DSC 5101 ANALYTICS IN MANAGERIAL ECONOMICS

Group Project 1

Estimation of Coffee Demand and Supply Functions

WANG XINRUI (Student ID: A0186103M, Email: xinrui.wang@u.nus.edu)
REN JIEWEN (Student ID: A0186102N, Email: jiewen.ren@u.nus.edu)
ZHANG AO (Student ID: A0176595L, Email: e0232199@u.nus.edu)
LI LIPING (Student ID: A0186040M, Email: liping.li@u.nus.edu)

1.Introduction

The objective of this project is to estimate demand and supply functions of coffee based on coffee consumption and production data from the Dutch market. Mathematical models such as Ordinary Least Square (OLS) regression and Two Stage Least Square (TSLS) regression were applied with the dataset provided.

Section 2 of this report will introduce and explain the methodology of prediction. Section 3 will interpret the regression results and analyze the significance and effectiveness of selected models. Section 4 will discuss the limitations of our models.

2. Methodology

2.1 Choice of Models

To estimate the demand and supply functions of coffee, we assume logarithm linearity for the variables in functions and apply the logarithmic form to OLS regression.

- Demand Function: $\ln\_Q_D = \alpha_D + \eta_D \cdot \ln\_P + control\ variables + \varepsilon_D$
- Supply Function: $\ln\_Q_S = \alpha_S + \eta_S \cdot \ln\_P + control\ variables + \varepsilon_S$

Simple Ordinary Least Squares (OLS) regression model was first applied to see the association between consumption of roasted coffee and its price using log-transformed values of quantity and price. Followed by OLS regression with control variables. The third model applied was the Two-Stage-Least-Squares (TSLS) regression, to eliminate the endogeneity problem existing in the previous 2 models.

2.2 Choice of Variables

We discounted prices by price index to get rid of inflation impact and converted variables into their logarithmic form for interpreting the parameters as elasticity. Refer to Appendix A step 1.

The following are the **key assumptions** for the regressions**:**

1. Coffee market is in equilibrium ($Q_D = qu = Q_S$)
2. Endogeneity of price: consumption and coffee price are determined at the same time. Unobserved factors that increase demand will tend to increase the price as the equilibrium moves up the supply curve

2.2.1 OLS Regression

By analyzing variables provided in the dataset, below demand and supply functions were derived for OLS regression:

- Demand Function:
$$\ln\_qu_d = \alpha_0 + \alpha_1 \cdot \ln\_cprice + \alpha_2 \cdot \ln\_incom + \alpha_3 \cdot \ln\_tprice + \alpha_4 \cdot q_1 + \alpha_5 \cdot q_2 + \alpha_6 \cdot q_3 + \varepsilon_d$$
- Supply Function:
$$\ln\_qu_s = \beta_0 + \beta_1 \cdot \ln\_cprice + \beta_2 \cdot \ln\_wprice + \beta_3 \cdot \ln\_bprice + \beta_4 \cdot q_1 + \beta_5 \cdot q_2 + \beta_6 \cdot q_3 + \varepsilon_s$$

For the demand function, **ln_incom** is chosen as a control variable in the demand function. The presumption is that with a larger income, consumers tend to pay a higher price for coffee, but income is not determined by the coffee market, therefore it's an exogenous variable. **ln_tprice** is a valid control variable as tea is a substitute of coffee, if tprice decreases, the demand for coffee tends to fall. As tea price is not determined by the coffee market, it's exogenous. **q_1, q_2, q_3** are season dummies, representing seasonal fluctuations, for example, length of daytime, that affect the coffee demand. We will only introduce three dummies into our model to control the impact of seasonality, since q_1, q_2 and q_3 are relative to the baseline of q_4, and the default quarter is q_4.

For the supply function, we chose ln_wprice, ln_bprice, q_1, q_2, q_3 as control variables. We included **ln_wprice**, **ln_bprice** because with higher production costs, producers tend to produce less coffee. Since raw material and labor cost are not determined by the coffee market, they are exogenous variables. With similar reasons stated above, q_1, q_2, q_3 are exogenous variables as well. Summary statistics of variables are in Appendix C.

2.2.2 TSLS Regression

Based on the 2 key assumptions stated above, ln_cprice and the error term ε are correlated since unobserved factors that increase demand will tend to increase the price as the equilibrium moves up the supply curve. Working with the above demand and supply equation, we will have ln_cprice in a **reduced form**:

$$\text{ln\_}cprice = c_0 + c_1 \cdot \text{ln\_}incom + c_2 \cdot \text{ln\_}tprice + c_3 \cdot \text{ln\_}bprice + c_4 \cdot \text{ln\_}wprice + c_5 \cdot q_1 + c_6 \cdot q_2 + c_7 \cdot q_3 + \varepsilon_c$$

For **Demand function**, $\boldsymbol{ln\_wprice}$ (price of labor per man hours) and $\boldsymbol{ln\_bprice}$ (price of coffee beans per kg) are selected as instrumental variables (IV). The presumption is that with a higher $wprice$ and $bprice$, price of coffee tends to increase, but that the production cost is not determined by the coffee market, therefore they are uncorrelated with coffee demand. For **Supply function**, $\boldsymbol{ln\_incom}$ (income per capita in current guilders) is selected as instrumental variables (IV). As coffee is not an inferior good, we assume that if income increases, the consumption of customers will increase, and product price will increase eventually. The income of customers is also uncorrelated with the coffee market.

