Anton Melnychuk ECON 3385 - Problem Set 2

January 26th, 2026

In [17]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.formula.api as smf

whale_data = pd.read_csv('whales_demand.csv')
whale_data.head()
```

Out[17]:

	year	price_sperm	price_oil	price_sperm_real	price_oil_real	sperm	oil	shipwrecks
0	1804	1.40	0.50	1.11	0.40	8.636	114.1065	0.064167
1	1805	0.96	0.50	0.68	0.35	6.390	42.1700	0.066090
2	1806	0.80	0.50	0.58	0.37	5.313	86.7370	0.067864
3	1807	1.00	0.50	0.77	0.38	0.270	65.1200	0.064410
4	1808	0.80	0.44	0.70	0.38	8.800	83.7500	0.064172

Question 1

a) For OLS to recover consistent estimates of β_1 when regressing equilibrium quantity on price, we need the error terms to be uncorrelated with price:

- $Cov(\epsilon_t^D, P_t^{oil}) = 0$ (demand shocks uncorrelated with price)
- $Cov(\epsilon_t^S, P_t^{oil}) = 0$ (supply shocks uncorrelated with price)

So that, if price moves independently of demand shocks, we can see how price affects quantity. Same regression will always recover consistent estimates of β_0.

b) Solving for equilibrium:

$$Q_y^{oil,D} = Q_y^{oil,S}$$

$$\beta_0 + \beta_1 P_y^{oil} + \epsilon_y^D = \gamma_0 + \gamma_1 P_y^{oil} + \epsilon_y^S$$

Rearranging:

$$(\beta_1 - \gamma_1) P_y^{oil} = \gamma_0 - \beta_0 + \epsilon_y^S - \epsilon_y^D$$

$$P_y^{oil} = \frac{\gamma_0 - \beta_0 + \epsilon_y^S - \epsilon_y^D}{\beta_1 - \gamma_1}$$

Since P_y^{oil} depends on both ϵ_y^S and ϵ_y^D, the conditions fail. Price is endogenous.

Question 2

Now we include shipwrecks (Z_y) in the supply equation:

Demand: $Q_y^{oil,D} = \beta_0 + \beta_1 P_y^{oil} + \epsilon_y^D$

Supply: $Q_y^{oil,S} = \gamma_0 + \gamma_1 P_y^{oil} + \gamma_2 Z_y + \epsilon_y^S$

where Z_y represents shipwrecks and ϵ_y^S captures other supply shifters.

Solving for equilibrium price:

$$P_y^{oil} = \frac{\gamma_0 + \gamma_2 Z_y - \beta_0 + \epsilon_y^S - \epsilon_y^D}{\beta_1 - \gamma_1}$$

Now price depends on shipwrecks Z_y in addition to the error terms.

Question 3

For 2SLS to recover consistent estimates, two conditions must hold:

1. Exclusion condition: $Cov(\epsilon_y^D, Z_y) = 0$

   - Shipwrecks must not affect demand directly
   - Shipwrecks only shifts supply

2. Relevance condition: $Cov(P_y^{oil}, Z_y) \neq 0$

   - Shipwrecks must affect price
   - From the equilibrium price equation, Z_y enters through γ_2

Why shipwrecks satisfies these:

From the equilibrium price equation, Z_y does not appear in the demand equation, so $Cov(\epsilon_y^D, Z_y) = 0$ holds. And, since P_y^{oil} depends on Z_y (through the $\gamma_2 Z_y$ term), shipwrecks affects price.

Question 4

Now regress price on shipwrecks to get predicted price $\hat{P_y^{oil}}$:

In [23]:

```
stage_1 = smf.ols('price_oil_real ~ shipwrecks', data=whale_data).fit()
print(stage_1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         price_oil_real   R-squared:                       0.266
Model:                            OLS   Adj. R-squared:                  0.259
Method:                 Least Squares   F-statistic:                     39.08
Date:                Mon, 26 Jan 2026   Prob (F-statistic):           8.30e-09
Time:                        12:35:59   Log-Likelihood:                 92.431
No. Observations:                 110   AIC:                            -180.9
Df Residuals:                     108   BIC:                            -175.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.2619      0.031      8.402      0.000       0.200       0.324
shipwrecks     2.5235      0.404      6.251      0.000       1.723       3.324
==============================================================================
Omnibus:                        5.480   Durbin-Watson:                   0.669
Prob(Omnibus):                  0.065   Jarque-Bera (JB):                4.881
Skew:                           0.450   Prob(JB):                       0.0871
Kurtosis:                       3.505   Cond. No.                         40.4
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specifi
ed.
```

Interpretation:

- **Intercept**: 0.2619 units expected real price when shipwrecks = 0
- **Shipwrecks coefficient**: 2.524 change in real price per 1 unit increase in shipwrecks. This makes sense, since 100% of ships lost would mean a significant increase in prices per barrel.

- **Positive coefficient** makes sense: more shipwrecks -> fewer ships -> less supply -> higher prices
- **F-statistic**: should be > 10 for a strong instrument -> good

Question 5

We regress quantity on predicted price from first stage

In [19]:

```
whale_data['p_hat_oil'] = stage_1.predict(whale_data)
stage_2 = smf.ols('oil ~ p_hat_oil', data=whale_data).fit()
print(stage_2.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                    oil   R-squared:                       0.055
Model:                            OLS   Adj. R-squared:                  0.046
Method:                 Least Squares   F-statistic:                     6.302
Date:                Mon, 26 Jan 2026   Prob (F-statistic):             0.0135
Time:                        12:13:36   Log-Likelihood:                -718.68
No. Observations:                 110   AIC:                             1441.
Df Residuals:                     108   BIC:                             1447.
Df Model:                           1                                         
Covariance Type:            nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    504.6173    114.904      4.392      0.000     276.857     732.378
p_hat_oil   -639.9404    254.911     -2.510      0.014   -1145.219    -134.662
==============================================================================
Omnibus:                       15.658   Durbin-Watson:                   0.125
Prob(Omnibus):                  0.000   Jarque-Bera (JB):                9.893
Skew:                           0.589   Prob(JB):                      0.00711
Kurtosis:                       2.121   Cond. No.                         19.1
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specifi
ed.
```

In [24]:

```
whale_data.describe()
```

Out[24]:

	year	price_sperm	price_oil	price_sperm_real	price_oil_real	sperm	oil
count	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000
mean	1858.500000	0.965320	0.480067	0.905273	0.446364	60.928918	218.971159
std	31.898276	0.426035	0.225719	0.295013	0.122427	50.578117	171.985297
min	1804.000000	0.400000	0.235000	0.420000	0.250000	0.270000	12.280000
25%	1831.250000	0.651250	0.338125	0.680000	0.350000	17.794000	80.013250
50%	1858.500000	0.850000	0.399500	0.840000	0.430000	45.036500	144.246500
75%	1885.750000	1.245000	0.538750	1.100000	0.510000	94.882500	349.836500

	year	price_sperm	price_oil	price_sperm_real	price_oil_real	sperm	oil
max	1913.000000	2.550000	1.450000	1.610000	0.780000	186.219000	594.675000

Question 6

Short summary:

Parameter	Coefficient	Standard Error
β_0 (Intercept)	504.62	114.90
β_1 (Price)	-639.94	254.91

- $\beta_1 = -639.94$ is negative, which is the correct sign for a demand curve. Higher prices lead to lower quantities demanded, consistent with the law of demand.
- In other words, a 1-unit increase in real price (in the units of the price variable) reduces quantity demanded by approximately 640 units. This means, the coefficient is economically meaningful. Small price increase -> a substantial decrease in quantity demanded.
- β_1 is statistically significant at the 5% level ($p < 0.05$), meaning we can reject the null hypothesis that price has no effect on quantity with 95% confidence. This provides strong evidence that the negative relationship between price and quantity is not due to random chance.
- The low $R^2 = 0.055$ (5.5% can be explained by instrument) is expected—the focus is on causal identification rather than model fit. The 2SLS estimate successfully identifies the demand curve.

Question 7

Now we run 2SLS on sperm whale oil.

In [25]:

```
# Reapeat Q4-Q6 for sperm whale oil
first_stage_sperm = smf.ols('price_sperm_real ~ shipwrecks', data=whale_data).fit()

whale_data['p_hat_sperm'] = first_stage_sperm.predict(whale_data)

second_stage_sperm = smf.ols('sperm ~ p_hat_sperm', data=whale_data).fit()

print("First Stage (Sperm):")
print(first_stage_sperm.summary())
print("\nSecond Stage (Sperm):")
print(second_stage_sperm.summary())
```

```
First Stage (Sperm):
                            OLS Regression Results
==============================================================================
Dep. Variable:       price_sperm_real   R-squared:                       0.169
Model:                            OLS   Adj. R-squared:                  0.162
Method:                 Least Squares   F-statistic:                     22.00
Date:                Mon, 26 Jan 2026   Prob (F-statistic):           8.01e-06
```

```
Time:                        12:56:52   Log-Likelihood:                -11.101
No. Observations:                 110   AIC:                             26.20
Df Residuals:                     108   BIC:                             31.60
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.5505      0.080      6.890      0.000       0.392       0.709
shipwrecks     4.8534      1.035      4.691      0.000       2.802       6.904
==============================================================================
Omnibus:                        2.947   Durbin-Watson:                   0.291
Prob(Omnibus):                  0.229   Jarque-Bera (JB):                2.915
Skew:                           0.352   Prob(JB):                        0.233
Kurtosis:                       2.624   Cond. No.                         40.4
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specifi
ed.

Second Stage (Sperm):
                            OLS Regression Results
==============================================================================
Dep. Variable:                  sperm   R-squared:                       0.103
Model:                            OLS   Adj. R-squared:                  0.095
Method:                 Least Squares   F-statistic:                     12.41
Date:                Mon, 26 Jan 2026   Prob (F-statistic):           0.000627
Time:                        12:56:52   Log-Likelihood:                -581.19
No. Observations:                 110   AIC:                             1166.
Df Residuals:                     108   BIC:                             1172.
Df Model:                           1
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept     182.0424     34.684      5.249      0.000     113.292     250.793
p_hat_sperm  -133.7867     37.977     -3.523      0.001    -209.064     -58.509
==============================================================================
Omnibus:                       11.971   Durbin-Watson:                   0.153
Prob(Omnibus):                  0.003   Jarque-Bera (JB):                5.696
Skew:                           0.337   Prob(JB):                       0.0580
Kurtosis:                       2.113   Cond. No.                         15.1
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specifi
ed.
```

Short summary:

Parameter	Coefficient	Standard Error
β_0 (Intercept)	182.04	34.68
β_1 (Price)	-133.79	37.98

- $\beta_1 = -133.79$ is negative and significant at the 1% level ($p < 0.01$), consistent with the law of demand: higher prices lead to lower quantities demanded.
- A 1-unit increase in the real price of sperm whale oil reduces quantity demanded by approximately 134 units.
- The elasticity exceeds that of regular oil (-1.3), showing that sperm oil consumers are more responsive to price changes.
- The $R^2 = 0.103$ indicates that roughly 10% of the variation in sperm whale oil quantity is accounted for by the instrumented price – nearly double the R^2 for regular oil (5.5%).

The coefficient magnitude differs from regular oil, but the scales are different (mean quantity for oil ≈ 219 vs sperm ≈ 61). The price elasticity of sperm oil is higher than regular oil, indicating more price-sensitive demand. This makes sense as sperm is a premium good with more substitutes available.

If non-sperm and sperm oil are substitutes, an increase in sperm price should increase demand for regular oil (positive cross-price effect). To model this, we could extend the demand equation:

$$Q_t^{oil} = \beta_0 + \beta_1 P_t^{oil} + \beta_2 P_t^{sperm} + \epsilon_t^D$$

This would require instrumenting both prices with supply shifters.