

Econ 490: Predictive Analytics

Final Project Guide: Model Simulation in R

Instructor: Dr. Rameses Armendariz

Author and Course Developer: Francis Fang¹

1 Project Introduction

In Predictive Analytics, you are learning how to employ models to make predictions, and one example of this learning process is your final project. For your final project, you are asked to calibrate Armendariz (2020) structural demand for dietary staples of a poor household and extrapolate some predictions using that model.

In this note, I explain to you how to calibrate that model and extrapolate those predictions using R. For the calibration process, we will use the method of moments to match the elasticity of the household demand for rice estimated by Jensen and Miller (2008) in Hunan, China. For the extrapolation of the predictions, we will do two types of extrapolations. The first type is a postdiction in which we will simulate the model and verify how well it replicates the inverted u-shape relationship between the Initial Staple Calorie Share (ISCS) and the demand elasticity that Jensen and Miller document with their randomized experiment. The second type are counterfactuals. The counterfactuals that we will extrapolate are: (1) the whole relationship between price and quantity demanded for the staple, (2) the household demand for total caloric intake, (3) the optimal consumption bundle of the household as the price of the staple changes, and (4) the optimal consumption bundle as income changes.

The model that we will calibrate is fully characterized by the following Data-Generating Process (DGP):

$$b = \frac{i[\beta p_m - p_b] + p_m p_b (1 - \beta) \bar{c}}{p_b(p_m - p_b)} + e$$

The variables in this model are the following:

- Endogenous variable: quantity demanded for the staple measured in calories (b).
- Exogenous variables:
 - Price of a calorie from the staple (p_b).
 - Price index of a calorie from all the other sources of food (p_m).
 - Income of the household (i).

¹ University of Illinois at Urbana-Champaign, email: haoyuf2@illinois.edu

- Quantity of caloric intake that makes the feeling of hunger become unbearable (\bar{c}).
- The importance of satiating hunger in the preferences of the household (β).
- e is a random variable. We ignore its probability distribution. All we know is that $E(e) = 0$.

2 Model Estimation Process

Step 1: Transform the model to match it to the data

From Table 1 in Jensen and Miller (2008, AER), we know that

- average caloric intake per household before the experimental treatment took place was $1,805(2.85) = 5,144$.
- the caloric share of the staple before the experimental treatment took place was 0.64.

We are going to use these numbers to calculate the income of the household, $i = p_b b + p_m m$, according to the economic model proposed in this project. Specifically, we are going to use these numbers to substitute for b and m . After the substitution, we find that the income of the household before the experimental treatment took place is

$$i = p_b 1,805(2.85)0.64 + p_m 1,805(2.85)0.36 = 1,805(2.85)[0.64p_b + 0.36p_m]$$

Now, we can substitute for income in the demand:

$$b = \frac{5,144.25(0.64p_b + 0.36p_m)[\beta p_m - p_b] + p_m p_b (1 - \beta) \bar{c}}{p_b(p_m - p_b)}$$

Step 2: Choose the moments that the model is going to match

First, we are going to normalize the model in terms of the price of the staple before the experimental treatment was introduced. Thus, we are going to set the original price of the staple equal to 1 ($p_b = 1$). Second, we need to find values for the remaining variables in the last equation: p_m , β , \bar{c} . Thus, at minimum, we need three moments.

1. The average consumption of b before the experimental policy was introduced.

$$\frac{5,144.25(0.64 + 0.36p_m)[\beta p_m - 1] + p_m(1 - \beta)\bar{c}}{(p_m - 1)} = 2.85(1,805)0.64$$

2. The average consumption of b after its price has increased by 1%.

$$\frac{5,144.25(0.64 + 0.36p_m)[\beta p_m - 1.01] + p_m 1.01(1 - \beta)\bar{c}}{1.01(p_m - 1.01)} = 2.85(1,805)0.64(1 + 0.00235)$$

3. The average consumption of b after its price has decreased by 1%.

$$\frac{5,144.25(0.64 + 0.36p_m)[\beta p_m - 0.99] + p_m 0.99(1 - \beta)\bar{c}}{0.99(p_m - 0.99)} = 2.85(1,805)0.64(1 - 0.00235)$$

Notice that the price of the staple in the first moment is equal to 1 (because that is the price we normalized), in the second moment is equal to 1.01, and in the third moment is equal to 0.99.

