OLS/SLR Analytics: Estimating Parameters

- What is OLS/SLR?
- The OLS/SLR (Simple Linear Regression) model setup
- OLS (Ordinary Least Squares) estimation: FOCs and SOCs
- OLS and sample statistics: Interpreting the OLS coefficients
- … $\hat{\beta}_1$: A Weighted average of slopes
- OLS predictions, residuals and SRFs
- … Properties of OLS/SLR SRFs and residuals
- Units of measurement and estimated coefficients
- Economic significance (meaningfulness): Beta regressions and elasticities
- Examples in Excel and Stata

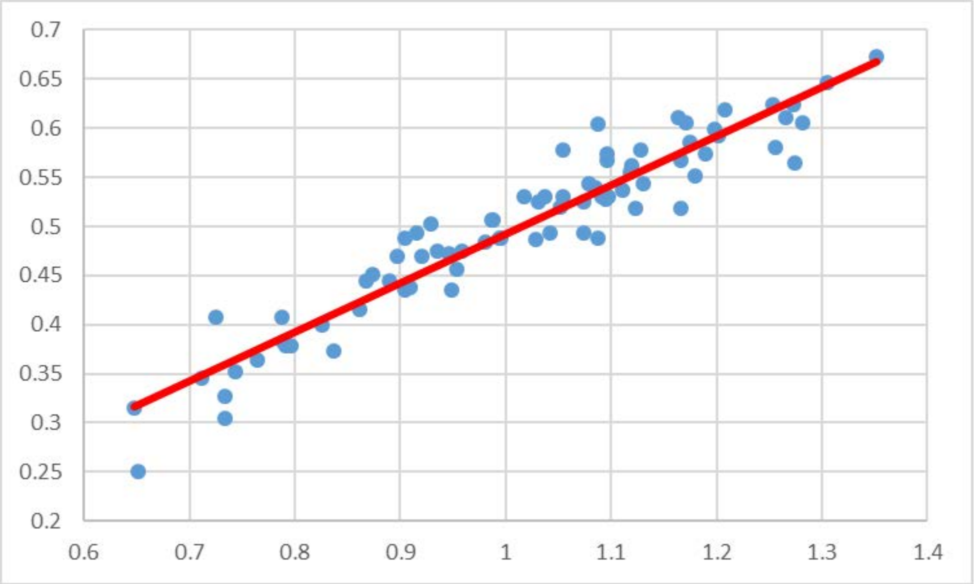
What is OLS/SLR?

- ***SLR (Simple Linear Regression) models:*** You believe that there is essentially a linear relationship between the *dependent* variable *y* (the *LHS* variable), and a single *independent* variable x (the *RHS* or *explanatory* variable… also called a *covariate* or *regressor*):

  $y_i \sim \beta_0 + \beta_1 x_i$ (*true* parameters β_0 and β_1 are unknown and to be estimated)

- ***OLS Estimation:*** The unknown (true) intercept and slope parameters of that linear model will be estimated using the data that you have, by minimizing SSRs (Sum of Squared Residuals)… generating predicted values (*predicteds*) that best fit the actual y's (*actuals*)

- ***OLS/SLR Analysis: What is it?*** … You might hear any/all of these answers …
  - Fitting a straight line to the data.
  - Exploring systematic relationships between the x's and the y's (correlation)
  - Developing a model that best explains the variation of the y's, as a (linear) function of the x's
  - Building a forecasting model that generates predicted y's for given x's
  - Estimating the unknown (true) intercept and slope parameter values
  - Forecasting the conditional mean of the distribution of the y's (conditional on x's)
  - Understanding/estimating the extent to which changes in x cause changes in y (causation)



The OLS/SLR (Simple Linear Regression) Model Setup

- ***Your dataset***: n observations of two variables (x, y): $\{(x_i, y_i)\}$ $i = 1, 2, \dots n$.
- ***Your belief***: Except for random noise in the data, there is a linear relationship between the x's and the y's: $y_i \sim \beta_0 + \beta_1 x_i$ (β_0: the true *y intercept* parameter; β_1: true *slope* parameter)
- ***Your challenge***: Use the data to estimate the unknown true parameters β_0 and β_1
- ***Your parameter estimates***: $\hat{\beta}_0$ and $\hat{\beta}_1$, and your predicted *y* values $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.
- *Your residuals*: For each observation, the residual, $\hat{u}_i$, is the difference between the actual y_i and the predicted value $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$: $\hat{u}_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$.
- ***Your success***: SSRs, the Sum of the Squared Residuals, are one measure of how well the predicteds fit the actuals: $SSR = \sum \hat{u}_i^2 = \sum (y_i - \hat{y}_i)^2$.

