EDS_241_Assignment_4

Marie Rivers

03/04/2022

In this assignment, we estimate the price elasticity of demand for fresh sardines across 56 ports located in 4 European countries with monthly data from 2013 to 2019. Each row in the data file is a combination of port location (where the fish is landed and sold) in a given year and month. You can ignore the fact that the sample is not balanced (the number of monthly observations varies across ports).

Variables used in this assignment include: - year - month - country - port (where the sardines are landed and sold) - price_euro_kg (price per kg in euros) - volume_sold_kg (quantity of sardines sold in kg)

Note: `log()` is used to denote the natural logarithm

Read Data

```
data <- read_csv(here("data", "EU_sardines.csv"))
```

Question a:

Estimate the bivariate regression of `log(volume_sold_kg)` on `log(price_euro_kg)`. What is the price elasticity of demand for sardines? Test the null hypothesis that the price elasticity is equal to -1.

```
data <- data %>%
  mutate(log_price = log(price_euro_kg)) %>%
  mutate(log_volume_sold = log(volume_sold_kg))

model_a <- lm_robust(formula = log_volume_sold ~ log_price, data = data)
huxreg("log(volume_sold_kg)" = model_a)
```

```
model_a_summary <- tidy(model_a) %>%
  select(term, estimate, std.error, p.value, conf.low, conf.high) %>%
  kable()
model_a_summary
```

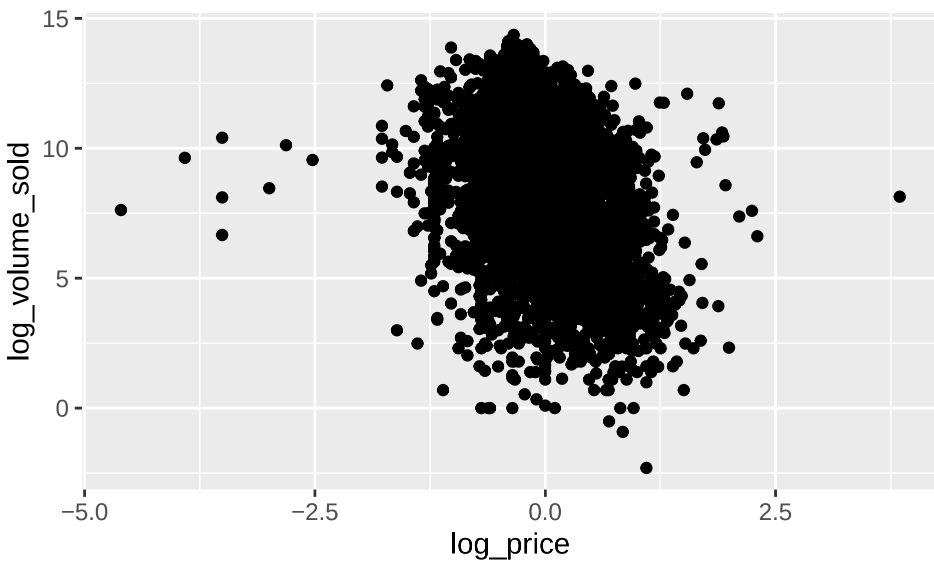
term	estimate	std.error	p.value	conf.low	conf.high
(Intercept)	7.759061	0.0430246	0	7.674709	7.843413
log_price	-1.545335	0.0781254	0	-1.698505	-1.392166

	log(volume_sold_kg)
(Intercept)	7.759 ***
	(0.043)
log_price	-1.545 ***
	(0.078)
N	3988
R2	0.104

*** p < 0.001; ** p < 0.01; * p < 0.05.

```r
log_price_coef_a <- round(model_a$coefficients[2], 2)
log_price_se_a <- round(model_a[[2]][2], 2)
```

```r
ggplot(data = data, aes(x = log_price, y = log_volume_sold)) +
  geom_point()
```



The price elasticity of demand for sardines based on a bivariate regression of `log(volume_sold_kg)` on `log(price_euro_kg)` is -1.55. Based on this value, it is estimated that a 1% increase in `log(price_euro_kg)` leads to a -1.55% change in `log(volume_sold_kg)`.

null hypothesis: The price elasticity is equal to -1.

$$\beta = -1$$

alternative hypothesis: The price elasticity is NOT equal to -1.

$$\beta \neq -1$$

```
ci_low <- model_a$conf.low[2]
ci_low
```

```
## log_price
## -1.698505
```

```
ci_high <- model_a$conf.high[2]
ci_high
```

```
## log_price
## -1.392166
```