The values in the right hand of the equations characterize the ATE that Jensen and Miller (2008, AER) document in Table 3. In particular, they document that the price elasticity of the staple is + 0.235, making it a Giffen good. 0.00235 is the price elasticity after it has been divided by 100.

Step3: Solve the system of equations defined in step 2.

The estimated values of the parameters are $\hat{\beta} = 0.0593$, $\bar{c} = 4,951.367$, $\hat{p}_m = 2.655$. Notice that the values that are approximated not accurate results. Thus, the estimated version of our model is the following:

$$b(p_b; p_m, i) = \frac{i[0.0593p_m - p_b] + p_m p_b (0.941)4,951.367}{p_b(p_m - p_b)}$$

Or, we can also present the demand model in terms of the price of the staple only:

$$b(p_b) = \frac{8,209.194[0.156 - p_b] + 12,370.272p_b}{p_b(2.655 - p_b)}$$

In this later estimated version of the model, I substituted for the income using this formula: $i = p_b b + p_m m$, where the value of b is 2.85(1,805)0.64, the value of m is 2.85(1,805)0.36, the value of p_m is 2.655, and the value of p_b is 1. When you substitute all these numbers, you find that **$i = 8,209.194$** . This is the average real income of a household in the experiment. This real income is measured in terms of staple calories. This is one of the main results of this estimation process. It provides an alternative technique to measure real income in terms of nutrition.

3 R Basics and Sample Codes

3 things you really need to know in R to conduct your final project are: importing raw data file, writing basic functions, and making the plots by the plot function. The following code are some sample codes that introduce you the R basics.

3.1 Basic Commands

1. Comment lines in R are denoted by #

```
# Make comments
```

2. Install and load packages

```
install.packages("libraryname") ; library(libraryname)
```

3. Read a data file(Make sure to put your datafile and coding file in same place)

```
data <- read.csv("datafilename.csv")
```

4. Make objects within dataframes accessible in R

```
attach(data)
```

5. Get help

```
?t.test
```

3.2 Basic Functions and Calculations

1. Generate a continuous sequence from a to b, with gaps between elements in c with seq

```
seq(a,b,c)
```

2. Assign a value to x by “<-”

```
x <- 888
```

3. Do calculations with assigned value

```
a<- 99
```

```
b <- 233
```

a+b
 a**b(which means a^b)
 a/b

4. Write a basic function by “<- function(a,b){ }”

```
f1 <- function(a,b){  
  a+b  
}
```

5. Use the written function f1

```
a <- 99  
b <- 233  
f1(a,b)  
or   f1(99,233)  
or   f1(a=99,b=233)
```

3.3 Plotting in R(plot)

Note: if you want to do ggplot, it's good! but we won't teach that in the lab

1. Use the Plot command

```
plot(y~x, data= ,main= "", xlab= "", ylab= "", xlim= c(a,b), ylim=c(e,f), type= "l", col=)
```

2. Add more lines

```
line(y~x, data=, lty=, col=)  
abline(a=, b=)
```

3. Add legend

```
legend(x, y, legend=c("line1 name ", "line2 name ", "line3 name"), col=c("line1 col", "line2 col", "line3 col"), lty=c(line1 type, line2 type, line3 type), cex=)
```

4 Testing the Model

Remember from previous part in Model Estimation Process, we have founded the approximately estimated version of model:

$$b(p_b: p_m, i) = \frac{i[0.0593p_m - p_b] + p_m p_b (0.941)4591.367}{p_b(p_m - p_b)}$$

Where the value of i for the average household in the experiment is $i= 8209.194$.

