# Logit Exercises

*Aimee Barciauskas, Roger Cusco, Hari S*

*November 11, 2015*

## 1. Specify a logit demand model model, in which you include observable product characteristics.

```
setwd('~/Box Sync/abarciausksas/15D011 Economics for the Era of Big Data/Christian Michel/PS3/')

library(foreign)

# Load data
data <- read.dta('cars.dta')

households <- unique(data$pop) / 3
outside_market_share <- (households - sum(data[,'qu'])) / households
data$inside_market_share <- data$qu/households

# hp: horsepower
# li: fuel consumption in liters per 100 km
# wi, he: width and height in cm
# cy: displacement (in cc)
# cla: factor(segment)

data$demand <- log(data$inside_market_share) - log(outside_market_share)

data$brd <- factor(data$brd)
data$cla <- factor(data$cla)

model.1 <- lm(demand ~ hp + li + wi + he + cy + cla + princ, data = data)
summary(model.1)
```

```
##
## Call:
## lm(formula = demand ~ hp + li + wi + he + cy + cla + princ, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7849 -0.8154  0.1880  0.9111  2.6624
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.304e+01  3.485e+00  -6.612 1.85e-10 ***
## hp          -6.744e-02  1.046e-02  -6.446 4.85e-10 ***
## li           3.085e-02  7.756e-02   0.398  0.69111
## wi           1.289e-01  1.691e-02   7.627 3.54e-13 ***
## he          -1.071e-02  1.984e-02  -0.540  0.58959
## cy          -1.235e-03  5.899e-04  -2.093  0.03724 *
## cla2         5.717e-01  2.791e-01   2.048  0.04143 *
```

```
## cla3         4.365e-01  3.388e-01   1.288  0.19865
## cla4         4.000e-01  4.865e-01   0.822  0.41165
## cla5         6.437e-01  6.317e-01   1.019  0.30905
## princ        1.946e+00  6.487e-01   3.000  0.00294 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.25 on 287 degrees of freedom
## Multiple R-squared:  0.3254, Adjusted R-squared:  0.3019
## F-statistic: 13.84 on 10 and 287 DF,  p-value: < 2.2e-16
```

## 2. Discuss endogeneity problems and how you are going to treat them.

Endogeneity problems arise when unobservable characteristics are correlated with the regressors. In this dataset, the effect of price may be correlated with unobservable characteristics like style. Therefore, any estimate of the effect of price may be biased by these unobservable characteristics.

We are going to treat them by using an instrumental variable which is correlated with price but not with unobservables. We expect the instrumental variable to be correlated with the demand but only through price. We have selected to sum the characteristics of products in other segments as the instrumental variable.

We will try this instrument in **3** and summarize our conclusions about endogeneity in price in the **Closing Remarks** section at the end of the document.

