Initial Question

*Is it sensible to purchase bottles of wine as an investment?*

- *vintage* - harvest year for wine grape
- bottled wine - often resold at auctions

- How would you determine if it is sensible to purchase wine as an investment?
Stochastic Model

- $n$ vintages of Bordeaux wines indexed by $i$
  - $P_i$ - log price of 1 bottle of vintage $i$
  - $A_i$ - age of wine (auction year - vintage)
  - $U_i$ - other forces that affect wine prices

Stochastic Model

$$P_i = \beta_0 + \beta_1 A_i + U_i$$

- $\beta_0$ - expected price of a newly released vintage
- $\beta_1$ - change in price for each year the wine ages

Potentially attractive investment: $\beta_1 > 0$
$b_1 = 3.5$

- Does this mean that the price of wine increases by 3.5% each year?
  - wine auctions that are not sampled may differ
  - wines from regions other than Bordeaux may differ

- these facts indicate that $b_1$ would take a different value for each sample

- if we had many samples, and averaged the values of $b_1$ obtained from these samples, this average would tend toward $\beta_1$
OLS Estimator is a Random Variable

- Deviations-from-means form

\[ B_1 = \frac{\sum_{i=1}^{n} X_i^* Y_i^*}{\sum_{i=1}^{n} (X_i^*)^2} \]

- \( X_i^* = X_i - \bar{X} \)
- \( Y_i^* = Y_i - \bar{Y} = \beta_1 (X_i - \bar{X}) + U_i - \bar{U} \)
- \( U_i^* = U_i - \bar{U} \)

\[ B_1 = \frac{\sum_{i=1}^{n} X_i^* (\beta_1 X_i^* + U_i^*)}{\sum_{i=1}^{n} (X_i^*)^2} \]

\[ B_1 = \beta_1 + \frac{\sum_{i=1}^{n} X_i^* U_i^*}{\sum_{i=1}^{n} (X_i^*)^2} \]
Framework: Expectation Operator Properties

- **Linear operator?** yes
  \[ \mathbb{E} \left( \beta_1 + \frac{\sum_{i=1}^{n} X_i^* U_i^*}{\sum_{i=1}^{n} (X_i^*)^2} \right) = \beta_1 + \mathbb{E} \left( \frac{\sum_{i=1}^{n} X_i^* U_i^*}{\sum_{i=1}^{n} (X_i^*)^2} \right) \]

- **Nonlinear operator?** no
  \[ \mathbb{E} \left( \frac{\sum_{i=1}^{n} X_i^* U_i^*}{\sum_{i=1}^{n} (X_i^*)^2} \right) \neq \frac{\mathbb{E} \left( \sum_{i=1}^{n} X_i^* U_i^* \right)}{\mathbb{E} \left( \sum_{i=1}^{n} (X_i^*)^2 \right)} \]

- We need to compute \( \mathbb{E} \left( \frac{\sum_{i=1}^{n} X_i^* U_i^*}{\sum_{i=1}^{n} (X_i^*)^2} \right) \)
Conditional Expectation

- $\mathbb{E}\left( \frac{Z}{W} \right)$ hard to compute
- $\mathbb{E}\left( \frac{Z}{W} \mid W \right)$ easier to compute
  - expectation conditional on (given) the value of $W$
  - $W$ is considered fixed at the value $w$

- the conditional expectation is

$$\mathbb{E}\left( \frac{Z}{W} \mid W \right) = \mathbb{E}\left( \frac{Z}{W = w} \right)$$
Conditional Expectation - OLS Estimator

\[ \mathbb{E}(B_1) = \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^{n} X_i^* U_i^*}{\sum_{i=1}^{n}(X_i^*)^2}\right) \]