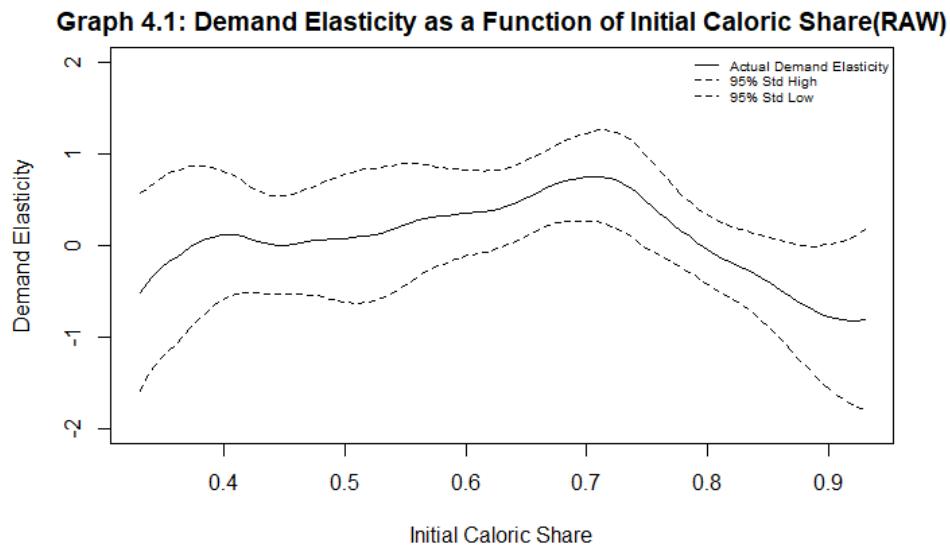
Now, we are going to simulate the model to compare its prediction with the data reported in Jensen and Miller(2008, AER).

From the given data, we can make Jensen and Miller's work to a graph showing demand elasticity as a function of initial Caloric Share by the such codes:

```
data_old <- read.csv("data_old.csv")

plot(y_i~x_i, type="l", main="Graph1: Demand Elasticity as a Function of Initial Caloric Share(RAW)", xlab="Initial Caloric Share", ylab="Demand Elasticity", data = data_old, ylim=c(-2, 2))
lines(std_hi_i~x_i, data=data_old, lty=2)
lines(std_low_i~x_i, data=data_old, lty=2)
legend(0.78, 2.14, bty="n", legend=c("Actual Demand Elasticity", "95% Std High", "95% Std Low"), col=c("black", "black", "black"), lty=c(1, 2, 2), cex=0.6)
```

The output graph would look like this(if you are not familiar with the R basic, please refer to previous sample codes in R or come to the office hour):



Our goal is to draw a calculated demand elasticity vs. Initial Caloric Share from our model to the graph to make comparisons. To simulate our model, we need to calculate model's predictions on our own, 4 parameters are important, they are:

b: represents the demand of bread, if budget i is fixed, it will correlate with price of meat and price of bread only.

m: represents the demand of meat

HSCS: represents the household share of bread

DE: represents the demand elasticity as the price change in bread(we will finally compare it with Jensen and Miller's work)

Then, following these steps to calculate above parameters:

0. Import the calculated parameters from the given model, remember the calculated $i = 8209.194$, $\bar{c} = 4531.367$, $\hat{\beta} = 0.0593$.

1. You will first create a sequence of p_m from 1.855 to 8.845 with gaps in 0.005 each, these can be wide and precise enough to cover the Initial Calories Share we want(from 0.3 to 1)

2. Calculate b_0 , b_1 by the formula we get before, b_0 , b_1 represent the demand of bread when price of bread is the exact value before the experiment($p_b=1$) and when raise the price at 10% level($p_b=1.1$). Remember the formula:

$$b(p_b: p_m, i) = \frac{i[0.0593p_m - p_b] + p_m p_b (0.941)4591.367}{p_b(p_m - p_b)}$$

Where the value of i for the average household in the experiment is $i= 8209.194$.

3.Calculate the demand of meat **before the experiment(m)** by the following formula:

$$m = \frac{i - p_b b}{p_m}$$

4. Calculate the demand elasticity(DE) by the following formula:

$$DE = \frac{(b_1 - b_0)/0.5(b_1 + b_0)}{(pb_1 - pb_0)/0.5(pb_1 + pb_0)}$$

5. Merge the provided dataset with the calculate demand elasticity by HSCS(the xi in the data provided) by the following code (note: the round function round calculated HSCS to 3 decimals, and make it matched with the given 3 digits xi in the dataset, the initial Caloric Shares) :

