Consumption in China: Household Consumption Behavior of Rich and Poor

Zi Yuan Yang Pradhayini Ganeshanathan *

April 6th, 2016

Abstract

This paper provides a model of consumption function for the Chinese economy for the year 2000. The main objectives of this paper is to investigate how consumption is determined by income, estimate the marginal propensity to consume (MPC) for the rich and the poor people, investigate whether there are significant differences in the consumption patterns of the rich and the poor and finally analyze how consumption behavior differs across the six Chinese provinces of Jilin, Shandong, Hubei, Guangdong, Sichuan and Shaanxi. We predominately use an ordinary least squares with approach as our conceptual frame work. We found that MPC is around 48% for the rich people and 60% for the poor people, and that depending on where you live, your consumption levels are very different.

*We would like to thank professor Gordon Anderson for putting up with us and for guiding us in the right direction when we were lost

1 Introduction

Since the 1930's, a considerable amount of econometric analysis has been carried out in deriving the consumption function. The concept is believed to have been introduced into macroeconomics by John Maynard Keynes in 1936 with the derivation of the consumption function which traces the simple relationship between income and consumption. The popular Keynesian Absolute Income Hypothesis states that, as the disposable income increases, the level of consumption also increases, though less than proportionately. As income increases, consumers split their income between consumption and savings (C + S = Y). Further, according to the Keynesians, the marginal propensity to consume (MPC), the change in consumption expenditure due to an $1 increase in income, is positive and is less than 1.

The purpose of this paper is to test how strong the relationship between consumption and income is, in the context of the Chinese economy. We further provide definitions for the rich and the poor and investigate whether there are any significant differences in the consumption behavior of the rich and the poor. We also attempt to find differences in consumption patterns, if any, across the six Chinese provinces (Shandong, Hubei, Guangdong, Sichuan, Shaanxi and Jilin). Through the analysis of the consumption function, we gain a good understanding of the level of savings, which helps the Chinese economy grow by accumulating wealth and expanding their production capacity. Further, the consumption function is indicative of the living standards of the Chinese population. We are motivated to find how the typical Chinese household[1] fare in terms of consumption and standard of living.

[1]We choose the typical Chinese household to be a family of 3 with 1 head of house, 1 spouse, and one child because of the one child policy.

2 Data

We have data on three coastal (Shandong, Hubei and Guangdong) and three interior (Sichuan, Shaanxi and Jilin) provinces of China for the year 2000, collected through the Urban Household survey (NHS). The Urban Household Survey promoted by the National Bureau of statistics (NBS) is a national survey conducted to collect individual and household data using questionnaires and the sampling frame designed to investigate the phenomena of urban unemployment and poverty in China.

An interesting feature of the data set is that all six provinces selected have different levels of regional GDP in the year 1999: Shanxi (GDP is 1506.78) (source: China's statistics yearbook), Sichuan (GDP equal to 3711.61 yuan), Hubei (GDP equal to 3857.99 yuan), Jilin (GDP equal to 1669.56 yuan), Shandong (GDP equal to 7662.10 yuan), Guandong (GDP equal to 8464.31 yuan). (The GDP is given in 100 million yuan). Furthermore, among the six provinces, two are ranked as poor economies (Jilin and Shanxi), two as upper-middle (Sichuan and Hubei) and two rich economies(Guandong and Shandong).

In addition to information on demographic characteristics such as age, gender and education levels, we have information on the basic conditions such as living expenditures for consumption, purchase of major commodities, durable consumer goods owned at the end of the year, housing conditions, and cash income and expenditures. We have detailed information on the sources of income and consumer expenditures.

We decided to do our analysis on the cross section data available for the year 2000. We are concerned about the current period's income and how that affects current consumption. Our data set is collected by the government and thus should be subjected to good statistical sampling methods. So, we assume our data set is independent and identically distributed.

2.1 Data problems

There were some minor issues with our cross sectional data set. To begin with, the data set contained categorical values that were not specified on the attached variables list, which describes the data set. For example, the categorical variable of *sex* had 1 representing male and 2 representing female. However, on the data set, it was not uncommon to find the entry value 0. There were multiple cases of misspecification of categorical values throughout the data set that had non-sensible values such as the value ten when 10 is not specified. We had no idea what it described. Unfortunately, no matter how good the econometrician is, there is no salvaging bad data. We got rid of those observations. In total, the observations that we got rid of account for less than 0.01 percent of our data. So, we were relatively comfortable.

