

The original data categorized the household expenditure into 11 categories: food (with two sub-categories: eating at home and dining out), clothing, entertainment, travel, education, housing and transportation, durable goods, medicine, personal care and insurance, social help, and others. Among them, I decide that expenses on eating at home, clothing, housing and transportation, and medicine are the essential goods, since they have a low elasticity of demand, i.e. people spend money on them no matter what.

Independent Variables

The independent variables are life satisfaction and confidence in the future of the head of the household; together they constitute the measurement of an individual's mental state.

The life satisfaction variable is measured by the response to the question "Are you satisfied with your life?" on a 5 point-scale: from 1 to 5 presenting "Very unsatisfied" to "Very satisfied". "Not applicable" is shown as -8, unknown as -2, and refuse as-1, which should all be treated as missing values.

The confidence in the future variable is accessed using the question "How confident are you about your future?". The responses are also on a 5 point-scale: from 1 to 5 presenting "Not confident at all" to "Very confident". Missing values are treated the same as those of life satisfaction variable.

Control Variables

Control variables are demographics of the head of the household and the characteristics of the household. Demographics variables of the head of the household include the age, gender, marital status, education level of him/her. The household characteristics are measured by the family size, the number of the old people, the number of the children, the total annual income, total current savings, the province, information on whether it's located in the urban area, and any big event (eg. a wedding, funeral, etc) happened in the household in the past 12 months.

Table 1 in the Data Appendix shows the full list of my variables and their descriptions.

Descriptive Statistics

Descriptive Statistics on Dependent Variables

To get a general idea of the consumption pattern of the households, Figure 1 – Figure 4 are created.

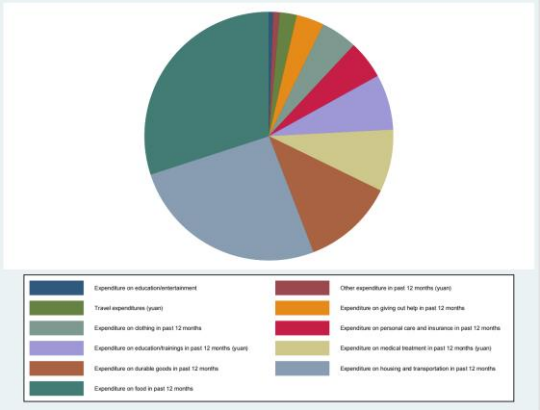

Figure 1 Expenditure on All Categories by Percentage

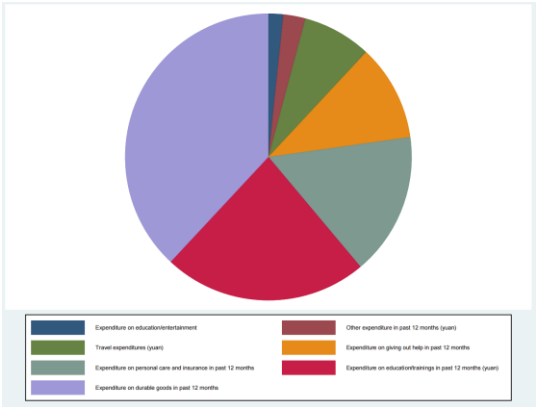

Figure 2 Expenditure on All Categories in Non-essential Goods by Percentage

As is shown, in average, households spend the most part of their expenses on food, closely followed by the housing and transportation. This quite conforms to the fact in China that people are coping with the soaring commodity prices and the rocketing housing expenses, the biggest headaches of contemporary Chinese households, and reaffirms the

burden of housing on Chinese (Bing, 2021)[1].

In terms of non-essential goods, they spend most part of their expenses on durable goods, followed by education, which are both considered long-term investments. This shows that even in non-essential spending, Chinese tend to invest rather than consume. This, in accordance with the insights of Hofstede's cultural model, suggests a high level of long-term orientation and a low level of indulgence of Chinese (See comparison with other countries in Figure 3).