OLS (Ordinary Least Squares): find the $\hat{\beta}_0$ and $\hat{\beta}_1$ that min SSRs

OLS = min SSRs

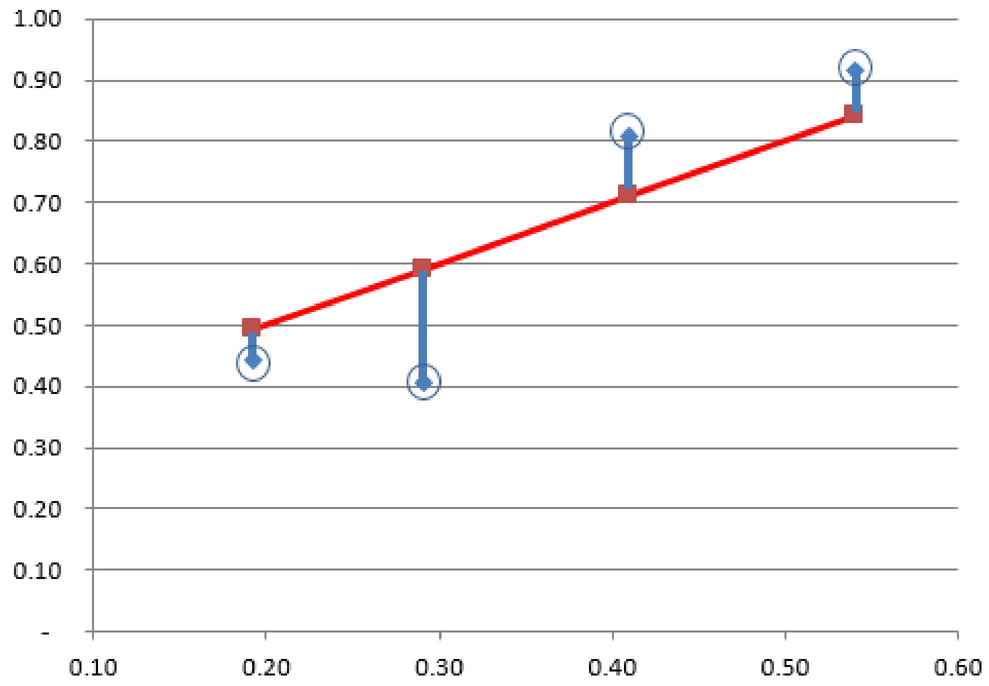
The OLS/SLR Model Setup: An Example

intercept	0.3
slope	1

SSRs
0.0520

id	x	y	pred	residual
1	0.19	0.44	0.49	(0.05)
2	0.29	0.41	0.59	(0.18)
3	0.41	0.81	0.71	0.10
4	0.54	0.92	0.84	0.08

id	resid^2
1	0.0025
2	0.0338
3	0.0100
4	0.0056
SSRs	0.0520



	I	II	III	IV
intercept	0.30	(0.07)	0.29	0.0773
slope	1.00	2.00	1.00	1.5879
SSRs	*0.052*	*0.0386*	*0.0507*	*0.0269*