3 Method and Model

We began our journey by considering how consumption patterns differ among the rich and the poor in China for the year 2000. We used a classical ordinary least squares regression model as our conceptual framework in attempting to answer our question.

$$Consumption_i = \beta_0 + \beta_1(Income_i) + \gamma_i(\mathbf{X}) + \psi_i(\mathbf{D}) + \varepsilon_i \tag{1}$$

Where $\mathbf{X}$ contains variables from our stepwise regression results (Appendix A Table 3), and $\mathbf{D}$ contains the set of dummy variables (Appendix A Table 2), which includes *TotIncFA* and provincial dummy variables. The $TotIncFA$ is equal to total income for those who generated income from assets, and zero to those without any income from assets. The dummy variable set also contains provinces, which we will discuss later.

We wanted to compare apples to apples. Thus, we did not want to compare consumption for a household of size two with a household of size of 7. So, we restricted the data set so that the household is 3 and the family of 3 contains 1 *head of house*, 1 *spouse*, and 1 *child*[2]. Further, we only looked at individuals with 4% or more of their income coming from assets. In other words, we restricted the ratio of income from assets to total income to 0.04. We believe this number is appropriate, because it leaves us with a decent sample size for both groups. It also helps us drop those people who only have a dollar or two as income from assets, which is too economically insignificant to be included as people with income from assets, and therefore in the rich condition. After the restriction of 4%, the household with the lowest income from assets has income from assets of 112 yuan per month. This way, we believe that on average the household consumption levels could be compared between the two groups of rich and the poor.

3.1 Defining rich and poor

In order to compare the rich and the poor people, we first had to provide definitions for who is rich and who is poor in our data set.

If households had income from assets, then we say they are in the rich condition. If households had no income from assets, then we say they are in the poor condition.

When households had income from assets, we perceived them to be in a rich condition, and when households had no income from asset, we considered them to be poor. Firstly, individuals who had income from asset are the ones who possibly saved in the past. With no direct observations on savings, we assumed that those who saved considered their savings to be a safety net, which increased their consumption confidence, and as a result, we expected consumption levels to be high. Secondly, incomes from assets provided the individuals with extra income, and therefore,

[2] We went into the *personal* data set and screened all 13317 observations to make sure no family size of 3 is any other than *head of house, spouse, and child.*

further savings became possible. Savings are investments, and investments make us richer. So, it is justifiable to consider those with asset income to be in a richer condition. So, there was evidence to believe that households with income from assets are in the richer condition.

To compare, the mean for the rich condition was 8266 yuan per month with 513 observations and the poor condition had 5509 yuan per month with 1192 observations.

3.2 Outliers

We looked at the distribution of incomes with household size equal to 3 and found that there was an outlier on the lower end, which we chose to discard. The household with the outlier had an income of 20 yuan per month whereas, the second lowest income household was an income of 800 yuan per month. We feel justified in getting rid of that observation so as to not pull our ordinary least squares (OLS) line down and bias our results.

3.3 Curvilinear Tests

From economic reasoning, there is reason to believe that the consumption function is curvilinear that is, it has the following model.

$$Consumption_i = \beta_0 + \beta_1(income_i) + \beta_2(income_i^2) + \varepsilon_i \tag{2}$$

We believe this because when you have low to no income, you still need to consume. As your income increases, your consumption initially increases at a slower rate. When you get richer, your consumption tends to increase faster. It is easy to think of cases where a doubling in income does not mean a doubling in consumption. We use the above regression to do a t-test on the variable β_2 to see if the parameter estimated is statistically significant[3]. (Appendix A Table 4)

[3] We use α=0.05 level for all of our tests throughout this paper

3.4 Stepwise Regression Method

We want to choose our explanatory variables for our model if and only if the the β_i that is being estimated is significant, which is to say the null hypothesis

$$H_0 : \beta_1 = 0$$

fails to be accepted in favor of the alternative by a t-test. We do this because we recognize that adding more explanatory variables will always increase explanatory power but will also increase the variance of β_i. We want to balance this trade off. So, we are looking to ignore those variables from our model that doesn't have a t-statistic significant at the $\alpha = 0.05$ level. We also made sure that there was no perfect col-linearity in our model when we did the stepwise method.