In **First stage of regression,** we will predict ln_cprice by using OLS on the reduced form, the predicted price will be independent from error term by construction. For the **Second stage of regression,** we will use predicted ln_cprice and perform the second OLS for ln_qu.

3. Result Interpretation

3.1 Demand Function

For model 1 (OLS regression between consumption and prices) and model 2 (OLS regression with control variables), results are shown below:

```
Call:
lm(formula = ln_qu ~ ln_cprice, data = rawdata)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.22672 -0.06680 -0.01993  0.06995  0.42112 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02998    0.22725   0.132    0.895
ln_cprice   -0.16509    0.08927  -1.849    0.068

(Intercept)  
ln_cprice   .
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1101 on 82 degrees of freedom
Multiple R-squared:  0.04004,	Adjusted R-squared:  0.02833 
F-statistic:  3.42 on 1 and 82 DF,  p-value: 0.068
```

```
Call:
lm(formula = ln_qu ~ ln_cprice + ln_incom + ln_tprice + q1 + 
    q2 + q3, data = rawdata)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.19669 -0.07418 -0.00985  0.06009  0.32008 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) -4.95797    3.79538  -1.306  0.19533   
ln_cprice   -0.25708    0.08872  -2.898  0.00489 **
ln_incom     0.59946    0.35972   1.666  0.09968 . 
ln_tprice    0.28142    0.50611   0.556  0.57979   
q1          -0.10870    0.03254  -3.340  0.00129 **
q2          -0.09255    0.03046  -3.038  0.00325 **
q3          -0.10429    0.03155  -3.306  0.00144 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09827 on 77 degrees of freedom
Multiple R-squared:  0.2816,	Adjusted R-squared:  0.2256 
F-statistic:  5.03 on 6 and 77 DF,  p-value: 0.000213
```

There is a significant change in the coefficient of ln_cprice after adding control variables in the model 2. Because part of the negative association between demand and price is offset by income, tea price and seasonality, the coefficient in the Demand model 1 would increase. The coefficient of ln_cprice means demand will decrease 0.257% when cprice increases 1%. R-squared is also higher, which means adding extra control variables improve fits of the model. Compared with model 1, results improved but the model still has an endogeneity problem.

Combinations of quantity and price that we observe reflect the forces on both demand and supply. Therefore, the relationship we estimate is a mix of shifting demand and supply curves. To deal with this endogeneity problem we will estimate the demand equation using IV by a procedure called two-stage least squares (TSLS). In addition, we can see that tprice has p-value of 0.57979, indicating there are strong evidences suggesting that $tprice$ is not correlated with coffee demand. Therefore, we should eliminate tprice from our model.

Model 3: Two Stage Least Squares (TSLS) regression (refer to Appendix A step 4)

```
## Call:
## lm(formula = ln_cprice ~ ln_incom + q1 + q2 + q3 + ln_wprice +
##     ln_bprice, data = rawdata)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.089143 -0.025028  0.001005  0.021611  0.099406
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.4260220  1.1667747  -2.079   0.0409 *
## ln_incom     0.2409270  0.1219488   1.976   0.0518 .
## q1          -0.0055528  0.0130770  -0.425   0.6723
## q2          -0.0191101  0.0122646  -1.558   0.1233
## q3          -0.0005488  0.0126336  -0.043   0.9655
## ln_wprice    0.7511677  0.4168538   1.802   0.0755 .
## ln_bprice    0.5010658  0.0203628  24.607   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03893 on 77 degrees of freedom
## Multiple R-squared:  0.9233, Adjusted R-squared:  0.9173
## F-statistic: 154.4 on 6 and 77 DF,  p-value: < 2.2e-16
```

```
## Call:
## lm(formula = ln_qu ~ p.cprice + ln_incom + q1 + q2 + q3, data = rawdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19598 -0.06153 -0.00987  0.06476  0.33733
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.25655    1.89521  -1.718 0.089708 .
## p.cprice    -0.27560    0.09266  -2.974 0.003906 **
## ln_incom     0.48663    0.26485   1.837 0.069965 .
## q1          -0.11158    0.03181  -3.507 0.000754 ***
## q2          -0.09240    0.03026  -3.054 0.003089 **
## q3          -0.10715    0.03079  -3.480 0.000823 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09764 on 78 degrees of freedom
## Multiple R-squared:  0.2816, Adjusted R-squared:  0.2355
## F-statistic: 6.114 on 5 and 78 DF,  p-value: 7.905e-05
```

Refer to Appendix A step 5 and 6, we have a p-value of 0.001385 for Hausman Test and a p-value of 0.73306 for Sargan Test. We therefore reject the null hypothesis in the Hausman Test, which indicates there is no endogeneity between our IV and the residuals. Since p value for Sargan test is significantly greater than 0.05, we can conclude that our instruments are valid.

