Section 4. Multiple Regression: Assumptions and Implications

1. **The LS assumptions for Multiple Regression**
2. **The two innocuous assumptions**
3. **The additional assumption on X's**
4. **Why assume Linear Population Regression?**
5. **Why optional homoskedasticity assumption?**

1.Multiple Regression

WHAT IF THE POPULATION OF POINTS IS NOT "LINEAR"?

The Multiple Regression Model

The multiple regression model is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i,\ i = 1, \ldots, n. \qquad (5.7)$$

where:

- Y_i is i^{th} observation on the dependent variable; $X_{1i}, X_{2i}, \ldots, X_{ki}$ are the i^{th} observations on each of the k regressors; and u_i is the error term.
- The population regression line is the relationship that holds between Y and the X's on average in the population:

\# →

$$E(Y \mid X_{1i} = x_1, X_{2i} = x_2, \ldots, X_{ki} = x_k) = \beta_0 + \underline{\beta_1 x_1} + \underline{\beta_2 x_2 + \cdots + \beta_k x_k}.$$

$E(u|X)=0$

- β_1 is the slope coefficient on X_1, β_2 is the coefficient on X_2, etc. The coefficient β_1 is the expected change in Y_i resulting from changing X_{1i} by one unit, holding constant $X_{2i}, \ldots, X_{ki}$. The coefficients on the other X's are interpreted similarly.
- The intercept β_0 is the expected value of Y when all the X's equal zero. The intercept can be thought of as the coefficient on a regressor, X_{0i}, that equals one for all i.

1a. OLS in Multiple Regression

The OLS Estimators, Predicted Values, and Residuals in the Multiple Regression Model

The OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are the values of $b_0, b_1, \ldots, b_k$ that minimize the sum of squared prediction mistakes $\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1i} - \cdots - b_k X_{ki})^2$. The OLS predicted values $\hat{Y}_i$ and residuals $\hat{u}_i$ are:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_k X_{ki},\ i = 1, \ldots, n, \text{ and} \tag{5.11}$$

$$\hat{u}_i = Y_i - \hat{Y}_i,\ i = 1, \ldots, n. \tag{5.12}$$

The OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ and residual $\hat{u}_i$ are computed from a sample of n observations of $(X_{1i}, \ldots, X_{ki}, Y_i)$, $i = 1, \ldots, n$. These are estimators of the unknown true population coefficients $\beta_0, \beta_1, \ldots, \beta_k$ and error term, u_i.

1b. Four MR Assumptions

The Least Squares Assumptions in the Multiple Regression Model

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$, $i = 1, \ldots, n$, where:

NOT CRITICAL

1. u_i has conditional mean zero given $X_{1i}, X_{2i}, \ldots, X_{ki}$, that is, $E(u_i | X_{1i}, X_{2i}, \ldots, X_{ki}) = 0$;
2. $(X_{1i}, X_{2i}, \ldots, X_{ki}, Y_i)$, $i = 1, \ldots, n$ are independently and identically distributed (i.i.d.) draws from their joint distribution;
3. $(X_{1i}, X_{2i}, \ldots, X_{ki}, u_i)$ have nonzero finite fourth moments; and
4. there is no perfect multicollinearity.

2. Two fairly innocuous assumptions

- Why worry about assumptions?
  Can opener joke
  One-armed economist joke
  Tools are more useful if assumptions are few and plausible
- #2) Random sampling
- #3) Finite 4^{th} moments

$Y = b_0 + b_1 X_1 + b_2 X_2 + e$

mean income

3. Multicollinearity: The additional assumption on X's

- Now that we have additional X's, the correlation of X's matters

  We saw that in the OVB formula
- If X's are perfectly correlated the OLS coefficients are undefined
- How would you know if the assumption were violated.. Stata will let you know.
- So this is also pretty innocuous.

4. Why assume (#1), linear pop. regression?

- What if population regression is not linear?
  - E.g. from car prices
- What does LS estimate if E(u|X) ≠ 0 ?

$$\beta^{BLP} = \frac{Cov(x, y)}{V(x)}$$

- Where did we need linear E(u|X) = 0 ?

BEST LINEAR PREDICTOR.