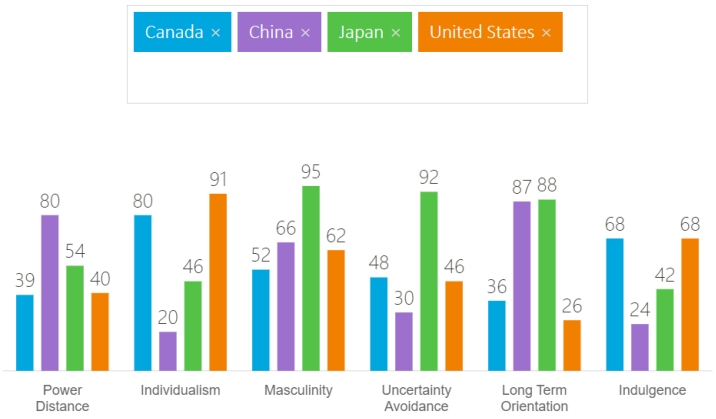

Figure 3 Comparison of 6 Dimensions of Cultural Characteristics

Source: Hofstede Insights Organisational Culture Consulting

In terms of the statistics of the y variable itself, Figure 4 shows that about 23% of the households spend less than 10% of their annual expenditure on non-essential goods. The minimum of the y variable is 0, maximum is 97.46%, and the mean is 31.13% (Figure 5).

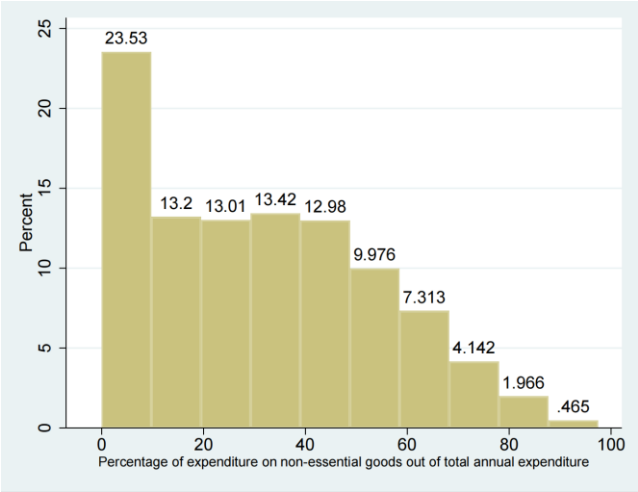

Figure 4 Histogram of Frequency Percent: Percentage of Expenditure on Non-essential Goods out of Total Annual Expenditure

[1] https://www.chinadaily.com.cn/a/202112/07/WS61ae9c26a310cdd39bc79b4d.html

Variable	Obs	Mean	Std. Dev.	Min	Max
noness_perc	9,463	31.13156	22.71741	0	97.46251

Figure 5 Summary of the Statistics of the Dependent Variable

Descriptive Statistics on Independent Variables

Figure 6 shows a summary of the statistics of all the independent variables. The two main independent variables *life_sat* and *future_con* both have a minimum of 1 and maximum of 5. Figure 7 shows that about 35% of people said they are 5/5 satisfied with their life and about 43.04% of people said they are 5/5 confident in their future. The majority of people reported a mental state (life satisfaction and confidence in the future) above 3/5.

Variable	Obs	Mean	Std. Dev.	Min	Max
life_sat	9,511	3.958364	.9735466	1	5
future_con	9,503	4.101336	.9671196	1	5
male	9,738	.5174574	.4997208	0	1
age	9,738	49.19758	15.45763	12	93
edu	4,395	1.061433	.6229953	0	2
married	9,560	.8382845	.3682089	0	1
finc	9,542	66531.64	83682.63	0	2000000
savings	9,340	55822.66	168693.9	0	7000000
familysize18	9,738	3.605874	1.916597	1	21
big_event	9,733	.132847	.3394268	0	1
child_n	9,738	.8882727	1.250591	0	10
old_n	9,738	.4385911	.7157164	0	3
urban18	9,582	.5185765	.4996809	0	1