residuals =
actuals - predicteds

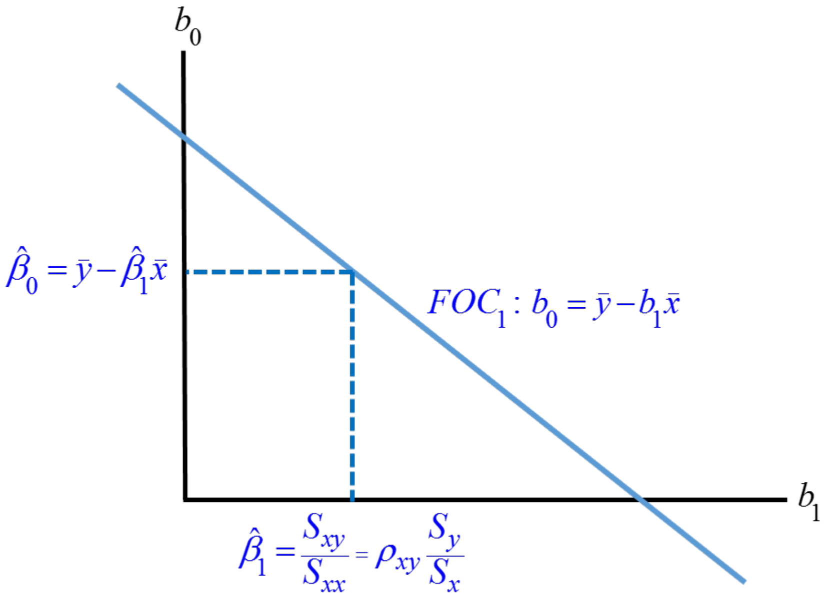
Model IV: Smallest SSRs ↔ *Best Fit*

OLS/SLR estimation (standardized variables): *FOCs and SOCs*

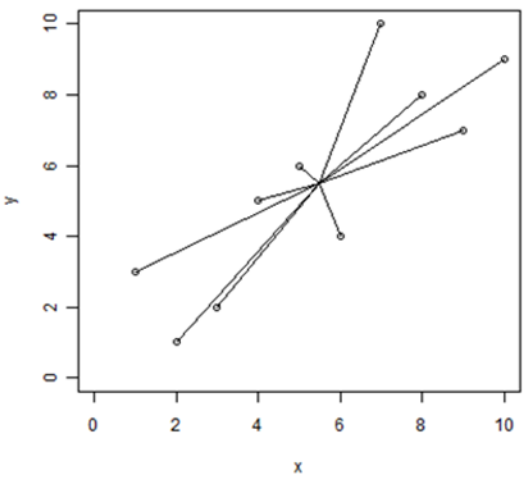
- ***Standardize x's and y's***: $\bar{x} = \bar{y} = 0$, $S_{xx} = S_{yy} = 1$, and $S_{xy} = \rho_{xy}$.
- ***Objective***: min $SSR = \sum\left(y_i - (b_0 + b_1 x_i)\right)^2$ wrt b_0 *and* b_1.
- **FOC 1**: Differentiating *SSR* wrt b_0:
  - $\frac{\partial SSR}{\partial b_0} = -2\sum(y_i - b_0 - b_1 x_i) \Rightarrow -2n\bar{y} + 2nb_0 + 2b_1 n\bar{x} = 0$, so
    $b_0 = \bar{y} - b_1\bar{x}$
  - Checking SOC: $\frac{\partial^2 SSR}{\partial b_0^2} = 2n > 0$, so we have a minimum at b_0 .
- ***OLS intercept estimate: 0!*** Since $\bar{x} = \bar{y} = 0$, $b_0^* = \bar{y} - b_1\bar{x} = 0$ is our best estimate for the intercept parameter.
- ***Now***: min $SSR = \sum\left(y_i - b_1 x_i\right)^2$ wrt b_1.
- **FOC 2**: Differentiating $SSR = \sum\left(y_i - b_1 x_i\right)^2$ wrt b_1:
  - $\frac{dSSR}{db_1} = -2\sum x_i\left(y_i - b_1 x_i\right) = 0$. So $\sum\left(x_i y_i\right) = b_1\sum x_i^2$, and
    $b_1 = \frac{\sum\left(x_i y_i\right)}{\sum x_i^2}$.
  - Checking SOC: $\frac{d^2 SSR}{db_0^2} = 2\sum x_i^2 > 0$, so we do indeed have a minimum at b_1 .
- ***OLS slope estimate: Correlation!*** Since x and y are standardized, we have several equivalent expressions for the estimated slope coefficient: $b_1^* = \frac{\sum\left(x_i y_i\right)}{\sum x_i^2} = \frac{\left[\sum\left(x_i y_i\right)\right]/(n-1)}{\left[\sum x_i^2\right]/(n-1)} = \frac{S_{xy}}{S_{xx}} = \rho_{xy}$.