Comparing results from model 2 and 3, we can see that R-squared improves to 0.2355 and standard errors reduce in model 3. Since it solves the $cprice$'s endogeneity problem, the TSLS is chosen as the final demand model:

$$\mathbf{ln}_{qu_d} = \boldsymbol{\alpha_0} - \mathbf{0.2756} \cdot \mathbf{ln}_{cprice} - \mathbf{0.4866} \cdot \mathbf{ln}_{incom} - \mathbf{0.1116} \cdot \boldsymbol{q_1} - \mathbf{0.0924} \cdot \boldsymbol{q_2} - \mathbf{0.10715} \cdot \boldsymbol{q_3} + \boldsymbol{\varepsilon_d}$$

The elasticity of coffee is -0.2756, implying that the consumption of coffee is relatively inelastic. The level of consumer income has a positive impact of the consumption of coffee, indicating that coffee is a normal good. In addition, since the coefficients are negative for q_1, q_2 and q_3, we can conclude that more coffee is consumed in quarter 4. We have also conducted the test the robustness for the TSLS model. We have also conducted the test the robustness for the TSLS model (refer to Appendix B). The robust standard errors above are modified by White heteroscedasticity correction and they do not deviate much from our previous model. It seems there are not serious "thick tail" problem and coefficient estimations are efficient. So we conclude that the model is robust.

3.2 Supply Function

In this session, we intend to identify the supply function. The procedures of processing data are the same with the demand function. Please Refer to Appendix A step 7 and 8.

First, we will do a simple OLS regression with the only price on quantity to see the relationship between them.

```
Call:
lm(formula = ln_qu ~ ln_cprice, data = rawdata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.22672 -0.06680 -0.01993  0.06995  0.42112

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02998    0.22725   0.132    0.895
ln_cprice   -0.16509    0.08927  -1.849    0.068 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1101 on 82 degrees of freedom
Multiple R-squared:  0.04004,	Adjusted R-squared:  0.02833
F-statistic:  3.42 on 1 and 82 DF,  p-value: 0.068
```

```
Call:
lm(formula = ln_qu ~ ln_cprice + ln_wprice + ln_bprice + q1 +
    q2 + q3, data = rawdata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.19597 -0.07093 -0.00717  0.05625  0.34610

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.22031    3.02203  -1.066 0.289929
ln_cprice   -0.01072    0.28278  -0.038 0.969858
ln_wprice    0.90787    0.93794   0.968 0.336103
ln_bprice   -0.09375    0.15408  -0.608 0.544696
q1          -0.12818    0.03092  -4.145 8.64e-05 ***
q2          -0.09487    0.03152  -3.010 0.003535 **
q3          -0.12093    0.03077  -3.931 0.000184 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09901 on 77 degrees of freedom
Multiple R-squared:  0.2707,	Adjusted R-squared:  0.2139
F-statistic: 4.764 on 6 and 77 DF,  p-value: 0.0003534
```

Compare the two OLS results above, the adjusted R-squared has improved from 0.02833 to 0.2139, so the latter model estimates the supply curve better. We can conclude that ln_wprice, ln_bprice and seasonality play a role in this model. Although result improves, the model still has an endogeneity problem, as reasons suggested above. And the coefficient of ln_cprice shows that when cprice increase 1%, the production of coffee will decrease 0.01%, which is inconsistent with the normal coffee market. The negative coefficient of ln_cprice also proved that our model has an endogeneity problem.