The stepwise selection technique is a modification of forward regression technique where one variable is added into the regression equation and then tested to see if the newly added variable changes the significance of the β_i in the equation. The step wise regression technique differs from the forward regression technique in that, variables in the model do not necessarily stay there as in the forward regression technique. After a variable is added, the stepwise method looks at all the variables already included in the model and deletes any variable that does not produce an F statistic significant at the $\alpha = 0.05$ level. Only after this check is made and the necessary deletions are done can another variable be added to the model. The stepwise process ends when none of the variables outside the model has an F statistic significant at the $\alpha = 0.05$ level and every variable in the model is significant at the $\alpha = 0.05$ level. (Appendix A Table 3)

3.5 Provincial Dummy Variables Set Up

We wanted to see if consumption deferred across provinces. In order to do so, we set up dummy variables for the provinces. The dummy variables are set up as the province equaled 1 if true and

0 otherwise. Our simplified regression model becomes

$$\mu = \beta_0 + \gamma(\mathbf{Z}) + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

Where μ is the grand mean for consumption given all Xs. Here we denote all Xs to be dummy variable, and X_1 to be Jilin, X_2 to be Shandong, X_3 to be Hubei, X_4 to be Guangdong, X_5 to be Sichuan, and **Z** to be the explanatory variables in our model.. We made Shanxi our reference category, so it is not in the model as it will cause co-linearity. (See Table 1 in appendix)

Let μ be the grand mean then from Table 1, we can see

$$\mu = \frac{1}{6}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6)$$

$$\mu = \beta_0 + \gamma(\mathbf{Z})$$

We can clearly see from the simplified model above that $\beta_0 + \gamma(\mathbf{Z})$ is the grand mean. All the β_i where $i \epsilon$ {1,2,3,4,5,6} is the difference between their μ_i and the grand mean.

3.6 White's Test for Homoscedasticity

We used the White's test for homoscedasticity to test the if the variances of the unobservable differs across $X_i s$. The hypothesis for White's test is

$$H_0 : Var(\varepsilon_i | X_i) = \sigma_u^2$$

4 Empirical Results

Our stepwise results gave us *number of rooms*, *income*, and *totalExpenditure* for the rich condition regression. We found it weird, at first, that *number of rooms* would be included in our model. We knew from that, that our model is misspecified in some way. Upon further investigation, we realized that the *number of rooms* is a proxy for household size [2]. We solved the problem by reweighting the explanatory and response variables by square root of household size. Rerunning the stepwise regression, we found that the *number of rooms* became statistically insignificant for our model.

The stepwise results for the poor condition has *income*, *totaExpenditure*, and *type of ownership*. We can tell that there is something wrong with our model as *type of ownership* is clearly a proxy for something else that we are omitting. We are not sure what it is proxying. So, we left it in the model without interpreting it. (See Appendix A Table 3 for output)

For the rich condition, our model fails the curvilinear test (Appendix A Table 4). However for the poor condition, our model suggests evidence for curvilinearity.

For our regression model with provincial dummy variables, we reject all β_i are the same with overall F-test statistic of 121.67. We also wwanted to know how provinces affect consumption (Appendix A Table 2). We concluded that only Guangdong had a statistically significant affect on consumption in our model.

Our model fails to accept the White's test for homoskedascity, which suggest that the variance of the unobservable are changing across the explanatory variables in our model.

5 Discussion

Our main goal in this paper was to 1) estimate the marginal propensity to consume for rich and poor people, 2) find differences in consumption levels for rich and poor people, and 3) estimate the impact of our 6 provinces on consumption.

Let's start with the first one. We used the methods described above to section our data and model our consumption function. The first function is for the rich condition and the second for the poor condition.

$$C_i = \beta_0 + \beta_1(IncWAsset) + \beta_2(totalExp) + \varepsilon_i$$

$$C_i = \beta_0 + \beta_1(IncWOAsset) + \beta_2(IncWOAsset^2) + \beta_3(totalexp) + \beta_4(ownership) + \varepsilon_i$$

Where C_i is consumption weighted with the square root of household size of 3, $IncWAsset$ total income of the rich condition with income from assets weighted with the square root of household size of 3, $IncWOAsset$ total income of the poor condition without income from assets weighted witht he square root of household size, $totalExp$ is total expenditure, and $ownership$ is the type of ownership of the household's residence. From our model, the marginal propensity to consume (MPC) for the rich condition is about 48% and for the poor condition is about 60% (Appendix B). From our analysis, the poor condition spends a significant amount more than the rich condition for every dollar earned. This is not surprising considering that people don't buy two times the amount of stuff just because their income has increased 2 times.