OLS/SLR estimation … more generally

		Standardized variables	Any x's and y's
FOC 1: $\frac{\partial SSR}{\partial b_0} = 0$	Intercept	$\hat{\beta}_0 = b_0^* = \bar{y} - b_1\bar{x} = 0$	$b_0^* = \bar{y} - b_1\bar{x}$
FOC 2: $\frac{dSSR}{db_1} = 0$	Slope	$\hat{\beta}_1 = b_1^* = \frac{\sum(x_i y_i)}{\sum x_i^2}$	$\hat{\beta}_1 = b_1^* = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$
OLS Estimated Coefficients	Intercept	$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \rho_{xy}\frac{S_y}{S_x} = \rho_{xy}$	$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}} = \rho_{xy}\frac{S_y}{S_x}$
	Slope	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 0$	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$
Predicteds		$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = \rho_{xy} x$	$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = \bar{y} + \hat{\beta}_1(x - \bar{x})$



Interpretation of OLS/SLR estimated slope coefficient



- $\hat{\beta}_1$ is a weighted average of slopes: The estimated slope coefficient is a weighted average of slopes of lines joining the various data points to the sample means $(\overline{x}, \overline{y})$:

$$\hat{\beta}_1 = \sum_i w_i \left[\frac{(y_i - \overline{y})}{(x_i - \overline{x})} \right] = \sum_i w_i \ slope_i \text{, where}$$

  - $w_i = \dfrac{(x_i - \overline{x})^2}{(n-1)S_{xx}} = \dfrac{(x_i - \overline{x})^2}{\sum_j (x_j - \overline{x})^2}$ (the slopes are weighted proportionally to $(x_i - \overline{x})^2$, the square of the various x-distances from the x mean), and
  - $slope_i = \dfrac{(y_i - \overline{y})}{(x_i - \overline{x})}$ is the slope of the line connecting (x_i, y_i) to $(\overline{x}, \overline{y})$

- Note that the data points are not weighted equally (that would be another estimator… but not OLS). Those that are farther away from $\overline{x}$ (in the x dimension) get disproportionately greater weight, which increases with the <u>square</u> of the x-distance from $\overline{x}$.

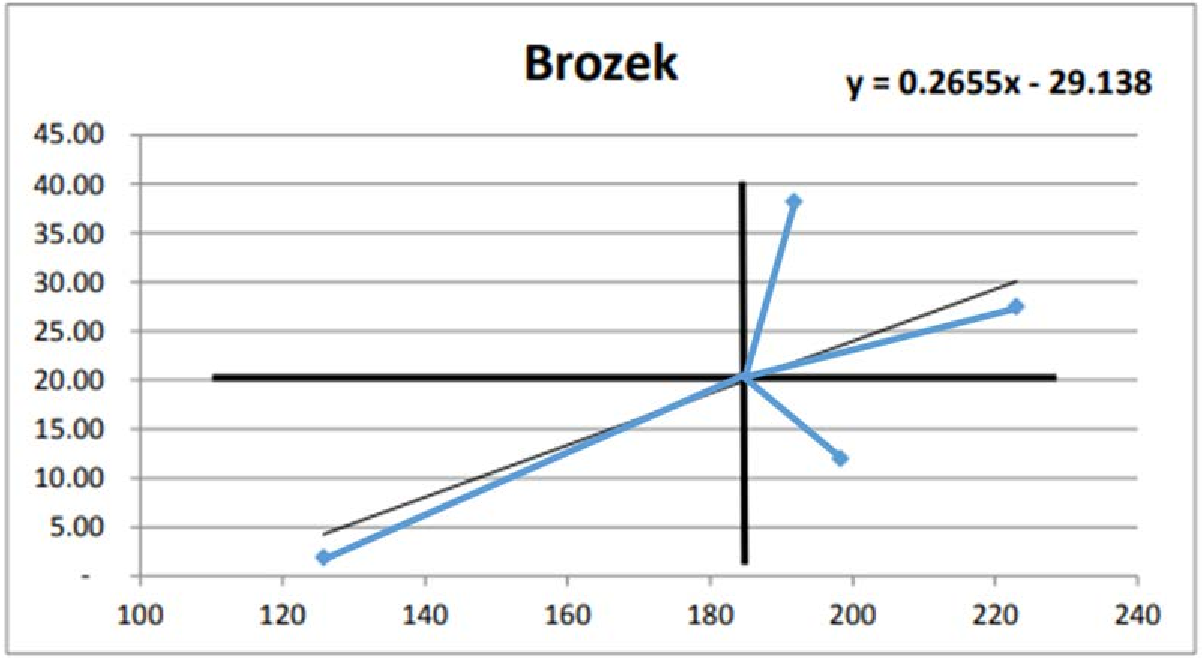
An Example: *A Weighted average of slopes*