Next, we will use income (income per capita in current guilders) as instruments to do the 2SLS regression to fitting the supply curve. The results are as below.

```
Call:
lm(formula = ln_cprice ~ ln_incom + q1 + q2 + q3 + ln_wprice +
    ln_bprice, data = rawdata)

Residuals:
      Min        1Q    Median        3Q       Max
-0.089143 -0.025028  0.001005  0.021611  0.099406

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.4260220  1.1667747  -2.079   0.0409 *
ln_incom     0.2409270  0.1219488   1.976   0.0518 .
q1          -0.0055528  0.0130770  -0.425   0.6723
q2          -0.0191101  0.0122646  -1.558   0.1233
q3          -0.0005488  0.0126336  -0.043   0.9655
ln_wprice    0.7511677  0.4168538   1.802   0.0755 .
ln_bprice    0.5010658  0.0203628  24.607   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03893 on 77 degrees of freedom
Multiple R-squared:  0.9233, Adjusted R-squared:  0.9173
F-statistic: 154.4 on 6 and 77 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = ln_qu ~ p.cprice + ln_wprice + ln_bprice + q1 +
    q2 + q3, data = rawdata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.19744 -0.06555 -0.00880  0.06082  0.34101

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.06282    4.07851   0.015  0.98775
p.cprice     1.46703    1.27654   1.149  0.25402
ln_wprice   -0.89404    1.78070  -0.502  0.61705
ln_bprice   -0.86470    0.66744  -1.296  0.19900
q1          -0.10502    0.03635  -2.889  0.00501 **
q2          -0.06147    0.04207  -1.461  0.14801
q3          -0.10917    0.03208  -3.403  0.00106 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09817 on 77 degrees of freedom
Multiple R-squared:  0.283,  Adjusted R-squared:  0.2271
F-statistic: 5.066 on 6 and 77 DF,  p-value: 0.0001992
```

Refer to Appendix A step 10, we have a p-value of 1.125e-32 for Hausman Test. Therefore, we reject the null hypothesis in the Hausman Test, which indicates there is no endogeneity between our IV and the residuals. We can conclude that our instruments are valid. And according to results of supply2.lm and 2sls, we can see that R-squared improves to 0.2271 and standard errors reduce in model 3. Thus, we chose 2SLS as the final supply model. The final supply function derived from 2SLS is shown as below:

$$\mathbf{ln\_qu_s = \beta_0 + 1.4670 \cdot ln\_cprice - 0.8940 \cdot ln\_wprice - 0.8647 \cdot ln\_bprice - 0.1050 \cdot q_1 - 0.0615 \cdot q_2 - 0.1092 \cdot q_3 + \varepsilon_s}$$

The elasticity of coffee is 1.4670, implying that the production of coffee is significantly elastic. The wprice and bprice impacts production of coffee negatively, which is the normal case in supply market. Since the coefficients are negative for q_1, q_2 and q_3, we can conclude that more coffee is produced in quarter 4. We have also conducted the test the robustness for the TSLS model and conclude that the model is robust (refer to Appendix B).

4. Limitations

The dataset we built our models on only contains 84 data points. Data used for prediction might not be representative of the actual coffee market. Besides, there is a lack of knowledge of the coffee market structure. From the dataset given, we are not able to find out whether the coffee market in Dutch is in perfect competition, whether it is an oligopoly market or a monopoly market. Therefore, we could only develop our models under the assumption that coffee market is in equilibrium and price of coffee is endogenous. Dutch is renowned as merchants and almost two-thirds of the economy is based on foreign trade, however external impacts from tariff and quota are not considered in the model. There might be some external forces that drive up or down the coffee price, but they are not measurable base on the information we have.

Appendix A

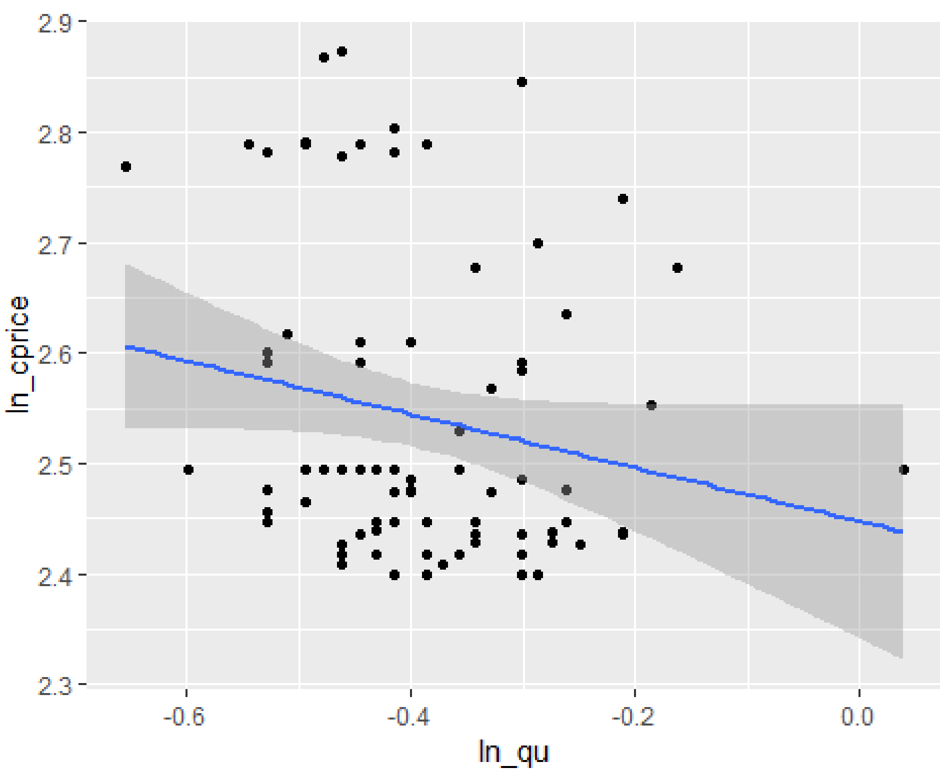
Estimation of Coffee Demand and Supply in Dutch