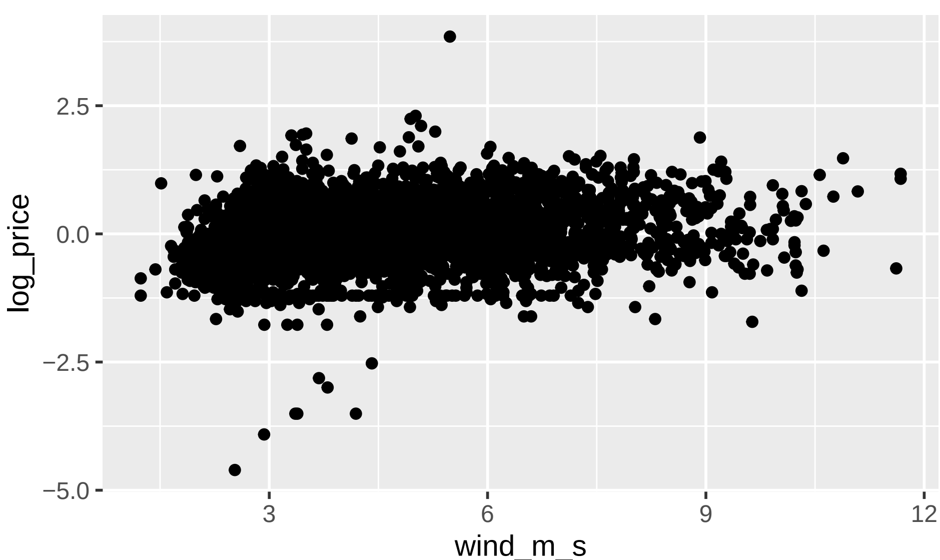
The 95% confidence interval ranges from `r`ci_low`to `ci_high`. This means that there is a 95% chance that this interval includes the true price elasticity value. Since the confidence interval does not contain -1, we reject the null hypothesis that price elasticity is equal to -1.

Question b:

Estimate the first-stage regression relating `log(price_euro_kg)` to `wind_m_s`. Interpret the estimated coefficient on wind speed. Does it have the expected sign? Also test for the relevance of the instrument and whether it is a "weak" instrument by reporting the proper F-statistic.

Note: See Lecture 8 and IV.R script. Use `wind_m_s` as an instrument for `log(price_euro_kg)`.

```
price_wind_plot <- ggplot(data = data, aes(x = wind_m_s, y = log_price)) +
  geom_point()
price_wind_plot
```




```
# first-stage regression
# in the first stage regression we get the change in price that is explained by wind
fs_b <- lm_robust(formula = log_price ~ wind_m_s, data = data)
huxreg("log(price)" = fs_b)
```

	log(price)
(Intercept)	-0.305 ***
	(0.027)
wind_m_s	0.067 ***
	(0.006)
N	3988
R2	0.038

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

```
# first-stage regression
wind_coef_b <- round(fs_b$coefficients[2], 2)
wind_se_b <- round(fs_b[[2]][2], 2)
```

The estimated coefficient on wind speed is 0.07. Based on this value, for each one unit increase in wind speed, the natural log of price increases by 0.07 percent. The estimated coefficient on wind speed does have the expected sign because it is likely that increased wind speed makes it harder to fish. When it is harder to fish, the supply of sardines could decrease and this would drive the price up.

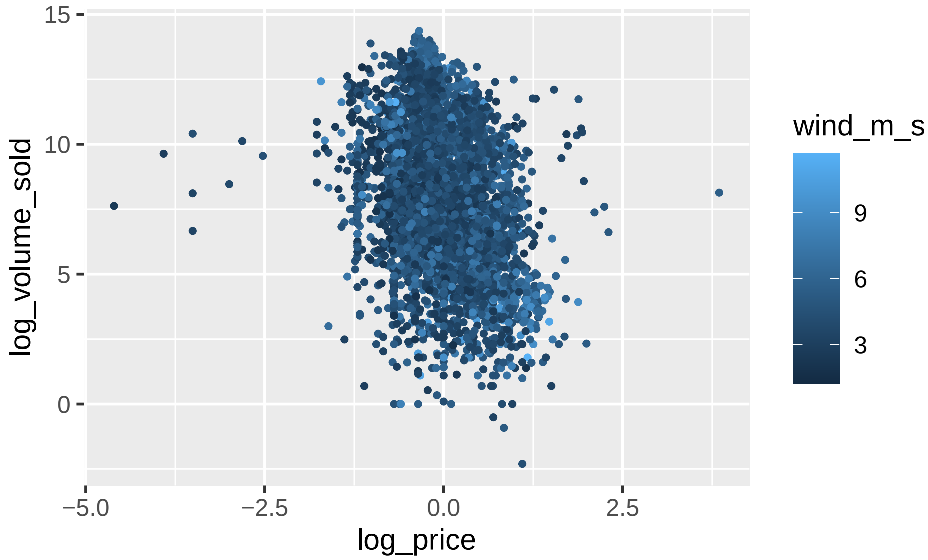
```
# F-test for non-weak and relevant instruments (Lecture 9, slides 13-14)
f_statistic_b <- linearHypothesis(fs_b, c("wind_m_s = 0"), white.adjust = "hc2")
f_statistic_b_value <- round(f_statistic_b$Chisq[2], 2)
```

The F-statistic for the instrument of wind in the model relating `log(price_euro_kg)` to `wind_m_s` is 144.65. Since this F-statistic is a lot greater than 10, wind speed is NOT a "weak" instrument. Wind speed is a valid instrument for explaining `log(price_euro_kg)`.