Secondly, we wanted to find whether or not there's a difference between consumption levels for rich and poor.

$$C_i = \beta_0 + \beta_1(Income) + \beta_2(TotIncFA) + \phi(\mathbf{D}) + \varepsilon_i$$

Where $Income$ is total income, and $TotIncFA$ is equal to total income for people when generate income from assets and 0 if otherwise. We wanted to test if β_2 is different than 0. As it turns out, we fail to accept the null hypothesis that β_2 is 0. So, we conclude that there is a difference between the rich condition and poor condition in terms of consumption.

Thirdly, we want to estimate the impacts of provinces on consumption for both condition people. We know that different provinces have different levels of income, inflation, macro and micro economic environments. We did an overall F-test with just the dummy variables for provinces and saw that the overall F-stat is 121.67, which suggest that there is a difference in consumption across provinces. From table 2, we can conclude that people who live in Guangdong have a statistically significant impact of around 490 yuan more than the national average. We should mention that our model is biased and inconsistent (more on this below). So, readers should take our parameter estimates with a grain of salt. With that being said to make the model more realistic, we should also attempt to explain the variation in consumption across provinces by inflation adjusting our models to take into account differences in inflation across provinces.

5.1 Confounding Variables and Inflated Type I Error

In our model

$$Consumption_i = \beta_0 + \beta_1(Income_i) + \beta_2(Numberofrooms) + \gamma\mathbf{D} + \varepsilon_i$$

where $numberofrooms$ is a proxy for household size and **D** is a catch all term for all the dummy variables in our model. We could think of many confounding variables for this model. For example, maybe people who like to use their brains more vigorously make more money but also consumes more food to sustain their brain power. So, *affinity to use brain* positively correlates with both the response variable and explanatory variable. So Keynes may not be right to suggest income affecting consumption. We need to be careful in interpreting causality.

However in our model, we also realized that there is an inflated type one error problem. The reason is that *number of rooms*, which proxies for household size is correlated with *income* and *consumption*. It is also measured with error[4]. This combination causes an inflated type I error rate that gets worst as the sample size increases[5][3] and, this problem occurs often in observational studies.

[4]We assume all observational studies are measured with error to some degree as our null hypothesis. It's better to err on the side of caution.

[5]This is shown by Brunner and Austin in a monte carlo study

5.1.1 Omitted Variables Bias

We recognize that there is omitted variables problem in our model, that is, our model may be missing explanatory variables that are highly correlated with our existing covariates. Omitted variables will bias our model so as to make our model biased and inconsistent. Since it is impossible to measure everything, there is most likely unobservables that's important to our model, but we cannot include it in our model.

5.1.2 Zero Conditional Mean Assumptions

Recall that zero conditional mean state that

$$E(\varepsilon_i|X_i) = 0$$

Since there is omitted variables bias, simultaneity, and measurement error, the zero conditional mean assumption does not hold. Since zero conditional mean is used to derive the fact that $\beta_i s$ are unbiased and consistent, our estimates are therefore biased and inconsistent.

To make OLS unbiased and consistent, we would have to first add in all the omitted variables and or proxies to solve the omitted variables problem.

5.2 Generalized Least Squares

We found that with White's test (Appendix A Table 5) that we reject that our model is homoskedastic. This leads to a problem for inference. We are fully aware that giving parameter estimates are not sufficient. If we are to give a parameter estimator, we better also give a confidence interval for those parameter estimates. The idea with generalized least squares is to transform a heteroskedastic model to a homoskedastic model to do inference. We concluded that our model is not homoskedastic. However, we cannot transform our model back from heteroskedastic to homoskedastic using generalized least squares because we do not have zero conditional mean.

$$var(\varepsilon_i|X_i) = E(\varepsilon_i^2|X_i) - E(\varepsilon_i|X_i)^2$$

The method of feasible generalized least squares requires zero conditional mean so that $E(\varepsilon_i|X_i)^2$=0. This way we can model the variance as a constant times a function. For example,

$$var(\varepsilon_i|X_i) = E(\varepsilon_i^2|X_i) = \sigma^2 e^{\delta_0+\delta_1 X_1+\ldots\delta_k Xk}$$

We assume that the variance is transformed by a function of the form $e^{\delta_0+\delta_1 X_1+\ldots\delta_k Xk}$. However, we do not have zero conditional mean. Therefore, we cannot use the methods of feasible generalized least squares.