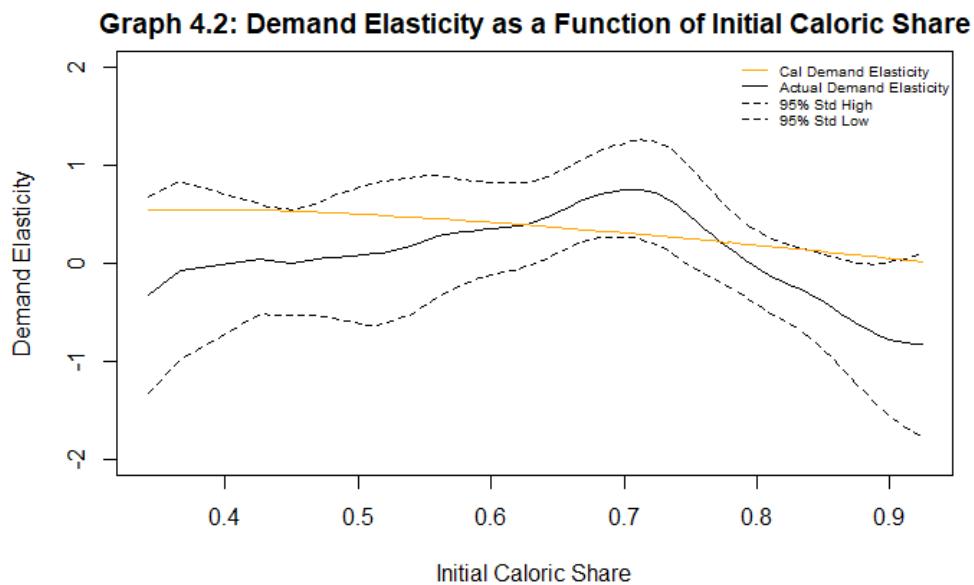
```

data <- data.frame(x_i=round(HSCS1, 3), de_el1)
data1 <- merge(data_old, data, by="x_i")

```

6. Attach our calculated demand elasticity to the original graph. Make comparisons and conclusion, what do you find?(Hint: try the line function)

What you can expect is a graph look like this(How much of the calculated demand elasticity falls into the 95% confidence interval of Jensen and Miller's work?):



5 Simulating Counterfactuals

From previous graph, you are able to test the validity for this model. Now, you are going to use this model to do simulations with counterfactuals.

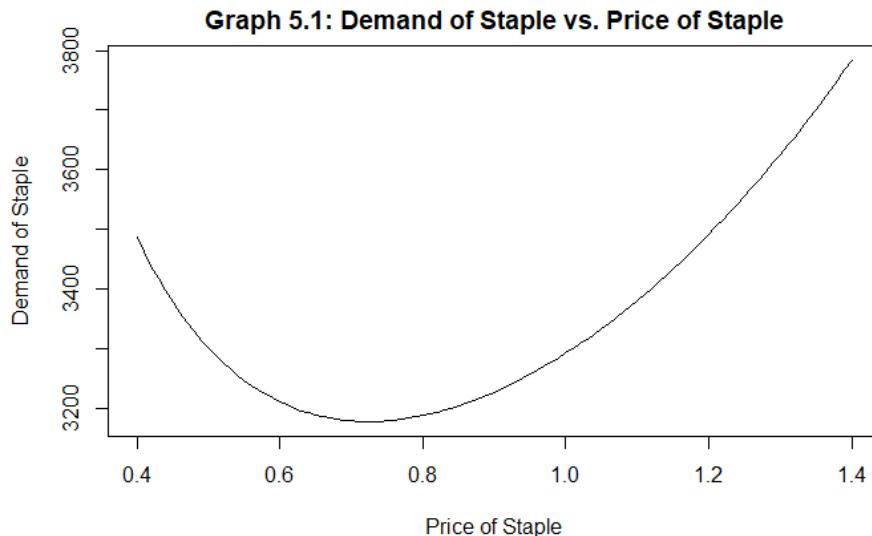
5.1 Quantity of Calories Demand from Staple versus the Price of Staple

In this graph, the main idea is fixing the income and price of meat to calculate the demand of the staple from the previous model.

You may follow these steps to generate the graph:

0. Import the calculated parameters from the given model, remember the calculated $i = 8209.194$, $\bar{c} = 4531.367$, $\hat{\beta} = 0.0593$.
1. Generate sequence for price of staple from 0.4 to 1.4, with gaps in 0.01 each.
2. Calculate the demand for staple with i , $cbar$, beta and price of meat fixed
3. Plot the graph. What do you find? (Hint: what is the lowest point for the graph? What is the meaning?)

What you can expect is a graph look like this:



5.2 Quantity demanded of calories as a function of the price of the staple

Based on previous graph, now you get the similar main idea for this simulation: fixing the income and price of meat to calculate the demand of calories from both the staples and the meat.