30 August 2018

Step 1: Data Cleaning

```
#Read data
rawdata = read.csv("Project1Data.csv",header=T)

#Check data
head(rawdata)

##       maand year month   qu cprice tprice oprice   incom q1 q2 q3 q4 bprice
## 1 Jan 1990 1990     1 0.55  12.12   18.6      1 1640.87  1  0  0  0   3.47
## 2 Feb 1990 1990     2 0.65  12.12   18.6      1 1538.60  1  0  0  0   3.40
## 3 Mar 1990 1990     3 0.66  12.12   18.6      1 1680.93  1  0  0  0   3.26
## 4 Apr 1990 1990     4 0.66  12.12   18.6      1 1656.20  0  1  0  0   3.46
## 5 May 1990 1990     5 0.64  12.12   18.6      1 1700.80  0  1  0  0   3.47
## 6 Jun 1990 1990     6 0.65  12.12   18.6      1 1732.67  0  1  0  0   3.68
##   wprice
## 1  28.15
## 2  28.15
## 3  28.33
## 4  28.49
## 5  28.55
## 6  28.55

tail(rawdata)

##        maand year month   qu cprice tprice oprice   incom q1 q2 q3 q4
## 79 Jul 1996 1996     7 0.64  15.63  19.53   1.17 2238.07  0  0  1  0
## 80 Aug 1996 1996     8 0.59  15.63  19.53   1.16 2224.40  0  0  1  0
## 81 Sep 1996 1996     9 0.74  15.63  19.53   1.17 2164.13  0  0  1  0
## 82 Oct 1996 1996    10 0.74  15.63  19.34   1.18 2238.53  0  0  0  1
## 83 Nov 1996 1996    11 0.72  15.39  20.09   1.18 2211.87  0  0  0  1
## 84 Dec 1996 1996    12 0.83  15.15  20.27   1.18 2297.20  0  0  0  1
##    bprice wprice
## 79   4.77  34.15
## 80   4.64  34.15
## 81   4.65  34.15
## 82   4.59  34.21
## 83   4.47  34.21
## 84   4.41  34.18

#Adjustment for Inflation
rawdata$cprice <- rawdata$cprice/rawdata$oprice
rawdata$tprice <- rawdata$tprice/rawdata$oprice
rawdata$bprice <- rawdata$bprice/rawdata$oprice
rawdata$wprice <- rawdata$wprice/rawdata$oprice
rawdata$incom <- rawdata$incom/rawdata$oprice

#Construction of variables in logs
```

```
rawdata$ln_qu <- log(rawdata$qu)
rawdata$ln_cprice <- log(rawdata$cprice)
rawdata$ln_tprice <- log(rawdata$tprice)
rawdata$ln_incom <- log(rawdata$incom)
rawdata$ln_bprice <- log(rawdata$bprice)
rawdata$ln_wprice <- log(rawdata$wprice)

#Plot data
library(ggplot2)
ggplot(rawdata,aes(ln_qu, ln_cprice)) + geom_point() + geom_smooth(method = "lm")
```



Step 2: OLS regression between consumption and prices

```
demand1.lm <- lm(ln_qu ~ ln_cprice,data=rawdata)
summary(demand1.lm)

## 
## Call:
## lm(formula = ln_qu ~ ln_cprice, data = rawdata)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.22672 -0.06680 -0.01993  0.06995  0.42112 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  0.02998    0.22725   0.132    0.895  
## ln_cprice   -0.16509    0.08927  -1.849    0.068 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.1101 on 82 degrees of freedom
## Multiple R-squared:  0.04004,    Adjusted R-squared:  0.02833
## F-statistic:  3.42 on 1 and 82 DF,  p-value: 0.068
```

Step 3: OLS regression with other control variables

```
demand2.lm <- lm(ln_qu ~ ln_cprice + ln_incom + ln_tprice + q1 + q2 + q3,data=rawdat
a )
summary(demand2.lm)

##
## Call:
## lm(formula = ln_qu ~ ln_cprice + ln_incom + ln_tprice + q1 +
##     q2 + q3, data = rawdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19669 -0.07418 -0.00985  0.06009  0.32008
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.95797    3.79538  -1.306  0.19533
## ln_cprice   -0.25708    0.08872  -2.898  0.00489 **
## ln_incom     0.59946    0.35972   1.666  0.09968 .
## ln_tprice    0.28142    0.50611   0.556  0.57979
## q1          -0.10870    0.03254  -3.340  0.00129 **
## q2          -0.09255    0.03046  -3.038  0.00325 **
## q3          -0.10429    0.03155  -3.306  0.00144 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09827 on 77 degrees of freedom
## Multiple R-squared:  0.2816, Adjusted R-squared:  0.2256
## F-statistic:  5.03 on 6 and 77 DF,  p-value: 0.000213
```