Since we do not know the variance of β_i with heteroskedascity, it does not make sense to do inference.

5.3 Simultaneity Problems

We can see that the structural model

$$Consumption_i = \beta_0 + \beta_1(Income_i) + \varepsilon_1$$

has a simultaneous problem. Namely the fact that consumption and income are determined together. If we, for a second, pretend to think that the equation above as causal, which it is not. In which case, the above equation says income effects consumption levels. This is not hard to imagine. As you make more money, you will have the ability to buy more stuff. On the other hand, we can think of

$$Income_i = \beta_0 + \beta_1(Consumption_i) + \varepsilon_2$$

We can also think of reasons why consumption levels would affect income. It is not hard to think of the case of real estate agents. Real estate agents usually have very nice cars to drive their clients in. If you talk to them, they will say that it is a must to have the state of the art toys to impress prospective clients with, to show them that the real estate firm or broker is successful and trustworthy. In this case, the consumption would affect income.

To solve the simultaneous problem we need instrumental variables to identify the equations. Specifically, we need an instrument in the first equation that is correlated with *consumption* in the second equation but not correlated with ε_2, and we need an instrument in the second equation that is correlated with *income* in the first equation but not with ε_1. The instruments also have to be different. However unfortunately, we were unable to find valid instruments at the time of writing this paper.

5.4 Errors in Variables

It is very likely from observational studies and especially large scale studies that require multiple teams of people to go out and collect data, that there is measurement error of one form or another. It is also highly possible that income measures are measured with error because people tend to have incentives to over report or under report their incomes. For these reasons, we believe our data set will have measurement error. Unfortunately, our model isn't Gauss Markov or more precisely $E(\varepsilon_i|X_i) \neq 0$. We cannot conclude if the estimated β_i would be upward biased or downward biased on expectation based on the theoretical sign of the β_i. (Appendix B)

6 Conclusion

We started our journey by asking ourselves how rich and poor households consume differently. We used econometric tools to generate estimators for key indicators such as MPC for the rich and poor groups, and tested consumption differences across provinces. Our econometric evidence illustrate that, as the level of income increases, consumption also increases, which is to be expected. We faced many problems such as heteroskedascity in our error, which prohibited us from giving a confidence interval estimate. We caution our readers to use our parameter estimates in making very important decisions and to remember to always take statistical models, by anyone, with a grain of salt. After all, all statistical models are wrong.

We hope that our research has encouraged others who are also studying consumption in China. More research is still needed in forecasting consumption. We hope our research can contribute to future forecasting of consumption in China.

References

[1] Alice. O. 2013 "An Estimation of the Consumption Function for Kenya Using Keynes' Absolute Income Hypothesis for the Period 1992-2011". Journal Of Emerging Trends In Economics And Management Sciences, 4(1), 103-105.

[2] Anderson, G. (2012). "The progress and characteristics of poor, middle and rich classes in urban China: Results from partial definition of class membership" JEL classification: C14; I32; O1.

[3] Brunner, J., P. Austin. (2009). "Inflation of type I error rate in multiple regression when independent variables are measured with error", The Canadian Journal of Statistics vol 37 No.1 p33-46.

[4] China Overview. Web. 06 Apr. 2016. http://www.worldbank.org/en/country/china/overview

[5] Fang, C., E. Wailes and Cramer G. (1998). "China's Rural and Urban Household data: Collection. Availability, and problems", CARD Working Paper 98-WP 202.