Figure 6 Summary of the Statistics of All the Independent Variables

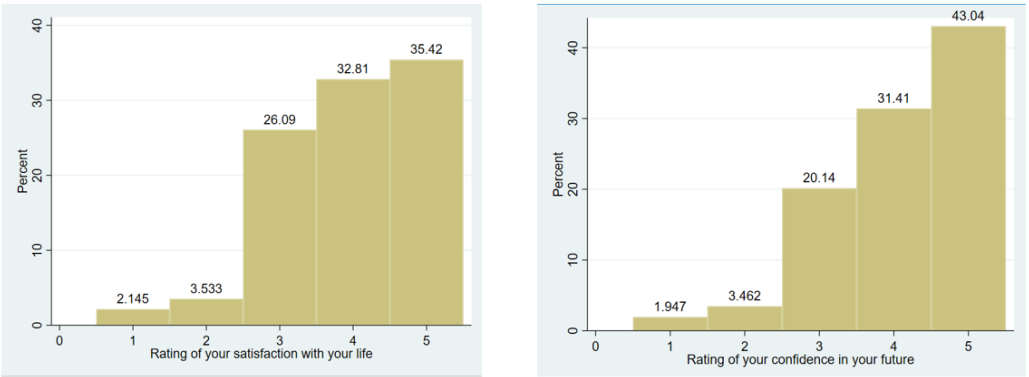

Figure 7 Histogram of Frequency Percent: Life Satisfaction and Confidence in the Future

Regression Analysis

Model and Methods

In 1789, Bentham first put forward a consumer utility model that took psychological factors into account. Bentham's model and its extension hold that consumers' utility is mainly the result of the positive and negative emotions. Loewenstein (1996) furthered this by incorporating irrational factors (including perceived mental state and negative emotions, etc.) into the utility model to investigate the influence of consumers' irrational factors on consumer behavior. Their model serves as the theoretical foundation for the regressions in paper.

To arrive at the most accurate results as possible, I conducted the analyses in following steps. First, only the main independent variables are taken into account (Model 1); then, demographical factors are added (Model 2); next, household characteristics are added (Model 3); finally, the squared terms of the two main independent variables are added to make fore the final model (Model 4).

The final regression model (referred to as Model 4.3 in later analyses) is as follows.

$$noness\_peerc = \beta_0 + \beta_1 life\_sat + \beta_2 sat2 + \beta_3 future\_con + \beta_4 future2 + \beta_5 male + \beta_5 age + \beta_6 edu + \beta_7 married + \beta_8 finc + \beta_9 savings + \beta_{10} familysize18 + \beta_{11} big\_event + \beta_{12} child\_n + \beta_{13} old\_n + \beta_{14} urban18 + \beta_{15} provcd18$$

Pre-regression Checks

Before the actual regression, checks are performed as a rough prediction of the regression results. As Figure 8 shows, the two main variables have a low linear correlation with the y variable, lower than that of the most control variables, which suggests low linear relationships. This problem will be addressed later by adding the squared terms of the two variables.

The correlation between the two main variables however, is the highest of all, indicating possibility of collinearity. To check on this, I performed a factor test on these two variables, the results of which showed a p-value of 0 and a KMO of 0.5 and therefore suggested no need to use the factor analysis.

	noness~c	life_sat	future~n	male	age	edu	married	finc	savings	famil~18	child_n	old_n
noness_perc	1.0000											
life_sat	-0.0645	1.0000										
future_con	-0.0122	0.4568	1.0000									
male	0.0559	-0.0126	0.0277	1.0000								
age	-0.3146	0.1469	-0.0108	0.0397	1.0000							
edu	0.3754	-0.0890	-0.0222	0.0391	-0.4998	1.0000						
married	-0.1399	0.1330	0.0561	-0.1570	0.2354	-0.1503	1.0000					
finc	0.2417	0.0260	0.0100	0.0031	-0.1329	0.3114	0.0750	1.0000				
savings	0.1366	0.0392	0.0039	0.0342	-0.0451	0.1815	0.0419	0.3749	1.0000			
familysize18	-0.1643	0.0753	0.0437	-0.0502	0.1586	-0.2271	0.4551	0.0490	0.0206	1.0000		
child_n	-0.0799	0.0388	0.0354	-0.0750	0.0271	-0.1451	0.3519	-0.0257	-0.0257	0.6496	1.0000	
old_n	-0.1753	0.0776	0.0086	0.0440	0.4375	-0.2577	0.0538	-0.0597	-0.0170	0.2746	0.0326	1.0000