You may follow these steps to generate the graph:

0. Import the calculated parameters from the given model, remember the calculated $i = 8209.194$, $\bar{c} = 4531.367$, $\hat{\beta} = 0.0593$.
1. Generate sequence for price of staple from 0.4 to 1.4, with gaps in 0.01 each.

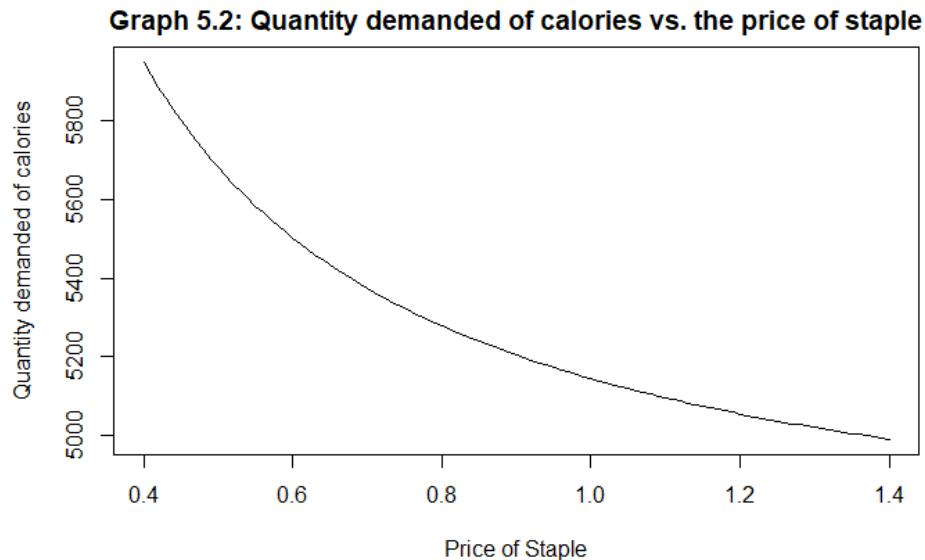
2. Calculate the demand of calories from staple and meat with i , $cbar$, beta and price of meat fixed. From the simulations on 5.1, you have learned how to calculate the demand for staples. Now, you may calculate the demand of meat by transforming this formula:

$$i = mp_m + bp_b$$

By adding the m and b you get, you will get the independent variable for graphing

3. Plot the graph. What do you find?(Hint: what is the lowest point for the graph? What is the meaning)?

What you can expect is a graph look like this:



5.3 Optimal consumption of the household as changing the prices of bread

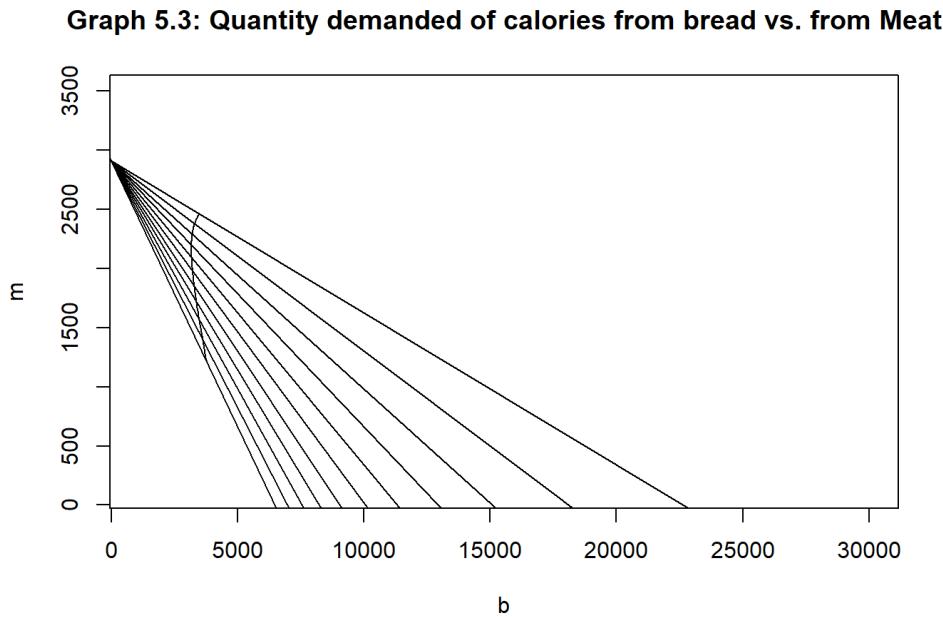
Remember the budget constraint and utilities function graph you learned from intermediate-micro class, it described how consumer's budget affect their utilities and decisions between two goods, which is staple(bread) and meat in our project. For the following two graphs, we will play with budget lines and consumer's decision when the price of staple and even their whole budget change. In this simulation ,we will first deal with the situation when the price of staple change.