Step 4: TSLS for demand

```
#library(AER)
#demand.2sls.form <- ivreg(ln_qu ~ ln_cprice + ln_incom + q1 + q2 + q3| ln_incom + q
1 + q2 + q3 + ln_bprice + ln_wprice, data=rawdata)
#summary(demand.2sls.form, diagnostics = TRUE)

#Run 2SLS on ln_cprice
cprice.reduced.form <- lm(ln_cprice ~ ln_incom + q1 + q2 + q3 + ln_wprice + ln_bpric
e,data=rawdata)
summary(cprice.reduced.form)

##
## Call:
## lm(formula = ln_cprice ~ ln_incom + q1 + q2 + q3 + ln_wprice +
##     ln_bprice, data = rawdata)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
```

```
## -0.089143 -0.025028  0.001005  0.021611  0.099406
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.4260220  1.1667747  -2.079   0.0409 *
## ln_incom     0.2409270  0.1219488   1.976   0.0518 .
## q1          -0.0055528  0.0130770  -0.425   0.6723
## q2          -0.0191101  0.0122646  -1.558   0.1233
## q3          -0.0005488  0.0126336  -0.043   0.9655
## ln_wprice    0.7511677  0.4168538   1.802   0.0755 .
## ln_bprice    0.5010658  0.0203628  24.607   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03893 on 77 degrees of freedom
## Multiple R-squared:  0.9233, Adjusted R-squared:  0.9173
## F-statistic: 154.4 on 6 and 77 DF,  p-value: < 2.2e-16
```

```
p.cprice= predict(cprice.reduced.form)
struc.2sls.form <- lm(ln_qu ~ p.cprice + ln_incom + q1 + q2 + q3,data=rawdata )
summary(struc.2sls.form)
```

```
##
## Call:
## lm(formula = ln_qu ~ p.cprice + ln_incom + q1 + q2 + q3, data = rawdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19598 -0.06153 -0.00987  0.06476  0.33733
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.25655    1.89521  -1.718 0.089708 .
## p.cprice    -0.27560    0.09266  -2.974 0.003906 **
## ln_incom     0.48663    0.26485   1.837 0.069965 .
## q1          -0.11158    0.03181  -3.507 0.000754 ***
## q2          -0.09240    0.03026  -3.054 0.003089 **
## q3          -0.10715    0.03079  -3.480 0.000823 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09764 on 78 degrees of freedom
## Multiple R-squared:  0.2816, Adjusted R-squared:  0.2355
## F-statistic: 6.114 on 5 and 78 DF,  p-value: 7.905e-05
```

Step 5: Hausman test for demand

```
hausman_test <- lm(struc.2sls.form$residuals ~ ln_wprice + ln_bprice,data=rawdata)
summary(hausman_test)
```

```
##
## Call:
## lm(formula = struc.2sls.form$residuals ~ ln_wprice + ln_bprice,
##     data = rawdata)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19789 -0.06576 -0.00895  0.06149  0.34062
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.943650   2.828266  -0.334    0.740
## ln_wprice    0.278700   0.833978   0.334    0.739
## ln_bprice    0.002705   0.042912   0.063    0.950
##
## Residual standard error: 0.09575 on 81 degrees of freedom
## Multiple R-squared:  0.001385,   Adjusted R-squared:  -0.02327
## F-statistic: 0.05616 on 2 and 81 DF,  p-value: 0.9454
```

```
print(summary(hausman_test)$r.squared)
```

```
## [1] 0.00138477
```

Step 6: Sargan Test for demand

```
p.demand.qu <- predict(struc.2sls.form,rawdata)
head(p.demand.qu)
```

```
##          1          2          3          4          5          6
## -0.4498000 -0.4740290 -0.4323626 -0.4250678 -0.4147352 -0.4150478
```

```
rawdata$qu_error_demand <- rawdata$ln_qu - p.demand.qu
sargan_test_demand <- lm(qu_error_demand ~ ln_bprice +ln_wprice,data = rawdata)
summary(sargan_test_demand)
```

```
##
## Call:
## lm(formula = qu_error_demand ~ ln_bprice + ln_wprice, data = rawdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19789 -0.06576 -0.00895  0.06149  0.34062
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.943650   2.828266  -0.334    0.740
## ln_bprice    0.002705   0.042912   0.063    0.950
## ln_wprice    0.278700   0.833978   0.334    0.739
##
## Residual standard error: 0.09575 on 81 degrees of freedom
## Multiple R-squared:  0.001385,   Adjusted R-squared:  -0.02327
## F-statistic: 0.05616 on 2 and 81 DF,  p-value: 0.9454
```

```
sargan_demand_stat = summary(sargan_test_demand)$r.squared * nrow(rawdata)
sargan_demand_pvalue = pchisq(sargan_demand_stat, 1, lower.tail = FALSE)
print(sargan_demand_pvalue)
```

```
## [1] 0.7330598
```