- We condition on \( X_1^* = x_1^*, X_2^* = x_2^*, \ldots, X_n^* = x_n^* \)

\[ \mathbb{E}(B_1|X_1^*, \ldots, X_n^*) = \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^{n} X_i^* U_i^*}{\sum_{i=1}^{n}(X_i^*)^2}|X_1^*, \ldots, X_n^*\right) \]
\[ = \beta_1 + \frac{\sum_{i=1}^{n} x_i^* \mathbb{E}(U_i^*|X_1^*, \ldots, X_n^*)}{\sum_{i=1}^{n}(x_i^*)^2} \]

- notation \( \mathbb{E}(B_1|X^*) \equiv \mathbb{E}(B_1|X_1^*, \ldots, X_n^*) \)

- If we assume \( \mathbb{E}(U_i^*|X^*) = 0 \)

\[ \mathbb{E}(B_1|X_1^*, \ldots, X_n^*) = \beta_1 \]
Law of Iterated Expectations

- Suppose \( W \) takes 2 values, \( \{w_1, w_2\} \) where
  - \( \mathbb{E}(Z \mid W = w_1) = 0 \)
  - \( \mathbb{E}(Z \mid W = w_2) = 0 \)

- What is \( \mathbb{E}(Z) \)?
  - \( \mathbb{E}(Z) = \mathbb{E}(Z \mid W = w_1) \cdot P(W = w_1) + \mathbb{E}(Z \mid W = w_2) \cdot P(W = w_2) \)
  - right side is \( \mathbb{E}_W \) of \( \mathbb{E}(Z \mid W) \)

- Law of Iterated Expectations
  \[
  \mathbb{E}(Z) = \mathbb{E}_W(\mathbb{E}(Z \mid W))
  \]

- If \( \mathbb{E}(Z \mid W) = c \) for all values of \( W \), what is \( \mathbb{E}(Z) \)?
  - \( \mathbb{E}(Z) = c \)
(Mean) Bias

\[ \text{Bias} = \mathbb{E}(B_1) - \beta_1 \]

if \( \mathbb{E}(U_i^*|X^*) = 0 \), then

- \( \mathbb{E}(B_1|X^*) = \beta_1 \) for all values of \( X^* \)
- \( \Rightarrow \mathbb{E}(B_1) = \beta_1 \)
- \( \Rightarrow \text{Bias} = 0 \)

the OLS estimator is (mean) unbiased, so we have no reason to believe our estimate of 3.5\% is systematically too low or too high
The median of a random variable is the midpoint of the outcomes

- Suppose $W$ takes the values $\{5, 10, 15\}$
  - $med(W) = 10$
- Suppose $W$ takes the values $\{5, 10, 15, 20\}$
  - $med(W) \in (10, 15)$  not unique
- Median is a robust measure of location
  - for $\{5, 10, 15\}$  \quad $med(W) = mean(W) = 10$
  - for $\{5, 10, 150\}$  \quad $med(W) = 10$  $mean(W) = 55$
Estimator Bias
Median Bias

- $B_1$ is mean unbiased if
  \[
  \mathbb{E}(B_1) = \beta_1 \\
  \text{Bias} = \mathbb{E}(B_1) - \beta_1
  \]

- $B_1$ is median unbiased if
  \[
  \text{med}(B_1) = \beta_1 \\
  \text{median Bias} = \text{med}(B_1) - \beta_1
  \]
Unbiased Estimators

- What assumption is needed to ensure the OLS estimator is unbiased?
  - \( \mathbb{E}(U_i^* | X^*) = 0 \)
- Does \( \mathbb{E}(U_i^* | X^*) = 0 \) imply \( \mathbb{E}(U_i^* | X_j^*) = 0 \) for \( j = 1, \ldots, n \)?
  - Yes
- Is \( \mathbb{E}(U_i^* | X_i^*) = 0 \) sufficient for the OLS estimator to be unbiased?
  - No. The condition must hold for all \( X_j^* \), not merely for \( j = i \)
- Is \( \mathbb{E}(X_i^* U_i^*) = 0 \) sufficient for the OLS estimator to be unbiased?
  - No. This condition doesn’t even imply \( \mathbb{E}(U_i^* | X_i^*) = 0 \)