Figure 8 Correlation Table of all the Variables

Regression Results

Model 1

First, only the main variables are taken into account as independent variables. The results are shown in Figure 9. As is shown, the coefficients are all statistically significant, but the R-squared is very small (0.0057). The negative coefficient for life satisfaction variable suggests that people with lower satisfaction are prone to spend more on non-essential goods (aligned with my hypothesis (2)); while the positive coefficient for confidence in future shows that people with higher levels of confidence in future tend to spend more in non-essential goods (rejected my hypothesis (3)). P-value of the whole model is 0, so this model is not meaningless; however, this result is significant but doesn't explain too much of the variation.

```
. reg noness_perc life_sat future_con

      Source |       SS           df       MS      Number of obs   =     9,252
-------------+----------------------------------   F(2, 9249)      =     26.62
       Model |  27224.3539         2  13612.1769   Prob > F        =    0.0000
    Residual |   4730100.7     9,249  511.417526   R-squared       =    0.0057
-------------+----------------------------------   Adj R-squared   =    0.0055
       Total |  4757325.05     9,251  514.249816   Root MSE        =    22.615

------------------------------------------------------------------------------
  noness_perc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
     life_sat |  -1.791797   .2782082    -6.44   0.000    -2.337146   -1.246447
  future_con |    1.72669   .2805648     6.15   0.000     1.176721    2.276659
       _cons |   31.23952   1.160386    26.92   0.000     28.96491    33.51413
------------------------------------------------------------------------------
```

Figure 9 Results for Model 1

Model 2

Obviously, consumers' decisions would also be impacted by their personal factors, so I added in the demographic variables in and got the results shown in Figure 10.

The R-squared increased a great deal (0.1628), but now the two main variables have very insignificant coefficients. This model shows that, in this case, the spending of on the household is not very much dependent on the individual's overall mental situation, but instead, more by the demographic features.

```
. reg noness_perc life_sat future_con male age edu married

      Source |       SS           df       MS      Number of obs   =     4,254
-------------+----------------------------------   F(6, 4247)      =    137.68
       Model |  330165.319         6  55027.5532   Prob > F        =    0.0000
    Residual |  1697402.28     4,247  399.670893   R-squared       =    0.1628
-------------+----------------------------------   Adj R-squared   =    0.1617
       Total |   2027567.6     4,253  476.738209   Root MSE        =    19.992

------------------------------------------------------------------------------
  noness_perc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
     life_sat |  -.2954077   .3616455    -0.82   0.414    -1.004422    .4136066
  future_con |   .0783423   .3824033     0.20   0.838    -.6713679    .8280526
         male |   2.018826   .6250426     3.23   0.001     .7934161    3.244236
          age |  -.2918841   .0299021    -9.76   0.000    -.3505078   -.2332604
          edu |   9.684162    .570513    16.97   0.000     8.565658    10.80267
      married |  -2.649447   .8032179    -3.30   0.001    -4.224174    -1.07472
       _cons |   39.50563   2.196965    17.98   0.000     35.19843    43.81283
------------------------------------------------------------------------------
```

Figure 10 Results for Model 2

Model 3

The spending is also impacted by the family's socio-economic situation. So 5 other variables are added in. Results are shown in Figure 11. The R squared has increased (0.2077) and most of the control variables have significant coefficients, but the coefficients for the main variables are not nearly significant. This could be because the two main variables are in a non-linear relationship with the y variable.

Source	SS	df	MS	Number of obs	=	3,961
				F(43, 3917)	=	23.88
Model	389107.922	43	9049.02144	Prob > F	=	0.0000
Residual	1484017.16	3,917	378.865755	R-squared	=	0.2077
				Adj R-squared	=	0.1990
Total	1873125.09	3,960	473.011385	Root MSE	=	19.464

noness_perc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
life_sat	-.4420903	.3702985	-1.19	0.233	-1.168086	.2839057
future_con	.097861	.3914005	0.25	0.803	-.6695071	.8652291
male	1.918545	.6405066	3.00	0.003	.6627871	3.174303
age	-.2845346	.0329968	-8.62	0.000	-.3492271	-.2198422
edu	6.641554	.6458839	10.28	0.000	5.375253	7.907854
married	-2.508771	.9430247	-2.66	0.008	-4.357637	-.6599052
familysize18	-1.002738	.2412462	-4.16	0.000	-1.475717	-.5297576
big_event	-.304458	.8925026	-0.34	0.733	-2.054272	1.445356
child_n	.7392842	.3121669	2.37	0.018	.1272592	1.351309
old_n	-.3891212	.6277982	-0.62	0.535	-1.619963	.841721
finc	.0000354	4.09e-06	8.66	0.000	.0000274	.0000434
savings	4.30e-06	2.05e-06	2.09	0.036	2.73e-07	8.33e-06
urban18	2.902972	.7051394	4.12	0.000	1.520497	4.285447