- The blue dots are the data points.
- The horizontal and vertical black lines are at the sample means.
- The blue lines are the lines connecting the data points to the sample means.
- The sloped black line shows the predicted Brozek values.
- Note that in the weighted averaging of slopes, one data point, (125.75, 1.90), gets two thirds of the weight, and when combined with (223, 27.50), those two data points get 95% of the weight.
- So even though there are four data points, the slope estimate is being largely driven by just two of the data points.

OLS Slope Estimate: Weighted Average of slopes

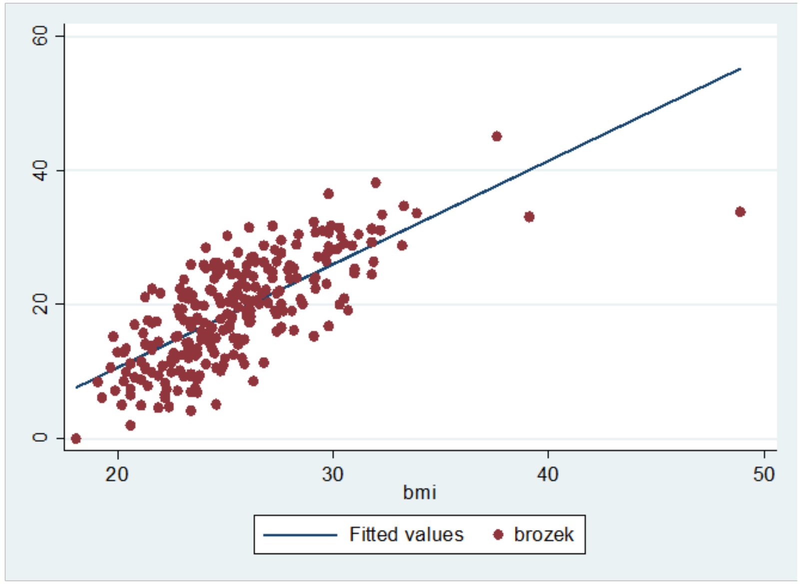
Means	184.688	19.900

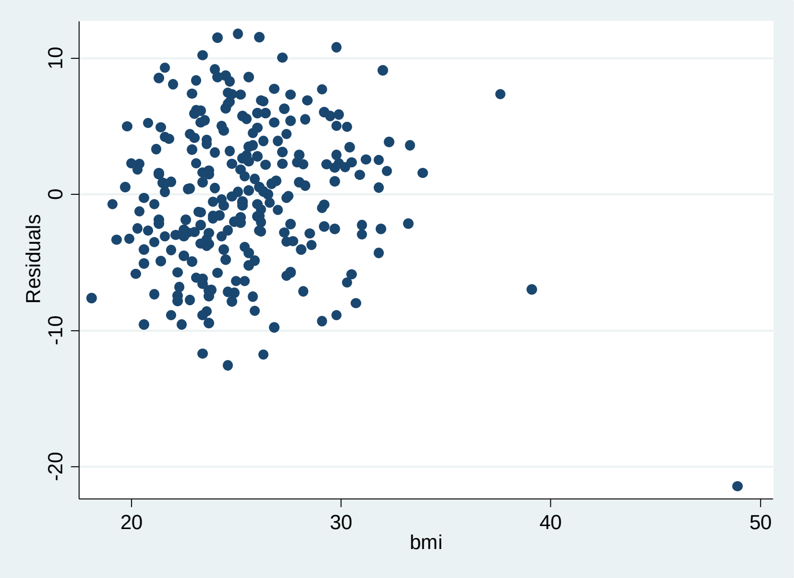
Case	wgt	Brozek	x-dist	y-dist	slope	(x-dist)^2	wgt	wgt*slope
172	125.75	1.90	58.94	18.00	**0.31**	3,473.63	**67%**	0.2050
36	191.75	38.20	(7.06)	(18.30)	**2.59**	49.88	**1%**	0.0250
10	198.25	12.00	(13.56)	7.90	**(0.58)**	183.94	**4%**	(0.0207)
205	223	27.50	(38.31)	(7.60)	**0.20**	1,467.85	**28%**	0.0563
						5,175.30	***100%***	***0.2655***