You may follow these steps to generate the graph:

- Import the calculated parameters from the given model, remember the calculated $i = 8209.194$, $\bar{c} = 4531.367$, $\hat{\beta} = 0.0593$.

1. Generate sequence for price of staple from 0.4 to 1.4, with gaps in 0.1 each.
2. Calculate the demand of meat in each situation, and graph the demand of bread versus the demand of meat
3. Add more lines showing different budget constraint
 Hint: try the “abline” function, think about the slope and y-intersection point for each budget line
 Hint2: You can add the budget lines one by one, but it looks just...not cool. How to add more lines in a smart way? try the for loop
4. Plot the graph, what do you find?

What you can expect is a graph look like this:



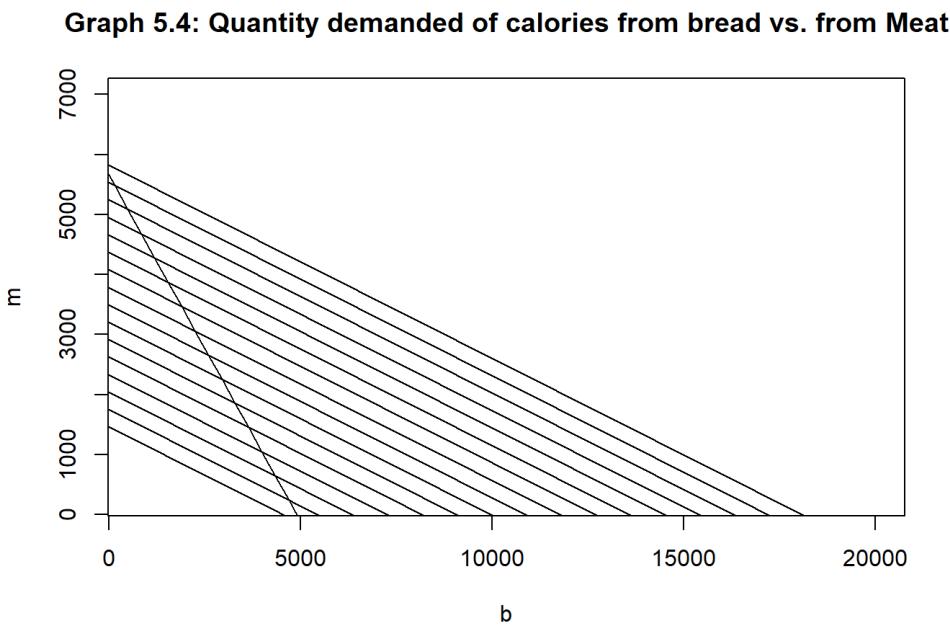
5.4 Optimal consumption of the household as changing the budget

In part 5.4, we will simulate the situation when both prices of bread and meat are fixed, and the budget change.

You may follow these steps to generate the graph:

0. Import the calculated parameters from the given model, remember the calculated $i = 8209.194$, $\bar{c} = 4531.367$, $\hat{\beta} = 0.0593$.
1. Set price of bread to a fix number 1.
2. Define a new parameter about multiples of income, it should be a sequence from 0.5 to 2 with gaps in 0.1. Then, we can change i easily within parameters multiply the old income
3. Calculate the demand of meat and demand of bread under different i
4. Plot the demand of bread vs. demand of meat
5. Add more lines showing different budget constraint. You should consider the slope and the y-axis intersection point in this situation.
6. Plot the graph, what do you find?

What you can expect is a graph look like this:



Reference

Jensen, R. T. and Miller, N. H. (2008). Giffen behavior and subsistence consumption. *American Economic Review*, 98(4):1553–77.

Armendariz, R. Y. (2020). Hunger and the Welfare Effects of Jensen and Miller's (2008) Experiment in Hunan, China.