Figure 11 Results for Model 3

Model 4

Based on the previous analyses, I added the squared terms of the two main variables first respectively (Model 4.1 and Model 4.2) and then together. The results are in Figure 12. The adding of the squared term didn't changed the R squared too much, but does add to the significance of the coefficients.

Interestingly, the results here show that the life satisfaction and confidence in the future are both quadratically related to the y variable. As the x variables increase, before a turning point, the value of the y variable decreases and then increases afterwards. Calculated from coefficients of the original variable and the squared term, the turning point for life satisfaction is around 3.2 and for confidence in the future it's 3.83.

One possible reason for this is that, before a certain level of life satisfaction and confidence in the future, consumers do spend more on non-essential goods to compensate themselves for the bad mental state but only to keep the overall utility stable; in other words, the utility lost due to the bad mental state could be offset by the spending on the non-essential goods and vice versa so that the consumers get a constant basic level of perceived utility. But after that certain level of life satisfaction and confidence in the future, as the mental state get better, consumers' expected utility also exceeds a threshold, in other words, they get "greedier", the utility brought by the rising life satisfaction and the confidence is not enough, and they want to spend more on the non-essential goods.

I think these two scenarios represent two stages of society development. In a less developed society, people only ask for so much, for example, they only want to make ends meet and they're satisfied, and they want to keep this balance in all time. As the society transformed into a more developed form, people becomes more hedonic and tend to embrace and long for all kinds of fun.

This phenomenon could also just be linked with the consumer behavior of different generations of people. The older generations (eg. people born in the 1950s or earlier) in China are in the former scenario: they're used to the basic standards of life and will even feel guilty if they "enjoy themselves too much". While the younger generations are in the latter: they want as much as possible to fulfill their desire and even consumerism.

VARIABLES	Model 4.1	Model 4.2	Model 4.3
life_sat	5.325***	-0.395	4.041**
	(1.866)	(0.370)	(1.998)
sat2	-0.817***		-0.630**
	(0.259)		(0.279)
future_con	0.171	6.304***	4.380*
	(0.392)	(2.222)	(2.379)
male	1.964***	1.986***	2.000***
	(0.640)	(0.640)	(0.640)
age	-0.278***	-0.278***	-0.276***
	(0.0330)	(0.0330)	(0.0330)
edu	6.428***	6.474***	6.363***
	(0.649)	(0.648)	(0.650)
married	-2.516***	-2.545***	-2.539***
	(0.942)	(0.942)	(0.942)
finc	3.45e-05***	3.50e-05***	3.45e-05***
	(4.09e-06)	(4.09e-06)	(4.09e-06)
savings	4.37e-06**	4.34e-06**	4.38e-06**
	(2.05e-06)	(2.05e-06)	(2.05e-06)
familysize18	-1.002***	-0.982***	-0.988***
	(0.241)	(0.241)	(0.241)
big_event	-0.301	-0.362	-0.341
	(0.891)	(0.892)	(0.892)
child_n	0.731**	0.728**	0.726**
	(0.312)	(0.312)	(0.312)
old_n	-0.398	-0.431	-0.425
	(0.627)	(0.627)	(0.627)
urban18	2.817***	2.820***	2.780***
	(0.705)	(0.705)	(0.705)
future2		-0.838***	-0.571*
		(0.295)	(0.318)
Constant	28.88***	27.23***	23.56***
	(4.555)	(5.189)	(5.434)
Observations	3,961	3,961	3,961
R-squared	0.210	0.209	0.210
Standard errors in parentheses			
*** p<0.01, ** p<0.05, * p<0.1			

Figure 12 Results for Model 4

As mentioned in the descriptive statistics in Data Description, the majority of the consumers reported a mental state (life satisfaction and confidence in the future) above 3/5 (exceeding the turning point), which suggests China in 2018 is in the latter scenario of society, and the satisfaction and confidence of consumers will surely boost industries that sells non-essential products/services like culture industries, entertainment etc.

