MLR Models: *Estimation & Inference*

- Comparison of SLR and MLR Estimation & Inference
- … *What's New?... Not Much!*
- SLR & MLR Standard Errors: *Not so different after all!*
- *… An Example*

SLR/MLR Compare: *Estimation & Inference I*

	SLR	**MLR**
Those (S/M)LR Assumptions	SLR.1: Linear Model $Y = \beta_0 + \beta_1 X + U$	MLR.1: Linear Model $Y = \beta_0 + \beta_x X_i + \beta_z Z_i + ... + U_i$
	SLR2: Random sampling	MLR2: Random sampling
	SLR3: Sampling variation in the independent variable	MLR3: No perfect collinearity
	SLR.4: Zero Conditional Mean $E(U \mid X = x) = 0$ for any x	MLR.4: Zero Conditional Mean $E(U \mid x, z, ...) = 0$ for any x, z, …
If these hold:	LUE: OLS is a LUE	LUE: OLS is an LUEs
	… But OLS is not alone!	… But OLS is not alone!

SLR/MLR Compare: *Estimation & Inference II*

	SLR	**MLR**
Add one more assumption	SLR.5: Homoskedastic errors (conditional on x)	MLR.5: Homoskedastic errors (conditional on the x's, z's, …)
If all 5 conditions hold:	$Var(B_1 \mid x's) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$ $= \frac{\sigma^2}{(n-1)S_{xx}}$	$Var(B_x \mid x, z \ldots) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} VIF_x$ $= \frac{\sigma^2}{(n-1)S_{xx}} VIF_x$
Variance Inflation Factors (VIF)		$VIF_x = \frac{1}{1-R_x^2}$; $R_x^2 = 1 - \frac{1}{VIF_x}$
	$\hat{\sigma}^2 = \frac{SSR}{n-2} = MSE$	$\hat{\sigma}^2 = \frac{SSR}{n-(k+1)} = MSE$
	MSE unbiased estimator of σ^2	MSE unbiased estimator of σ^2
Gauss-Markov Thm.	SLR.1-SLR.5 $\rightarrow$ BLUE	MLR.1-MLR.5 $\rightarrow$ BLUE

SLR/MLR Compare: *Estimation & Inference III*

	SLR	**MLR**
SRF (Sample Regression Function)	$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$	$\hat{y} = \hat{\beta}_0 + \hat{\beta}_x x + \hat{\beta}_z z + \ldots$
PRF (Population Regression Function)	$E(Y \mid X = x) = \beta_0 + \beta_1 x$	$E(Y \mid x, z, \ldots) = \beta_0 + \beta_x x + \beta_z z + \ldots$
2. Inference		
Add one last Assumption: Distribution of Y_i	SLR.6: $U_i \sim N(0, \sigma^2)$ and indept. of X	MLR.6: $U_i \sim N(0, \sigma^2)$ and indept. of the RHS variables
Degrees of Freedom	n-2	n-k-1
Distribution of Estimator:	$B_1 \sim N\left(\beta_1, \text{var}(B_1)\right)$, where $\text{var}(B_1) = \dfrac{\sigma^2}{(n-1)S_{xx}}$	$B_x \sim N\left(\beta_x, \text{var}(B_x)\right)$, where $\text{var}(B_x) = \dfrac{\sigma^2}{(n-1)S_{xx}} VIF_x$

SLR/MLR Compare: *Estimation & Inference IV*

	SLR	**MLR**
Standard Deviation of B_1	$sd(B_1) = \dfrac{\sigma}{S_x\sqrt{n-1}}$	$sd(B_1) = \dfrac{\sigma}{S_x\sqrt{n-1}}\sqrt{VIF_x}$
Standard Error of B_1 (est. of $sd(B_1)$)	$se(B_1) = \dfrac{RMSE}{S_x\sqrt{n-1}}$	$se(B_x) = \dfrac{RMSE}{S_x\sqrt{n-1}}\sqrt{VIF_x}$
Distribution of Estimator: *t statistic*	$\dfrac{B_1 - \beta_1}{se(B_1)} \sim t_{n-2}$	$\dfrac{B_x - \beta_x}{se(B_x)} \sim t_{n-k-1}$

SLR/MLR Compare: *Estimation & Inference V*

	SLR	**MLR**
Confidence Intervals (C is confidence level)	$[B_1 \pm c\, se(B_1)]$, $P[\lvert t_{n-2} \rvert < c) = C$	$[B_x \pm c\, se(B_x)]$, $P[\lvert t_{n-k-1} \rvert < c) = C$
Null Hypothesis:	$H_0 : \beta_1 = 0$	$H_0 : \beta_x = 0$
t stat (under H_0)	$tstat = \frac{B_1}{se(B_1)} \sim t_{n-2}$	$tstat = \frac{B_x}{se(B_x)} \sim t_{n-k-1}$
p value (for given data set and OLS estimates)	$p = P[\lvert t_{n-2} \rvert > tstat_{\hat{\beta}_1})$	$p = P[\lvert t_{n-k-1} \rvert > tstat_{\hat{\beta}_x})$
Hypothesis Test (for given…) :	Reject H_0 if $\lvert t_{\hat{\beta}_1} \rvert > c$ or if $p < \alpha$	Reject H_0 if $\lvert t_{\hat{\beta}_j} \rvert > c$ or if $p < \alpha$

SLR & MLR Standard Errors: *Not so different after all!*

- At first glance the SLR and MLR standard error formulas appear to differ slightly, with a $\sqrt{VIF_x}$ adjustment included in the MLR formula:
  - SLR - $se_{\hat{\beta}_x} = \dfrac{RMSE}{\sqrt{\sum (x_i - \overline{x})^2}} = \dfrac{RMSE}{S_x \sqrt{n-1}}$
  - MLR - $se_{\hat{\beta}_x} = \dfrac{RMSE}{\sqrt{\sum (x_i - \overline{x})^2}} \sqrt{VIF_x} = \dfrac{RMSE}{S_x \sqrt{n-1}} \sqrt{VIF_x}$
- But in fact, as you'll see below, both expressions can be rewritten as: $se_{\hat{\beta}_x} = \dfrac{RMSE}{\sqrt{SSR_x}}$, where SSR_x is the SSR from the collinearity regression in which the RHS variable *x* has been regressed on the other RHS variables.

MLR Standard Errors: *An Example*

	Full Model	Collinearity Regressions		
	(1)	(2)	(3)	(4)
	rtotgross	wk1	wk2	wk3
wk1	0.540***		0.261***	-0.115***
	(0.0253)		(0.00234)	(0.00336)
wk2	0.745***	2.361***		0.792***
	(0.0761)	(0.0212)		(0.00603)
wk3	4.778***	-1.146***	0.872***	
	(0.0798)	(0.0334)	(0.00664)	
_cons	-0.601**	0.110	0.0817*	0.287***
	(0.227)	(0.102)	(0.0340)	(0.0323)
N	7730	7730	7730	7730
R-sq	0.921	0.886	0.959	0.908
rss (SSR)	2,333,804	471,875	52,173	47,388
rmse	17.38	7.815	2.598	2.476

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

- $wk1: \quad se_1 = \frac{RMSE}{\sqrt{SSR_1}} = \frac{17.38}{\sqrt{471,875}} = .02530$
- $wk2: \quad se_2 = \frac{RMSE}{\sqrt{SSR_2}} = \frac{17.38}{\sqrt{52,173}} = .07609$
- $wk3: \quad se_3 = \frac{RMSE}{\sqrt{SSR_3}} = \frac{17.38}{\sqrt{47,388}} = .07984$