Step 7: OLS regression between supply and price

```
#Naive Supply Function Estimation
supply1.lm <- lm(ln_qu ~ ln_cprice,data=rawdata)
summary(supply1.lm)
```

```
##
## Call:
## lm(formula = ln_qu ~ ln_cprice, data = rawdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.22672 -0.06680 -0.01993  0.06995  0.42112
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.02998    0.22725   0.132    0.895
## ln_cprice   -0.16509    0.08927  -1.849    0.068 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1101 on 82 degrees of freedom
## Multiple R-squared:  0.04004,    Adjusted R-squared:  0.02833
## F-statistic:  3.42 on 1 and 82 DF,  p-value: 0.068
```

Step 8: OLS regression with other control variables

```
#Supply Function OLS Estimation
supply2.lm <- lm(ln_qu ~ ln_cprice + ln_wprice + ln_bprice + q1 + q2 + q3,data=rawda
ta)
summary(supply2.lm)
```

```
##
## Call:
## lm(formula = ln_qu ~ ln_cprice + ln_wprice + ln_bprice + q1 +
##     q2 + q3, data = rawdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19597 -0.07093 -0.00717  0.05625  0.34610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.22031    3.02203  -1.066 0.289929
## ln_cprice   -0.01072    0.28278  -0.038 0.969858
## ln_wprice    0.90787    0.93794   0.968 0.336103
## ln_bprice   -0.09375    0.15408  -0.608 0.544696
## q1          -0.12818    0.03092  -4.145 8.64e-05 ***
## q2          -0.09487    0.03152  -3.010 0.003535 **
## q3          -0.12093    0.03077  -3.931 0.000184 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09901 on 77 degrees of freedom
```

```
## Multiple R-squared:  0.2707, Adjusted R-squared:  0.2139
## F-statistic: 4.764 on 6 and 77 DF,  p-value: 0.0003534
```

Step 9: TSLS for supply

```
#Run 2SLS on ln_cprice
cprice.reduced.form2 <- lm(ln_cprice ~ ln_incom + q1 + q2 + q3 + ln_wprice + ln_bpri
ce,data=rawdata)
summary(cprice.reduced.form2)

##
## Call:
## lm(formula = ln_cprice ~ ln_incom + q1 + q2 + q3 + ln_wprice +
##     ln_bprice, data = rawdata)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.089143 -0.025028  0.001005  0.021611  0.099406
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.4260220  1.1667747  -2.079   0.0409 *
## ln_incom     0.2409270  0.1219488   1.976   0.0518 .
## q1          -0.0055528  0.0130770  -0.425   0.6723
## q2          -0.0191101  0.0122646  -1.558   0.1233
## q3          -0.0005488  0.0126336  -0.043   0.9655
## ln_wprice    0.7511677  0.4168538   1.802   0.0755 .
## ln_bprice    0.5010658  0.0203628  24.607   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03893 on 77 degrees of freedom
## Multiple R-squared:  0.9233, Adjusted R-squared:  0.9173
## F-statistic: 154.4 on 6 and 77 DF,  p-value: < 2.2e-16

p.cprice2= predict(cprice.reduced.form2)
supply.2sls.form <- lm(ln_qu ~ p.cprice + ln_wprice + ln_bprice + q1 + q2 + q3 ,data
=rawdata)
summary(supply.2sls.form)

##
## Call:
## lm(formula = ln_qu ~ p.cprice + ln_wprice + ln_bprice + q1 +
##     q2 + q3, data = rawdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19744 -0.06555 -0.00880  0.06082  0.34101
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.06282    4.07851   0.015  0.98775
## p.cprice     1.46703    1.27654   1.149  0.25402
## ln_wprice   -0.89404    1.78070  -0.502  0.61705
```

```
## ln_bprice   -0.86470    0.66744  -1.296  0.19900
## q1          -0.10502    0.03635  -2.889  0.00501 **
## q2          -0.06147    0.04207  -1.461  0.14801
## q3          -0.10917    0.03208  -3.403  0.00106 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09817 on 77 degrees of freedom
## Multiple R-squared:  0.283,  Adjusted R-squared:  0.2271
## F-statistic: 5.066 on 6 and 77 DF,  p-value: 0.0001992
```

