Demand Model Simulation in R with Endogenous Prices and Unobservable Quality

Daniel Toro Gonzalez

Universidad Tecnológica de Bolívar

18. September 2014

Online at http://mpra.ub.uni-muenchen.de/59117/
MPRA Paper No. 59117, posted 7. October 2014 14:04 UTC
Demand Model Simulation in R with Endogenous Prices and Unobservable Quality

Daniel Toro Gonzalez*

Abstract

The aim of the present routine is to simulate a demand equation with endogenous prices and unobservable product quality and to retrieve the original parameters using the Control Function (CF) approach. The CF approach is a very useful and simple method to obtain unbiased estimates. The present R code helps to understand the underlying structure of the endogeneity problem in demand estimations. Results support the important bias correction of the CF approach.

Key Words: R, Demand, Endogeneity, Simulation, Endogenous Prices

JEL: C13, C15, D49, E27

Introduction

The exercise consists in the repeated estimation of three models by OLS, the original model designed by simulation of the variables, non-observable quality model which omits the quality variable from the specification, and the Control Function (CF) model which is estimated in two stages.

* Associate Professor at the Universidad Tecnológica de Bolívar in Cartagena - Colombia. 2014.
Routine

The software can be downloaded from the R Project for Statistical Computing at the web site www.r-project.org.

# SIMULATE A Demand model with ENDOGENOUS prices
# Daniel Toro Gonzalez 2014
# Universidad Tecnológica de Bolivar

rm(list=ls()) # Remove objects from the workspace
# Clearing memory space

install.packages("AER") # In case is not installed remove
install.packages("systemfit") # In case is not installed remove
library(AER) # Load library AER
library(systemfit) # Load library Systemfit

n <- 50 # Sample size (Change at will)
S <- 1000 # Number of samples (Change at will)

B0 <- 100 # Intercept of Demand
B1 <- -2 # Price coefficient
B2 <- +5 # Income coefficient
B3 <- +3 # Quality coefficient

parameters <- c(B0, B1, B2, B3) # Save the original population parameters for printing purposes

# Creates empty matrices to store the parameters in each trial
results0 <- matrix(nrow=S, ncol=4)
results1 <- matrix(nrow=S, ncol=3) # There is no coefficient for X
results2 <- matrix(nrow=S, ncol=4)

# The following Loop generates S number of repetitions (or samples)

---

1 Created by Pretty R at inside-R.org
for (i in 1:S) {
    set.seed(i)
    # Errors
    e1 <- rnorm(n, mean=0, sd=1)  # Error e1~N(0,k) k=1, this term is included in the demand equation
    e2 <- rnorm(n, mean=0, sd=1)  # Error e2~N(0,k) k=1, this term will be included in the endogenous price equation

    # Quality (Exogenous variable, non observable)
    X <- rnorm(n, mean=10, sd=1)  # Quality X~N(10,k) k=1

    # Price of Labor (Exogenous variable, observable)
    W <- rnorm(n, mean=18, sd=4)  # Minimum wage by hour W~N(18,k) k=1

    # Average Exogenous Income US$5000
    I <- rnorm(n, mean=5000, sd=1000)  # Income I~N(5000,500)

    # GENERATING ENDOGENOUS PRICES
    # Endogenous Prices depend on labor cost (W), Quality (X) and a random term e2~N(0,k) k=1
    # The values of the parameters (10, 0.5, 2) can be modified at will
    P <- 10+0.5*W+2*X+e2

    # Generating quantity values for the demand function
    # using the unobservable population parameters (B0, B1, B2, B3)
    # and the simulated values of the variables (P, I, X, e1)
    Q <- B0+B1*P+B2*I+B3*X+e1  # Demand Curve

    #BASE ESTIMATION
    # Reg0 is the estimation of the demand model as if quality is an observable variable
    # this model replicates closely the real data generating process
    reg0 <- lm(Q ~ P + I + X)
    results0[i,] <- coef(reg0)

    #ENDOGENOUS PRICE ESTIMATION
    # Reg1 is the estimation omiting Quality which is unobservable but highly correlated with the prices
    # this model yields bias estimators for the price parameter
    reg1 <- lm(Q ~ P + I)
    results1[i,] <- coef(reg1)
### CONTROL FUNCTION ###

# Stage No.1
# Assume W is observable so we can estimate P = D0+D1W+u2
# P=10+0.5*W+2*X+e2

# ESTRUCTURAL PRICE EQUATION #

```
regP <- lm(P ~ W)
CF <- resid(regP)  # The Control Function Variable is # the error term of the Estructural Price Equation
```

# DEMAND EQUATION WITH CONTROL #
# Stage No.2
# We estimate the demand model using CF instead of X,
# hence: Q = B0+B1*P+B2*I+B3*CF

```
reg2 <- lm(Q ~ P + I + CF)
results2[,] <- coef(reg2)
```

------

### PRINT RESULTS ###

```{r}
colnames(results0) <- c("INTERC","PRICE", "INCOME", "QUALITY")
colnames(results1) <- c("INTERC","PRICE", "INCOME")
colnames(results2) <- c("INTERC","PRICE", "INCOME", "QUALITY")
summary(results0)  # ORIGINAL MODEL
summary(results1)  # UNOBSERVABLE QUALITY
# Bias in price coefficient (%)
(mean(results1[,2])-mean(results0[,2]))/mean(results0[,2])*100
summary(results2)  # CONTROL FUNCTION
# Bias in price coefficient (%)
(mean(results2[,2])-mean(results0[,2]))/mean(results0[,2])*100
```

### Results ###

The first set of coefficients corresponds to the original model, including all the variables price, income and quality, to explain the demand. For each of the one thousand samples, the set of three parameters is estimated and the summary statistics of the vector of parameters is presented in the next table. It is clear that since all the relevant variables of the model are included, the parameter values on average are very close to the original parameter values (B0=100, B1=-2, B2=5, B3=3).
The second set of coefficients corresponds to the omitted quality model. In this case the variables explaining the demanded quantities are price and income. As in the previous case, for each of the one thousand samples the set of two parameters is estimated and the summary statistics of the vector of parameters is presented in the next table. In this case since not all the relevant variables of the model are included, the parameter values, specifically the price coefficient is biased on average. Compared to the original parameter of -2, the observed parameter is biased in -33% (-1.33).

Finally, when the control function method is implemented by using the errors of the structural price equation to replace the variable Quality, the results show that the price coefficient is no longer biased.

Conclusion

The CF strategy allows the researcher to control for the unobservable factors and to correctly identify the price parameter.