OLS Examples

OLS Regression

• Problem
  – The Kelley Blue Book provides information on wholesale and retail prices of cars. Following are age and price data for 10 randomly selected Corvettes between 1 and 6 years old. Here, age is in years, and price is in hundreds of dollars.

<table>
<thead>
<tr>
<th>age</th>
<th>6</th>
<th>6</th>
<th>6</th>
<th>2</th>
<th>2</th>
<th>5</th>
<th>4</th>
<th>5</th>
<th>1</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td>205</td>
<td>195</td>
<td>210</td>
<td>340</td>
<td>299</td>
<td>230</td>
<td>270</td>
<td>243</td>
<td>340</td>
<td>240</td>
</tr>
</tbody>
</table>
### OLS Regression

#### Coding Sheet for Corvette Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Possible Values</th>
<th>Source</th>
<th>Mnemonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of Corvettes</td>
<td>Years</td>
<td>Kelley Blue Book, various issued</td>
<td>age</td>
</tr>
<tr>
<td>Price of Corvettes</td>
<td>Hundred of Dollars, Nominal</td>
<td>IBID.</td>
<td>price</td>
</tr>
</tbody>
</table>
Note from Stat 101:

\[ s = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n-1}} = \sqrt{\frac{S_{YY}}{n-1}} = \sqrt{\frac{25681.6}{9}} = 53.41827 \]
### OLS Specification

Form for OLS command:

**LS dependent variable C independent variable(s)**

or

Equation `eq1.LS` **dependent variable C independent variable(s)**

<table>
<thead>
<tr>
<th>Section 1</th>
<th>Section 2</th>
<th>Section 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1</td>
<td>Section 2</td>
<td>Section 3</td>
</tr>
</tbody>
</table>
Equation for $\hat{Y}$:

Estimated Price $= 371.6019 - 27.90291 \times \text{AGE}$
SSE = Sum of Squared Residuals = 1623.709
s = Standard Error of Regression = 14.2463

Note:

\[ s^2 = \frac{SSE}{n-2} = \frac{1623.709}{10-2} = 202.9636 \]
\[ s = \sqrt{202.9636} = 14.2463 \]

Standard Errors of Estimates

Note:

\[ s_{AGE} = \frac{s}{\sqrt{S_{xx}}} \]
\[ s_{AGE} = \frac{14.2463}{\sqrt{30.9}} = 2.562889 \]
Both Highly Significant

\[ t_{\text{AGE}} = \frac{-27.90291}{2.562889} = -10.88729 \]
No Explanatory Variable

Sum is zero

SSE
\[ \text{SST} = \sum (Y - \bar{Y})^2 = S_{YY} = 25681.60 \]

Note:

\[ R^2 = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSE}}{S_{YY}} \]

\[ = 1 - \frac{1623.709}{25681.60} \]

\[ = 0.936775 \]

Note:

\[ \eta_{\text{Price}} = \frac{4.1}{257.2} \times -27.90291 \]

\[ = -0.444798 \]

Very Inelastic
OLS Regression

Elasticity Summary Table

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Mean</th>
<th>Elasticity</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>-27.90291</td>
<td>4.1</td>
<td>-0.444798</td>
<td>Inelastic</td>
</tr>
</tbody>
</table>

Interpretation: Price of a corvette is inelastic with respect to the age of the corvette so that a 1% increase in age decreases the price by only 0.4%. Other factors are at play regarding the lower prices, but age is certainly a major factor as evidenced by the $R^2$ of 0.94.

Other Examples

Problem 1: CAPM
Problem 1: CAPM

• Problem
  – Estimate $\beta$, the systematic risk, for CAPM
  • CAPM is
    \[
    E(r_i \mid \Omega_t) = r_f + \beta_i [E(r_m \mid \Omega_t) - r_f]
    \]
  • An empirical version is
    \[
    \tilde{r}_i = \alpha + \beta_i \tilde{r}_m + \varepsilon_i
    \]
    $\varepsilon \sim N(0, \sigma^2)$
    \[
    \tilde{r}_i = r_i - r_f
    \]
    \[
    \tilde{r}_m = r_m - r_f
    \]
    $\alpha = 0$
Problem 1: CAPM

- *beta* is proportionality factor
  \[ \beta_i = \frac{E(r_i \mid \Omega_t) - r_f}{E(r_m \mid \Omega_t) - r_f} \]
  - Excess returns for security over riskless return is proportional to excess returns in market over riskless returns
  - Excess is extra return to compensate for risk for not holding the market portfolio
    - Premium on security is proportional to premium on the market portfolio

Problem 1: CAPM

- If
  \[ \beta_i = \frac{E(r_i \mid \Omega_t) - r_f}{E(r_m \mid \Omega_t) - r_f} > 1 \Rightarrow \text{Premium}_i > \text{Premium}_m \]
  then asset \( i \) is viewed as riskier than the market
Problem 1: CAPM