Appendix A: Tables and Figures

Table 1: Dummy Variables Using Effect Coding

City	X_1	X_2	X_3	X_4	X_5	μ=β_0+$\gamma(\mathbf{Z})$+β_1X_1+β_2X_2+β_3X_3+β_4X_4+β_5X_5
Jilin	1	0	0	0	0	μ_1=β_0+$\gamma(\mathbf{Z})$+β_1
Shandong	0	1	0	0	0	μ_2=β_0+$\gamma(\mathbf{Z})$+β_2
Hubei	0	0	1	0	0	μ_3=β_0+$\gamma(\mathbf{Z})$+β_3
Guangdong	0	0	0	1	0	μ_4=β_0+$\gamma(\mathbf{Z})$+β_4
Sichuan	0	0	0	0	1	μ_5=β_0+$\gamma(\mathbf{Z})$+β_5
Shanxi (reference)	-1	-1	-1	-1	-1	μ_6=β_0+$\gamma(\mathbf{Z})$-β_1-β_2-β_3-β_4-β_5

Here we denote all Xs to be dummy variable, and X_1 to be Jilin, X_2 to be Shandong, X_3 to be Hubei, X_4 to be Guangdong, and X_5 to be Sichuan. $\mathbf{Z}$ to be a set of explanatory variables in our model.

Table 2: How Provinces affect Consumption

Variable	Parameter Estimate	Standard Error	t-value	p-value
Intercept	347.16639	164.02260	2.12	0.0344
Income	0.77288	0.02241	34.48	<.0001
Jilin	-130.61627	185.96851	-0.70	0.4826
Shandong	-273.32407	192.06429	-1.42	0.1549
Hubei	102.36017	186.97144	0.55	0.5841
Guangdong	489.92163	240.65529	2.04	0.0419
Sichuan	-95.34255	184.87264	-0.52	0.6061
Overall F-test			F-stat=370.45	<.0001

* All significance at α=0.05 level.

Table 3: Stepwise Output

Rich Condition

Variable	F-value	p-value
Intercept	0.66	0.4175
Number of rooms	7.25	0.0072
Income	363.55	<.0001
Total expenditure	1641.53	<.0001

* All significance at α=0.05 level.

Poor Condition

Variable	F-value	p-value
Intercept	69.31	<.0001
Income	10.49	0.0012
Total expenditure	586.51	<.0001
Ownership	9.9	0.0017

* All significance at α=0.05 level. Where *Ownership* is the type of ownership of the residence of the household.

Table 4: Curvilinear Test

Model: $C_i = \beta_0 + \beta_1(IncWAsset_i) + \beta_2(IncWAsset_i^2) + \beta_3(totalExp_i) + \varepsilon_i$

Variable	Parameter Estimate	t-value	p-value
Intercept	226.09	1	0.3174
IncWAsset	0.25729	3.59	.0004
IncWAsset Squared	0.0000044	1.28	0.2014
totalExp	0.30495	25.24	<.0001

* All significance at α=0.05 level. $IncWAsset$ is total income of people with income from assets (ie the rich condition). $totalExp$ is total expenditures of the household.

Model: $C_i = \beta_0 + \beta_1(IncWOAsset_i) + \beta_2(IncWOAsset_i^2) + \beta_3(totalexp_i) + \beta_4(ownership_i) + \varepsilon_i$

Variable	Parameter Estimate	t-value	p-value
Intercept	308.4326	3.6	0.0003
IncWOAsset	0.53928	16.26	<.0001
IncWOAsset Squared	-0.0000155	-6.26	<.0001
totalExp	0.19157	24.71	<.0001
Ownership	-81.57928	-3.69	0.0002

* All significance at α=0.05 level. $IncWOAsset$ is total income for people without assets (ie the poor condition). $totalexp$ is total expenditures of the household. $Ownership$ is the type of ownership of the residence of the household.

Table 5: White's test for homoskedascity Output

Chi-Square	p-value
45.77	<.0001

Table 6: Consumption Function Parameter Estimates

Model: $C_i = \beta_0 + \beta_1(IncWAsset_i) + \beta_2(totalExp_i) + \varepsilon_i$

Variable	Parameter Estimate	Standard Error	t-value	p-value
Intercept	5.28542	145.95330	0.04	0.9711
IncWAsset	0.33938	0.03193	10.63	<.0001
totalExp	0.30384	0.01206	25.20	<.0001
Overall F-test			F-stat=964.26	<.0001

* All significance at α=0.05 level. $IncWAsset$ is total income of people with income from assets (ie the rich condition). $totalExp$ is total expenditures of the household.

