Set2_Task1

Task 1

How could the relationship between price and demand be affected by endogeneity?

Endogeneity describes the problem, that a regressor x may predict the dependent variable very well, but has no causal relationship with it. There are two mechanisms creating endogeneity: Omitted variables and reverse causality.

Reverse causality: In an ideal economic model the demand is a function of the price. The higher the price, the lower the demand and vice versa. But in reality, the price can also be determined by the demand i.e. when a buyer knows that he is the only one interested in the product. Then, we have a reverse causality.

As the lecture focuses more on omitted variables, we will also do so in the following.

The relationship between price and demand (quantity of sold fish) in the fish market may be affected by endogeneity, when the price correlates well with the demand, but is not the causality. Instead, a omitted variable is the common cause for price and demand.

This omitted variable could be the amount of received fish, i.e. the supply. In this case the quantity and price are determined by the amount of fish received at the day. When there is not much fish available, the seller just cannot sell more fish resulting in a lower quantity with higher prices. On the other hand, when there is a lot of fish available the seller can sell more, resulting in a higher quantity with lower prices. This plausability argument is supported by the following data:

```
price_demand = lm(daily_data$qty~ daily_data$price)
price_demand_totr = lm(daily_data$qty~ daily_data$price + daily_data$totr)
round(summary(price_demand)$coefficients, 5)
```

```
##                   Estimate Std. Error  t value Pr(>|t|)
## (Intercept)       9615.381   1038.694  9.25718  0.00000
## daily_data$price -3709.017   1098.750 -3.37567  0.00102
```

```
round(summary(price_demand_totr)$coefficients, 5)
```

```
##                     Estimate Std. Error  t value Pr(>|t|)
## (Intercept)       3581.90280  802.71462  4.46224  0.00002
## daily_data$price -1402.96103  713.79407 -1.96550  0.05192
## daily_data$totr      0.62142    0.04802 12.93994  0.00000
```

If we include the number of received fish (totr) in the model, the effect of the price is reduced from an initial -3709 (without totr) to -1403 (with totr). Also the p-value of price in the model including totr is above 5%, indicating that it is not significant. The effect of the received fish on the other hand is very significant (very small p-value for totr).

Implications: The quantity of sold fish is not determined by the price. Instead it is determined by the supply of fish. The buyers of fish are not sensitive to the price (meaning the price elasticity is low) but to the availability of fish.

Is weather data a suitable instrument in this context?

In her dataset Graddy classified the weather as stormy when a certain wave height and wind speed are exceeded. The wave height and wind speed are the moving averages of the last three days' wind speed and wave height before the trading day. Her argument is, that storms are an important determinant of the supply as strong winds and high waves make it difficult to catch fish. If supply is high, quantities rise and prices falls and vice versa. This is supported by the following data:

```r
cor(daily_data$price, daily_data$stormy)
```

```
## [1] 0.4227539
```

```r
# Fit linear models holding day-of-week constant
# Day1..Day4 are dummies for weekdays
lm_qty <- lm(qty ~ stormy + day1 + day2 + day3 + day4, data = daily_data)
lm_price <- lm(price ~ stormy + day1 + day2 + day3 + day4, data = daily_data)
# Calculate predicted differences between clear (stormy=0) and stormy
# (stormy=1) days
# For quantity:
coef_qty <- coef(lm_qty)
diff_qty <- coef_qty["stormy"]

# For price:
coef_price <- coef(lm_price)
diff_price <- coef_price["stormy"]

# Overall averages
mean_qty <- mean(daily_data$qty)
mean_price <- mean(daily_data$price)

# Print results
diff_qty
```

```
##  stormy
## -2370.7
```

```r
diff_price
```

```
##   stormy
## 0.322184
```

```r
mean_qty
```

```
## [1] 6334.666
```

```r
mean_price
```

```
## [1] 0.8845243
```

On a stormy day the average quantity of sold fish shrinks by 2370.7 pounds and the price rises by 32.22 cents. The average price on all days was 88.45 cents and average quantity was 6334.67 pounds. The correlation of price and stormy weather is also positive. Therefore the requirement of relevance is fulfilled.

```r
cor(daily_data$price, daily_data$stormy)
```

```
## [1] 0.4227539
```

Another requirement for instruments is exogeneity. So, to predict prices for fish, we need a variable that is independent of supply. Storms are such a variable as the supply cannot influence the weather. Therefore, the data and plausibility support that stormy weather can be used as an instrumental variable.

Nevertheless, it makes more sense to use the supply in the form of total amount of received fish (totr) directly, because that is the omitted variable.

Re-run lin-log model with instrumental variables

```
y = daily_data$qty # linear
x = daily_data$price_log # log
main_model = lm(y~x)
b = coef(main_model)["x"]
se = summary(main_model)$coefficients["x", "Std. Error"]
```

First we run the lin-log model with stormy as an instrumental variable as proposed by Graddy:

```
# use stormy_weather as an instrumental variable:
v = daily_data$stormy
iv_reg_stormy = ivreg(formula = y ~ x | v)
b_iv = coef(iv_reg_stormy)["x"]
se_iv = summary(iv_reg_stormy)$coefficients["x", "Std. Error"]

# hausman test:
hausman = ((b_iv - b)^2)/(se_iv ^2 - se^2)
hausman = unname(hausman)
1-pchisq(hausman,df=1)
```

```
## [1] 0.176175
```

The Hausman test ($p = 0.176$) does not reject the null hypothesis of identical estimates. This implies that the instrument based on stormy weather conditions is not strong enough to provide a statistically significant improvement over OLS. As a result, the OLS estimate appears adequate for this dataset, and the evidence for price endogeneity is weak.

The stormy variable is a binary variable indicating if wave heigt and windspeed are above a certain threshold. This omits information. Therefore we now use windspeed directly as an instrumental variable. It fulfills the requirement of relevance as it is correlated with price:

```
cor(daily_data$price, daily_data$windspd)
```

```
## [1] 0.4150659
```

As windspeed also describes the weather the same argument as above holds regarding exogeneity.

```
# use totr as an instrumental variable:
v = daily_data$windspd
iv_reg_wind = ivreg(formula = y ~ x | v)
b_iv = coef(iv_reg_wind)["x"]
se_iv = summary(iv_reg_wind)$coefficients["x", "Std. Error"]

# hausman test:
hausman = ((b_iv - b)^2)/(se_iv ^2 - se^2)
hausman = unname(hausman)
1-pchisq(hausman,df=1)
```

```
## [1] 0.3155061
```

Again, the Hausman test ($p = 0.316$) indicates no evidence of price endogeneity. OLS and IV estimates do not differ significantly, so the OLS estimate of the demand function is likely consistent. Using wind speed as an instrument does not improve the model statistically.

Is there an endogeneity problem in the data? Do you see other endogeneity problems not captured by your instrument?

According to the Hausman tests using weather data as an instrument, we do not have an endogeneity problem in the data. But as seen above the total amount of received fish (totr), i.e. the supply, is an omitted variable that changes the results significantly. Therefore we conclude that we have an endogeneity problem, that is not captured by the instruments.