**SLR Estimation and Inference: Key Concepts RoadMap**

1) **SLR Estimation**
   a) Those OLS Estimates: min SSRs
   b) Estimators (ex ante) v. estimates (ex post)
   c) The Simple Linear Regression (SLR) Conditions 1-4
      i) **SLR.1 – *Linear DGM*:** $Y = \beta_0 + \beta_1 X + U$ (*X*, *U* and *Y* are random variables)
      ii) **SLR.2 – *Random sampling*:** $Y_i = \beta_0 + \beta_1 X_i + U_i$, $i = 1, ..., n$ generate data: $\{(x_i, y_i)\}$
      iii) **SLR.3 – *Variation in the x's***
      iv) **SLR.4 – *Conditional mean of U is 0*:** $E(U_i \mid X_i = x_i) = 0$
         (1) So: $E(U) = 0$ and $Cov(X, U) = 0$
   d) Under SLR.1 – SLR.4 (skip proofs for now; focus on $B_1$)
      i) OLS estimators, $B_0$ and $B_1$, are linear (conditional on the x's)
      ii) OLS estimators are unbiased, $E(B_0) = \beta_0$ *and* $E(B_1) = \beta_1$
      iii) So $B_0$ and $B_1$ are **LUEs** … but $B_1$ is not alone; there are lots of unbiased linear estimators of $\beta_1$ ; How to choose between them?

         *LUE*s: $B_1 = \sum w_i \left( \frac{Y_i - \overline{Y}}{x_i - \overline{x}} \right)$, where $w_i = \left[ \frac{(x_i - \overline{x})^2}{(n-1)S_{xx}} \right]$ and $\sum w_i = 1$
   e) OLS estimators have a variance… (driven by random nature of the sample; pictures!)
   f) The Simple Linear Regression (SLR) Condition 5
      i) **SLR.5 – *Homoskedasticity*:** $Var(U_i \mid X_i = x_i) = \sigma^2$

2) **Gauss-Markov Theorem**
   a) **Under SLR.1 – SLR.5, OLS estimators are BLUE!** (Who saw this coming?)

**SLR Estimation and Inference: *RoadMap***

**3) SLR Estimation, cont'd**

a) Under SLR.1 – SLR.5

i) $Var(B_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$ (skip proof for now), and so $sd(B_1) = \frac{\sigma}{\sqrt{\sum (x_i - \bar{x})^2}}$

ii) MSE is an unbiased estimator of Var(U | x's): $E(MSE) = E\left(\frac{SSR}{n-2}\right) = \sigma^2$ (skip proof)

(1) $Var(B_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} = E\left(\frac{MSE}{\sum (x_i - \bar{x})^2}\right)$ … use $\hat{\sigma}$ =RMSE to estimate $\sigma$

(2) Use $StdErr(B_1) = se(B_1) = \frac{RMSE}{\sqrt{\sum (x_i - \bar{x})^2}}$ to estimate $sd(B_1)$

**4) SLR Inference**

a) **SLR.6 – *Normality***: $U_i \sim N(0, \sigma^2)$ and is independent of X's

b) *Under SLR.*1 – SLR.6:

i) $\{Y_i \mid X_i = x_i\} \sim Normal(\beta_0 + \beta_1 x_i, \sigma^2)$ since $Y_i = \beta_0 + \beta_1 X_i + U_i$ and $U_i \sim N(0, \sigma^2)$

ii) $B_1 \sim Normal(\beta_1, Var(B_1))$ where $Var(B_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$

iii) $\frac{B_1 - \beta_1}{sd(B_1)} \sim Normal(0,1)$ … could use for inference if only we knew $\sigma$

iv) The ***t statistic***: $\frac{B_1 - \beta_1}{se(B_1)} \sim t_{n-2}$ … used for inference (estimate $\sigma$ with RMSE)

c) Use the t statistic for Inference (Confidence Intervals and Hypothesis Testing)

5) Heteroskedasticity and ***robust*** (White corrected) standard errors