Model: $C_i = \beta_0 + \beta_1(IncWOAsset_i) + \beta_2(IncWOAsset_i^2) + \beta_3(totalexp_i) + \beta_4(ownership_i) + \varepsilon_i$

Variable	Parameter Estimate	Standard Error	t-value	p-value
Intercept	308.4326	85.62627	3.6	0.0003
IncWOAsset	0.53928	0.03317	16.26	<.0001
IncWOAsset Squared	-0.0000155	0.00000248	-6.26	<.0001
totalExp	0.19157	0.00775	24.71	<.0001
Ownership	-81.57928	22.13295	-3.69	0.0002
Overall F-test			F-stat=792.29	<.0001

* All significance at α=0.05 level. $IncWOAsset$ is total income for people without assets (ie the poor condition). $totalexp$ is total expenditures of the household. $Ownership$ is the type of ownership of the residence of the household.

Appendix B: Derivations

Marginal Propensity to Consume

Getting our marginal propensity to consume was a bit tricky. For the rich condition, our model was

$$Consumption_i = \beta_0 + \beta_1(IncWAsset_i) + \beta_2(totalExp_i) + \varepsilon_i$$

Note that $totalExp = \phi_0 + \phi_1(Consumption_i) + \phi_2(Tax) + \epsilon_i$

$$\begin{aligned} Consumption_i &= \beta_0 + \beta_1(IncWAsset_i) + \beta_2(\phi_0 + \phi_1(Consumption_i) + \phi_2(Tax) + \epsilon_i) + \varepsilon_i \\ &= \beta_0 + \beta_1(IncWAsset_i) + \beta_2\phi_0 + \beta_2\phi_1(Consumption_i) + \beta_2\phi_2(Tax) + \beta_2\epsilon_i + \varepsilon_i \end{aligned}$$

$$(1 - \beta_2\phi_1)Consumption_i = \beta_0 + \beta_1(IncWAsset_i) + \beta_2\phi_0 + \beta_2\phi_2(Tax) + \beta_2\epsilon_i + \varepsilon_i$$

So we can see that MPC for the rich condition is estimated by something like $\frac{\beta_1}{(1-\beta_2\phi_1)}$. However for the poor condition, the model is quadratic.

$$Consumption_i = \beta_0 + \beta_1(IncWOAsset_i) + \beta_2(IncWOAsset_i^2) + \beta_3(totalexp_i) + \beta_4(ownership_i) + \varepsilon_i$$

We do the same for total expenditure that $totalExp = \phi_0 + \phi_1(Consumption_i) + \phi_2(Tax) + \epsilon_i$. We rescale every term by $\frac{\beta_j}{(1-\beta_3\phi_1)}$. The catch is that the derivative of the quadratic function with respect to $IncWOAsset$ is $\beta_1 + 2(\beta_2)(IncWOAsset)$. We used the sample average to estimate $IncWOAsset$. The sample average was 3180.87.

Using our best guesses from the above relationships. We estimate the MPC for the rich condition is around 48% and the poor condition to be around 60%.

Error In Variable Derivation in Matrix Form

Assuming *income* is measured with error, and let x be a vector of income and y be a vector of consumption.

$$x = x^* + \varepsilon$$

is measured with some error ε.

$$y = x\beta + u$$

$$y = x^*\beta + u + \varepsilon\beta$$

Let $v = u + \varepsilon\beta$

$$\begin{aligned}
E(\hat{\beta}) &= (x^{*'}x^*)^{-1}x^{*'}y \\
&= \beta + (x^{*'}x^*)^{-1}x^{*'}v \\
&= \beta + (x^{*'}x^*)^{-1}x^{*'}(u + \varepsilon\beta) \\
&= \beta + (x^{*'}x^*)^{-1}(x - \varepsilon)^{'}(u + \varepsilon\beta) \\
&= \beta + (x^{*'}x^*)^{-1}(x^{*'}u + \varepsilon^{'}u) + (x^{*'}x^*)^{-1}(x^{*'}\varepsilon\beta - \varepsilon^{'}\varepsilon\beta) \\
&= \beta + (x^{*'}x^*)^{-1}(x^{*'}u - \varepsilon^{'}\varepsilon\beta)
\end{aligned}$$

We get the above equation because $cov(\varepsilon, u) = 0$, and x is independent of ε. Since $E(u|x) \neq 0$ in our model, we cannot conclude if our estimated parameter will be upward biased or downward biased on expectation based on the theortical sign of β.