Step 10: Hausman test for supply

```
#Hausman test
hausman_test <- lm(supply.2sls.form$residuals ~ ln_wprice + ln_bprice,data=rawdata)
summary(hausman_test)
```

```
##
## Call:
## lm(formula = supply.2sls.form$residuals ~ ln_wprice + ln_bprice,
##     data = rawdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19744 -0.06555 -0.00880  0.06082  0.34101
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.029e-16  2.827e+00       0        1
## ln_wprice    1.858e-16  8.337e-01       0        1
## ln_bprice   -9.330e-18  4.290e-02       0        1
##
## Residual standard error: 0.09571 on 81 degrees of freedom
## Multiple R-squared:  1.125e-32,  Adjusted R-squared:  -0.02469
## F-statistic: 4.557e-31 on 2 and 81 DF,  p-value: 1
```

```
print(summary(hausman_test)$r.squared)
```

```
## [1] 1.125091e-32
```

Appendix B

Robustness test for TSLS models

```
. ivregress 2sls ln_qu (ln_cprice=ln_bprice ln_wprice) q1 q2 q3 ln_incom, robust

Instrumental variables (2SLS) regression          Number of obs   =         84
                                                  Wald chi2(5)    =      35.64
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2783
                                                  Root MSE        =      .0943

------------------------------------------------------------------------------
             |               Robust
       ln_qu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   ln_cprice |  -.2755988    .0775266    -3.55   0.000    -.4275481   -.1236495
          q1 |  -.1115757    .0333443    -3.35   0.001    -.1769294   -.0462221
          q2 |  -.0924008    .0287438    -3.21   0.001    -.1487376    -.036064
          q3 |  -.1071489    .0341876    -3.13   0.002    -.1741553   -.0401424
    ln_incom |    .486626    .2164286     2.25   0.025     .0624337    .9108183
       _cons |  -3.256544     1.56825    -2.08   0.038    -6.330257   -.1828298
------------------------------------------------------------------------------
Instrumented:  ln_cprice
Instruments:   q1 q2 q3 ln_incom ln_bprice ln_wprice
```

Figure B-1 Robustness test for TSLS demand model

```
. ivregress 2sls ln_qu (ln_cprice=ln_incom) ln_wprice ln_bprice q1 q2 q3 , robust

Instrumental variables (2SLS) regression          Number of obs   =         84
                                                  Wald chi2(6)    =      20.55
                                                  Prob > chi2     =     0.0022
                                                  R-squared       =     0.0121
                                                  Root MSE        =     .11033

------------------------------------------------------------------------------
             |               Robust
       ln_qu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   ln_cprice |   1.467024    1.298291     1.13   0.258    -1.077579    4.011626
   ln_wprice |  -.8940325    2.015433    -0.44   0.657    -4.844208    3.056143
   ln_bprice |  -.8647009    .6824172    -1.27   0.205    -2.202214    .4728122
          q1 |  -.1050228    .0405329    -2.59   0.010    -.1844659   -.0255798
          q2 |  -.0614696    .0449553    -1.37   0.172    -.1495803    .0266411
          q3 |  -.1091727    .0393682    -2.77   0.006     -.186333   -.0320125
       _cons |   .0627872    4.698137     0.01   0.989    -9.145391    9.270966
------------------------------------------------------------------------------
Instrumented:  ln_cprice
Instruments:   ln_wprice ln_bprice q1 q2 q3 ln_incom
```

Figure B-2 Robustness test for TSLS supply model

Appendix C

Control Variables for demand and supply functions:

Functions	Control Variables
Demand	ln_incom, ln_tprice, q1, q2, q3
Supply	ln_wprice, ln_bprice, q1, q2, q3

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
maand	0				
year	84	1993	2.012012	1990	1996
month	84	6.5	3.472786	1	12
qu	84	.6815476	.0796301	.52	1.04
cprice	84	14.02976	2.555492	12	20
tprice	84	19.19155	.5254499	18.41	20.27
oprice	84	1.089048	.0574691	1	1.18
incom	84	1956.982	183.0193	1538.6	2297.2
q1	84	.25	.4356134	0	1
q2	84	.25	.4356134	0	1
q3	84	.25	.4356134	0	1
q4	84	.25	.4356134	0	1
bprice	84	4.030119	1.233319	2.44	7.18
wprice	84	31.79143	1.839375	28.15	34.21
ln_qu	84	-.3897157	.1116681	-.6539265	.0392207
ln_cprice	84	2.542223	.1353501	2.398729	2.873515
ln_tprice	84	2.870185	.0307959	2.796661	2.923374
ln_wprice	84	3.373586	.0130576	3.337547	3.403971
ln_bprice	84	1.268638	.2537721	.8243394	1.849082
ln_incom	84	7.490851	.0481092	7.338628	7.574934