<table>
<thead>
<tr>
<th>beta</th>
<th>Interpretation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Moves With Market</td>
<td>Conglomerates (AT&amp;T)</td>
</tr>
<tr>
<td>&gt;1</td>
<td>More Volatile</td>
<td>Companies Sensitive To Macro Events (Autos)</td>
</tr>
<tr>
<td>0&lt;beta&lt;1</td>
<td>Less Volatile</td>
<td>Mildly Sensitive To Macro Events (Electric Utilities)</td>
</tr>
<tr>
<td>&lt;0</td>
<td>Move Counter To Market</td>
<td>Goldmining Stocks</td>
</tr>
</tbody>
</table>

Coding Sheet for CAPM Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Possible Values</th>
<th>Source</th>
<th>Mnemonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess returns on an index of 104 stocks in the cyclical consumer goods sector in the UK. Monthly, 1/80-12/99.</td>
<td>Percentages*</td>
<td>IBID.</td>
<td>rendcyco</td>
</tr>
<tr>
<td>Excess returns on an index of 104 stocks in the noncyclical consumer goods sector in the UK. Monthly, 1/80-12/99.</td>
<td>Percentages*</td>
<td>IBID.</td>
<td>rendncco</td>
</tr>
<tr>
<td>Excess returns on an index of 104 stocks in the information tech sector in the UK. Monthly, 1/80-12/99.</td>
<td>Percentages*</td>
<td>IBID.</td>
<td>rendit</td>
</tr>
<tr>
<td>Excess returns on an index of 104 stocks in the telcom sector in the UK. Monthly, 1/80-12/99.</td>
<td>Percentages*</td>
<td>IBID.</td>
<td>rendtel</td>
</tr>
</tbody>
</table>

*Note: See calculations note.
Problem 1: CAPM

- Excess returns are calculated as follows
  - Let \( p_i \) be the closing price of the index at the last trading day in month \( i \) and let \( r_i \) be the one-month interest rate at the start of month \( i \). Then the return \( v_i \) of the index is \( v_i = (p_i - p_{i-1})/p_{i-1} \) and the excess return is \( v_i - r_i \). The reported numbers are 100\((v_i - r_i)\).
Excess Returns Distributions
Stock Market Sectors (UK)
1980 - 1999

Note: Monthly data

Excess Returns (%)

Excess Returns
Overall Stock Market Index (UK)
1980 - 1999

Note: Monthly data
Excess Returns
Overall Stock Market Index (UK)
1980 - 1999

Series: RENDMARK
Sample 1980M01 1999M12
Observations 240

Mean       0.808884
Median   1.204026
Maximum  13.46098
Minimum -27.86969
Std. Dev.   4.755913
Skewness  -1.157801
Kurtosis   8.374932
Jarque-Bera  342.5191
Probability  0.000000

Note: Monthly data
Excess Returns
Cyclical Consumer Goods Sector (UK)
1980 - 1999

Series: RENDCYCO
Sample 1980M01 1999M12
Observations 240

Mean 0.499826
Median 0.309320
Maximum 22.20496
Minimum -35.56618
Std. Dev. 7.849594
Skewness -0.230900
Kurtosis 4.531597
Jarque-Bera 25.59050
Probability 0.000003

Excess Returns
Noncyclical Consumer Goods Sector (UK)
1980 - 1999

Note: Monthly data
Excess Returns
Information Technology Sector (UK)
1980 - 1999

Note: Monthly data

Excess Returns
Telecommunications Sector (UK)
1980 - 1999

Note: Monthly data
Problem 1: CAPM

- Interpretation
  - The intercept is highly insignificant suggesting that the line goes through the origin.
  - The slope is highly significant and positive at the 0.0 level indicating that the market rate of return (excess returns) have a positive effect on the returns for the cyclical consumer goods sector in the UK.
  - The R² is 0.50 suggesting that 50% of the variation in returns in the cyclical consumer goods sector is accounted for by the market excess returns.
  - The model is significantly different from the naïve model as indicated by the p-value for F.
Problem 1: CAPM

- Interpretation
  - Since this is a CAPM, the slope has the interpretation of systematic risk
    - Since it is greater than 1.0, this indicates that the cyclical consumer goods sector is riskier than the market as a whole
    - When the market rises, the excess returns in this sector will rise more than the market

<table>
<thead>
<tr>
<th>Model Portfolio</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.477 (0.2188)</td>
</tr>
<tr>
<td>RendMark</td>
<td>1.17* (0.0000)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.5035</td>
</tr>
<tr>
<td>$F_C$</td>
<td>241.3363 (0.0000)</td>
</tr>
</tbody>
</table>

Notes: p-value in parentheses; *=significant
Other Examples

Problem 2: Bank Wages
Problem 2: Bank Wages

• Problem
  – Find the relationship between the salary of employees at a major US bank and their years of education completed

| Coding Sheet for Bank Salary Data |
|-----------------------------|-----------------|------------------|------------------|
| Variable                     | Possible Values | Source           | Mnemonic         |
| Current yearly salary of bank employees in US | Nominal dollars | SPSS, version 10 (2000) | salary          |
| Number of years of education completed | Years           | IBID.            | educ             |
| Natural log of salary        | logs            | Calculated as natural log of salary | logsalary      |
Distribution of Current Yearly Salary Bank Employees (US)