## 3. Estimate the demand equation of the logit model with and without instrumenting for price.

We estimated the demand equation of the logit model without instrumenting for price in part 1.

```r
# Want to regress price on our instrumental variables which will be defined as the sum of the character
calculate_iv <- function(row, characteristic) {
  # other_products not in our segment
  segment <- row['cla']
  other_products <- subset(data, cla != segment)
  sum(other_products[,characteristic])
}

data$iv.hp <- apply(data, 1, function(row) calculate_iv(row, 'hp'))
data$iv.li <- apply(data, 1, function(row) calculate_iv(row, 'li'))
data$iv.wi <- apply(data, 1, function(row) calculate_iv(row, 'wi'))
data$iv.he <- apply(data, 1, function(row) calculate_iv(row, 'he'))
data$iv.cy <- apply(data, 1, function(row) calculate_iv(row, 'cy'))

price_model <- lm(princ ~ hp + li + wi + he + cy + iv.hp + iv.li + iv.wi + iv.he + iv.cy, data = data)
data$priceiv <- fitted.values(price_model)
# Sanity check
summary(price_model)
```

```
##
## Call:
## lm(formula = princ ~ hp + li + wi + he + cy + iv.hp + iv.li +
##     iv.wi + iv.he + iv.cy, data = data)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29456 -0.06834 -0.00809  0.05460  0.48518
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.143e+00  8.311e-01  -6.188 2.09e-09 ***
## hp           5.716e-03  8.888e-04   6.431 5.25e-10 ***
## li          -4.010e-03  7.042e-03  -0.569 0.569533
## wi           3.169e-03  1.524e-03   2.079 0.038521 *
## he           2.716e-03  1.795e-03   1.513 0.131410
## cy           2.018e-04  5.225e-05   3.863 0.000139 ***
## iv.hp        1.105e-03  2.773e-04   3.985 8.55e-05 ***
## iv.li       -1.280e-02  2.884e-03  -4.437 1.30e-05 ***
## iv.wi       -4.290e-04  1.804e-04  -2.378 0.018061 *
## iv.he        8.732e-04  2.721e-04   3.209 0.001481 **
## iv.cy               NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1136 on 288 degrees of freedom
## Multiple R-squared:  0.8826, Adjusted R-squared:  0.8789
## F-statistic: 240.6 on 9 and 288 DF,  p-value: < 2.2e-16
```

```r
# Instrumented price model
model.2 <- lm(demand ~ hp + li + wi + he + cy + priceiv, data = data)
summary(model.2)
```

```
##
## Call:
## lm(formula = demand ~ hp + li + wi + he + cy + priceiv, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6567 -0.7334  0.1657  1.0192  2.5218
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.540e+01  3.532e+00  -7.192 5.41e-12 ***
## hp          -5.910e-02  1.302e-02  -4.540 8.26e-06 ***
## li           2.469e-02  7.770e-02   0.318    0.751
## wi           1.434e-01  1.527e-02   9.390  < 2e-16 ***
## he          -1.073e-02  2.051e-02  -0.523    0.601
## cy          -1.119e-03  6.392e-04  -1.751    0.081 .
## priceiv      1.343e+00  1.103e+00   1.218    0.224
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.271 on 291 degrees of freedom
## Multiple R-squared:  0.2932, Adjusted R-squared:  0.2786
## F-statistic: 20.12 on 6 and 291 DF,  p-value: < 2.2e-16
```

## 4. Price elasticities

```r
alpha <- coefficients(model.2)['priceiv'][[1]]

# Own price elasticity is the elasticity of demand for project j given a change in price j
data$elasticity.own <- -alpha*(1-data$inside_market_share)*(data$princ)
# Cross price elasticity is the elasticity of demand for project x given a change in price j
# We can calculate the cross price elasticity for every product
data$elasticity.cross <- alpha*(data$inside_market_share)*data$princ
summary(data$elasticity.own)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.9060 -1.0880 -0.7665 -0.8925 -0.5443 -0.3320
```

```r
summary(data$elasticity.cross)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 6.131e-06 1.410e-04 4.316e-04 9.316e-04 1.307e-03 5.761e-03
```

```r
## Marginal cost
data$marginal_cost <- data$princ * (1 + 1/data$elasticity.own)
summary(data$marginal_cost)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.49740 -0.33970 -0.17410 -0.08013  0.06584  1.41900
```

General patterns:

- Cross-price elasticity is positive with values between near 0 and 0.005022. The interpretation is that demand for product j when the price in product k increses is increasing but near zero.
- Own-price elasticity is negative with values between -2.91 and -0.3. The intpretation is that demand for product j when the price in product j increses is decreasing.
- Marginal cost is positive between -0.497 and 1.7.

## Closing Remarks

We found that there was not a great improvement in fit between the instrumented price model when compared with the original model. Further, we found price to show near zero correlation with the residuals. We conclude there is nothing to be gained from this instrumental variable and try a different instrumental variable with additional data available to instrument for price.