Question c:

Estimate the TSLS (two stage least squares) estimator of the price elasticity of demand for sardines using `wind_m_s` as an instrument for `log(price_euro_kg)`. What is the estimated price elasticity of demand for sardines?

```
ggplot(data = data, aes(x = log_price, y = log_volume_sold)) +
  geom_point(aes(color = wind_m_s), size = 0.75)
```




```r
# Lecture 8, slide 13
tsls_c <- ivreg(log_volume_sold ~ log_price | wind_m_s, data = data)
#summary(tsls_c)
huxreg("log(volume_sold)" = tsls_c)
```

	log(volume_sold)
(Intercept)	7.755 ***
	(0.043)
log_price	-1.088 **
	(0.370)
N	3988
R2	0.095

*** p < 0.001; ** p < 0.01; * p < 0.05.

```r
log_price_coef_c <- round(tsls_c$coefficients[2], 2)
```

The estimated price elasticity of demand for sardines is -1.09 based on the two-stage least squares (TSLS) estimator of the price elasticity of demand for sardines using wind speed as an instrument for the natural log of price. Based on this value, it is estimated that a 1% increase in `log(price_euro_kg)` leads to a -1.09% change in `log(volume_sold_kg)`.

Question d:

Repeat the exercise in (c), but include fixed effects for each year, month, and country. Report the estimated price elasticity of demand and the F-statistic testing for relevant and non-weak instruments.

Hint: you can use the command `as.factor(country) + as.factor(year) + as.factor(month)` to the `ivreg` function in R.

```
tsls_d <- ivreg(log_volume_sold ~ log_price + as.factor(year) + as.factor(month) + as.factor(country) |
#summary(tsls_d)
huxreg("log(volume_sold)" = tsls_d)
```

```
log_price_coef_d <- round(tsls_d$coefficients[2], 2)
```

The estimated price elasticity of demand is -1.25 when fixed effects for each year, month, and country are included in the two-stage least squares regression. Based on this value, it is estimated that a 1% increase in `log(price_euro_kg)` leads to a -1.25% change in `log(volume_sold_kg)`. The magnitude of the price elasticity for the fixed effect model is greater than for the model that just used wind as an instrument (part c) and less than for the model that didn't not include any instruments (part a).

```
# run just the first stage with the new regressors
fs_d <- lm_robust(log_price ~ wind_m_s + as.factor(year) + as.factor(month) + as.factor(country), data =
```

```
f_statistic_d <- linearHypothesis(fs_d, c("wind_m_s = 0"), white.adjust = "hc2")
f_statistic_d_value <- round(f_statistic_d$Chisq[2], 2)
```

The F-statistic for the instrument of wind in the model relating `log(price_euro_kg)` to `wind_m_s` is 77.66 with fixed effects for year, month, and country. Since this F-statistic is a lot greater than 10, wind speed is NOT a "weak" instrument. Wind speed is a valid instrument for explaining `log(price_euro_kg)`.

The F-statistic in part d is lower than the F-statistic in part b because we are now controlling for other things that explain price such as seasonality, country effects, and yearly fluctuations. There is less variation left for wind to explain, but wind still explains some of the variation in price.

xxx…Do we need to get F-statistic for all variables (year, month, country) or just wind? Do we need to identify which are strong and which are weak instruments?

	log(volume_sold)
(Intercept)	7.337 ***
	(0.208)
log_price	-1.250 **
	(0.464)
as.factor(year)2014	0.146
	(0.153)
as.factor(year)2015	0.185
	(0.152)
as.factor(year)2016	0.213
	(0.153)
as.factor(year)2017	0.074
	(0.152)
as.factor(year)2018	-0.091
	(0.155)
as.factor(year)2019	0.036
	(0.197)
as.factor(month)2	0.069
	(0.210)
as.factor(month)3	0.516 *
	(0.205)
as.factor(month)4	0.914 ***
	(0.203)
as.factor(month)5	1.149 ***
	(0.204)
as.factor(month)6	1.145 ***
	(0.202)
as.factor(month)7	1.400 ***
	(0.210)
as.factor(month)8	1.264 ***
	(0.217)
as.factor(month)9	1.311 ***
	(0.213)
as.factor(month)10	0.721 **