Sample 1 474 F GENDER=1
Observations 258

Mean 41441.78
Median 32850.00
Maximum 135000.0
Minimum 19650.00
Std. Dev. 19499.21
Skewness 1.629931
Kurtosis 5.702920
Jarque-Bera 192.7741
Probability 0.000000
Natural Log of Current Year Salary

Distribution of Log of
current Yearly Salary
Bank Employees (US)

Series: LOGSALARY
Sample 1474 IF GENDER=1
Observations 258

Mean       10.54461
Median   10.39967
Maximum  11.81303
Minimum  9.885833
Std. Dev.   0.398579
Skewness   0.840303
Kurtosis   2.805992
Jarque-Bera  30.76731
Probability  0.000000
Problem 2: Bank Wages

- Our model is
  \[ \text{Salary}_i = Ae^{\beta_1 \text{Educ}_i}e^{\varepsilon_i} \]
  or
  \[ \ln(\text{SALARY}_i) = \beta_0 + \beta_1 \text{EDUC}_i + \varepsilon_i \]
- Interpret the slope parameter as the percentage increase in salary (S) due to one additional year of education
  \[ \frac{d \ln(S)}{dx} = \frac{dS/S}{dx} \]
  Salary rises 9% for each additional year of education
Problem 2: Bank Wages

• Interpretation
  – The intercept and slope parameters are highly significant at the 0.0 level
    • This indicates that education has a significant, positive effect on salary
      – The higher the level of education, the higher the salary
    • For each extra 1 year of education, salary rises 9%
  – Almost 49% of the variation in (log) salary is accounted for by education
  – The model is significantly different from the naïve model as indicated by the p-value for F

Problem 2: Bank Wages

<table>
<thead>
<tr>
<th>Model Portfolio</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>9.22*</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Educ</td>
<td>0.092*</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.4680</td>
</tr>
<tr>
<td>$F_C$</td>
<td>225.2383</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
</tr>
</tbody>
</table>

Notes: p-value in parentheses; *=significant
Problem 3: Population and Economic Growth

- Population growth is an important factor for long-term economic growth
  - Malthus assumed (correctly) that population will grow geometrically (e.g., 1, 3, 9, 27, 81, …) while the food supply will increase arithmetically (e.g., 10, 10, 30, 40, …)
    - Population would double every 24 years
  - As a result of the faster growth of population, population growth would overtake food supply growth – but he didn’t specify when
Problem 3: Population and Economic Growth

- Digression on compound growth
  - Compounding
    \[ FV_t = PV_0 (1 + g)^t \]
  - Time to double
    - What is the implied population growth rate, \( g \), for Malthus?
    - How long to double at 1.9%?
Problem 3: Population and Economic Growth

- Current population projections
  - The Year 2000 population growth rate, \( g \), was about 1.4%
    - World population in 2000 was about 6.1 Billion people
    - Absolute growth: 85.4 Million in one year
Problem 3: Population and Economic Growth

- The future
  - There is some hope ahead
    - The UN forecasts a drop in fertility in developing countries
    - Developed countries will grow by less than 1% annually
    - Predicted world’s population in 2300
      - Old: 12 Billion
      - New: 9 Billion
Problem 3: Population and Economic Growth

- Malthus had checks on population growth
  - Contraceptives
  - Abortion
  - Self-restraint
  - Wars

- Malthus wrote right at the time of the Industrial and Green Revolutions, so he could not see their implications

Problem 3: Population and Economic Growth

- Data from the World Development Indicators, World Bank (2004)
  - Calculated average annual growth for GDP and population

\[ g_i = \left( \frac{\text{Value in 2003}_i}{\text{Value in 1963}_i} \right)^{\frac{1}{40}} - 1, \ i = 1, \ldots, 107 \text{ countries} \]
<table>
<thead>
<tr>
<th>Variable</th>
<th>Possible Values</th>
<th>Source</th>
<th>Mnemonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real GDP in country (1995 SUS), 1963 and 2003</td>
<td>Real dollars</td>
<td>World Development Indicators (World Bank, 2004)</td>
<td>gdp63, gdp03</td>
</tr>
<tr>
<td>Average annual growth in GDP</td>
<td>Decimal values</td>
<td>Calculated</td>
<td>g</td>
</tr>
<tr>
<td>Population in country, 1963 and 2003</td>
<td>Count</td>
<td>World Development Indicators (World Bank, 2004)</td>
<td>pop63, pop03</td>
</tr>
<tr>
<td>Average annual growth in population</td>
<td>Decimal values</td>
<td>Calculated</td>
<td>p</td>
</tr>
</tbody>
</table>
Real GDP and Population Average Annual Growth
1963 - 2003

Highly insignificant

Terrible $R^2$
Real GDP and Population Average Annual Growth
1963 - 2003

Real GDP Average Annual Growth

Population Average Annual Growth by Country (n = 107)