Implications for the Post-Epidemic Businesses

To put forward better suggestions for post-epidemic society, I used the newly released individual data from CFPS 2020. Since only the individual data are available, I conducted a T-test on values of the two main independent variables from the two years.

Surprisingly, the results (Figure 13) show that consumers in 2020 are significantly more confident in the future than those in 2018, and no significant difference in their life satisfaction. So, opposite to the hypothesis (4), people after the pandemic have even better mental state. A possible explanation is that China actually got the COVID under control just within three months after it broke out in China, and almost all business operations went back

to normal after April, 2020, while in the same time, pandemic in other countries were still severe. At the time of the survey, Chinese could be encouraged by the society's quick recovery from the COVID and therefore showed more confidence.

Two-sample t test with equal variances

Variables	G1(2018)	Mean1	G2(2020)	Mean2	MeanDiff	p-Value
life_sat	30270	4.011	22704	4.001	0.010	0.236
future_con	30238	4.125	22664	4.141	-0.016	0.050*

Figure 13 T-test of the Two Main Variables from Year 2018 and Year 2020

This points out a possible direction for start-up businesses and the overall economy. To maintain the development, the economy should pay more attention on how to entertain the consumers to meet the consumers' demand. The consumers are no longer looking for things with just the basic functions, but are looking for "more fun". Under the situation of pandemic, the online businesses have undoubtedly been in a favorable position thanks to their nature of the combination of functions and entertainment.

Conclusion

Concluding Remarks

To sum up, this paper found that first, consumers' life satisfaction and confidence in the future are both in a quadratic relationship with the percentage of household expenditure on non-essential good, so as consumers mental state (life satisfaction and confidence) increases, the percentage of their spending on non-essential goods out of total spending first decreases and then increases; second, consumers in 2020 are significantly more confident in the future than those in 2018, suggesting rising demand in the non-essential goods and services.

Limitations and Future Steps

There are at least three limitations in this study.

First, the limitation of the data. The data for the dependent and independent variables are not the perfect indicators for this research topic. I used the household economic data to measure the consumption pattern, the perfect measurement of which should be individual economic data.

Second, the limitation of the data processing. I didn't used data from different periods. The utility in Bentham's model is actually the utility across different periods, and the complete model should use panel data and include the lagged variables. And therefore, the results of this paper cannot imply a constant relationship between variables in different times.

Third, the limitation of the model. There is great potential of endogeneity caused by omitting of important variables and reciprocal causation.

Future steps in this study are: first, panel data should be used so that lagged variables could be included; second, more variables including instrumental variables should be found in an attempt to reduce the issue of endogeneity.

Data Appendix

Variables

Variable Name	Description
fid18	Family ID 2018
provcd18	Province ID
urban18	Urban area (Census Bureau's definition)
pid	Person ID
finc	Family's total income in past 12 months (yuan)
food	Expenditure on food in past 12 months
dineout	Expenditure on dining out in past 12 months
total_exp	Total expenditure in past 12 months
eatathome	Expenditure on eating at home in past 12 months
clothing	Expenditure on clothing in past 12 months
entertainment	Expenditure on education/entertainment
travel	Travel expenditures (yuan)
housing_trans	Expenditure on housing and transportation in past 12 months
durable	Expenditure on durable goods in past 12 months
edu_exp	Expenditure on education/trainings in past 12 months (yuan)
med	Expenditure on medical treatment in past 12 months (yuan)
percare_insurance	Expenditure on personal care and insurance in past 12 months
help_out	Expenditure on giving out help in past 12 months
otherexp	Other expenditure in past 12 months (yuan)
savings	Total amount of cash and deposits held by family now (yuan)
big_event	Whether had banquets and ceremonies in past 12 months?
noness_perc	Percentage of expenditure on non-essential goods out of total annual expenditure
ent_perc	Percentage of entertainment expenditure
male	The gender of the respondent
age	Age
edu	The education level of the respondent
married	Whether the respondent is married
attitude	Respondent's average attitude towards life. A bigger the number means more posit
life_sat	Rating of your satisfaction with your life
future_con	Rating of your confidence in your future
familysize18	Number of Family members (defined by T1)
child_n	The number of children in the household
old_n	The number of old people in the household

Table 1 Full Variable List

Data Cleaning

This section is a synthesis of the comments in the do file. For more details, see “Final Project – Yuxin Gong 101259039.do”

Data Cleaning of “famconfi” file

This is the only file that contains the variable "family size". I kept the family size information

for each family ID from this dataset. The codebook result shows there are 58,504 unique observations in this dataset and no missing value for pid variable. I kept one observation for one household and kept the variable “familsize2018”.