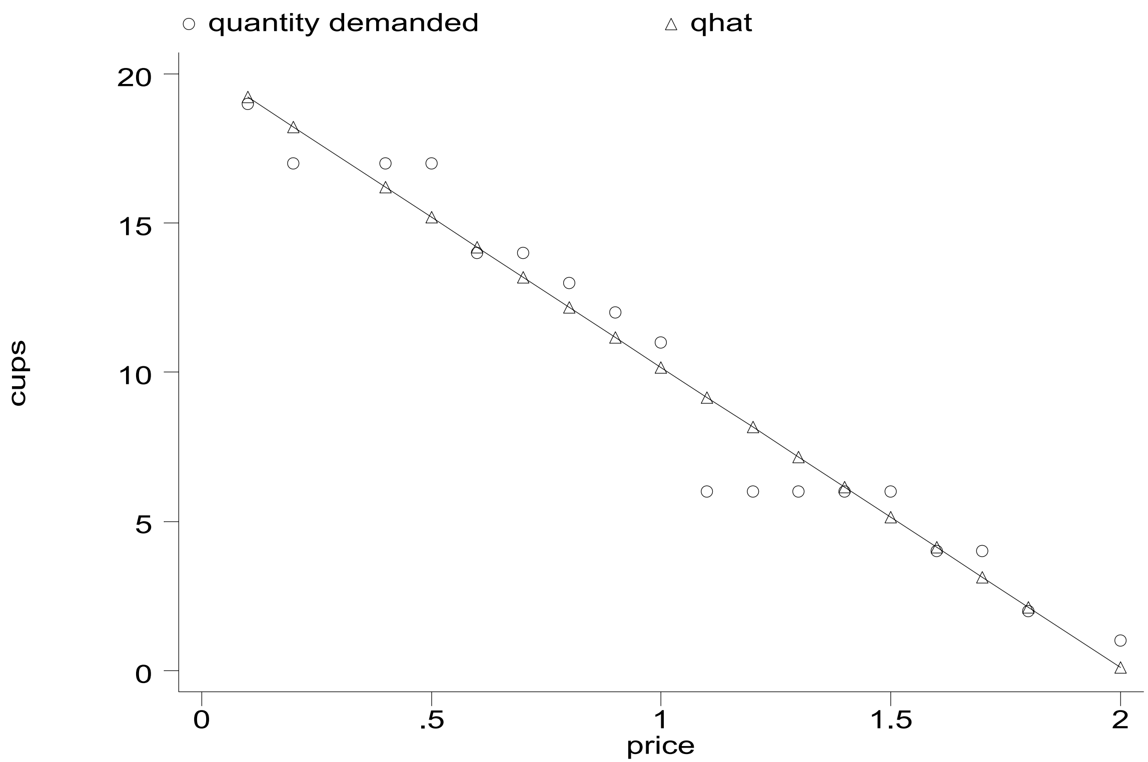
  - For unbiased estimates

  Why do we need unbiased estimates?

  - We can get consistency and CI even without assumption #1.

  What are we estimating if

Coffee Demand – Looks linear



5. Why Assume Homoskedasticity? A) Efficiency

- $V(u_i)$ doesn't depend on i
- Examples
- Why assume that?

  - Gauss-Markov Theorem tells you that OLS estimators $b_0 \ldots b_k$ are minimum variance (among unbiased estimates) if you assume homoskedasticity and linear pop. Regression in addition to other assumptions.

NICE RESULT BUT HOMOSKEDASTICITY IS GENERALLY NOT RELEVANT ASSUMPTION.

5. Why Assume Homoskedasticity? B) Simpler Std. Errors

- The std. Error formula under heteroskedasticity is a mess (White)
- Under homoskedasticity it's simpler

Bottom line: Linearity and homoskedasticity are restrictive assumptions and we avoid them if possible.

HOMOSKEDASTIC $V(u_i) = \sigma^2$

HETEROSKEDASTIC $V(u_i) = \sigma_i^2$

FIGURE 4.7 **An Example of Heteroskedasticity**

Like Figure 4.4, this shows the conditional distribution of test scores for three different class sizes. Unlike Figure 4.4, these distributions become more spread out (have a larger variance) for larger class sizes. Because the variance of the distribution of u given X, var($u|X$), depends on X, u is heteroskedastic.