However, the collapse(sum) result shows the total number of people in all surveyed households is 53,121, which is less than the num of total observations. The reason is unknown because no information in the data documentation recorded how familysize18 variable was collected. To ensure the best accuracy, I decide to drop those families whose recorded familysize18 is not the same as the actual recorded number of family members.

I generated the “familysize” file for later merge.

Data Cleaning of “person” file

The "person" file includes the variables used to generate the main x variables and some of the control variables. In this section, I use this dataset to generate the main x variables (mental health, life satisfaction, and confidence in the future) and control variables (demographics, the number of children, and the number of old people).

(i) To generate the variable for the number of children/old people in a household. Since the children's answers are in a different dataset, to make sure the accuracy of the number of the children, I will append the childproxy dataset to the person dataset. Before I append, I made sure the variables needed for this process have the same names (i.e. the variables fid18 and age), so that I don't lose data. After checking on the data on age, I found that in this survey, for many questions with numerical answers, the researchers use -8 representing "missing”, which will disrupt the later-on analyses. So I always checked on those when cleaning numeric variables and made sure they’re all replaced. I generated the “child_old_n” file for later merge.

(ii) Next, I generate the variables “attitude toward life”. Generate "attitude toward life" variable using variables qm101m - qm110m: For these variables, 1-strongly disagree, 2-disagree, 3-agree, 4-strongly agree, 5-neutral. I need to re-order them from 1-5 respectively representing 1-strongly disagree, 2-disagree, 3-neutral, 4-agree, 5-strongly agree, so that the bigger the ultimate number of mental health, the better the mental health. Except for qm103m and qm105m that basically measure how bad the respondent is feeling, the others measure how good the respondent is feeling. So I choose to reverse the response of qm103m and qm105m. I finally generated the variable “attitude” using the average of qm101m - qm110m. Summarize “attitude” and the results show that the mean is 3.44, only 1806 observations. This could lead to a sharp decrease in the number of observations in later-on regression analysis because of the next steps of merging.

(iii) Next, I generate the variables "life satisfaction", and "confidence in future". For the original data of life_sat and future_con, there are three indicating values: "Not applicable" (-8) , unknown (-2) and refuse (-1), which should be treated as missing values so I make sure I replace them. As of now, I have all the main independent variables ready. These are saved in the file “main_inde”.

(iv) Finally, I keep the all independent variables as well as the control variables in this dataset. For control variables married, edu, male, there are three indicating values: "Not applicable" (-8), unknown (-2) and refuse (-1), which should be treated as missing values so I make sure I replace them. The clean file is saved as “person_clean”.

Data Cleaning of “famecoon” file

Now I generate the dependent variables. I made sure all missing values are replaced with "." (since all values here are numeric) and made sure the variables that I later merge on have the same names. I generated annual expenditure using the variables indicating monthly expenditure and then used the annual expenditure to calculate the percentage spent on non-essential goods. The clean file is saved as "famecon_clean".

Data Preparation

Merge Filesz

I use "famecon_clean" as the master file and merge on fid18 and pid using "person_clean", and merge on fid18 using "familysize" and "child_old_n". The clean file is saved as "cfps18_clean" file.

Data Preparation for CFPS 2020 Individual Data

The preparation is similar to that in the "data cleaning of 'person' file".

(3458 words, 7 Pages excluding figures and table)

References

Adolfo Crosta, et al. (2021). Plos One. Psychological factors and consumer behavior during the COVID-19 pandemic

Kristina M. Durante, et al. (2016). American Marketing Association. The Effect of Stress on Consumer Saving and Spending

Samuel Lins, et al. (2021). Journal of Psychiatric Research. Anxiety, depression, and stress: Can mental health variables predict panic buying?