OLS predictions, residuals and SRFs

- OLS coefficient estimates will generate predicted values, $\hat{y}'s$, and residuals, $\hat{u}'s$:
  - ***Predicted values***: For given x_i, the predicted y_i value given the estimated coefficients is: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ (recall again that we use "hats" for predicted or estimated values).
  - ***Sample Regression Function (SRF)***: The predicted values from the estimated equation, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, comprise the ***Sample Regression Function***.
  - ***Residuals***: And for the given predicted y_i value, the residual, $\hat{u}_i$, is as above the difference between the actual and predicted values: $\hat{u}_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$.





Properties of OLS/SLR SRFs and residuals

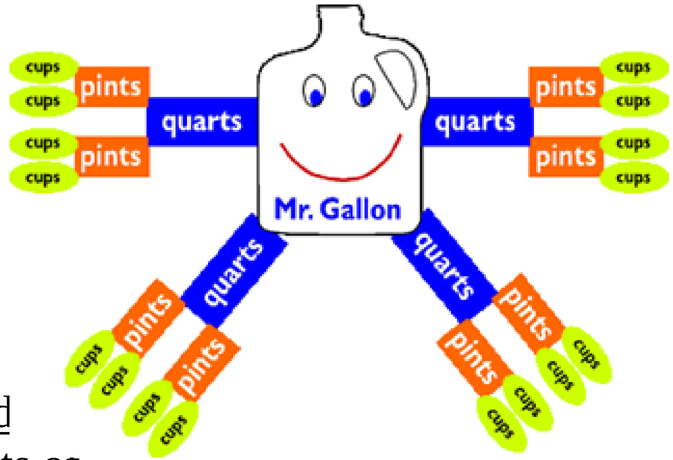
- The SRF passes through the sample means, $(\overline{x},\overline{y})$: at $\overline{x}$, $\hat{y}=\hat{\beta}_0+\hat{\beta}_1\overline{x}=\overline{y}$
- $\rho_{y\hat{y}}=\rho_{yx}$: Correlation (predicteds, actuals) = Correlation (x's, y's)

predicteds and residuals are uncorrelated

- Average residuals: The average residual is zero: $\frac{1}{n}\sum\hat{u}_i=\overline{y}-\left(\hat{\beta}_0+\hat{\beta}_1\overline{x}\right)=0$
- Correlation I - $\rho_{x\hat{u}}=0$: The sample correlation between the $x_i\text{'}s$ and the $\hat{u}_i\text{'}s$ is zero.
- Correlation II - $\rho_{\hat{y}\hat{u}}=0$: The sample correlation between the predicteds ($\hat{y}_i$'s) and the residuals ($\hat{u}_i$'s) is zero
- Decomposition I - $y_i=\hat{y}_i+\hat{u}_i$ and $\rho_{\hat{y}\hat{u}}=0$: OLS essentially decomposes the actual $y_i\text{'}s$ into the sum of two uncorrelated parts, *predicteds* and *residuals*.
- Decomposition II - $S_{yy}=S_{\hat{y}\hat{y}}+S_{\hat{u}\hat{u}}$: Since the predicted and residuals have zero covariance, the variance of their sum is the sum of their variances. Var(actuals) = Var(predicteds) + Var(residuals)

Units of measurement and estimated coefficients

- The magnitudes of the estimated coefficients depend on units of measurement, and typically tell you little about the ***meaningfulness*** of the estimated effect/relationship.
- Changes in the units of measurement of the RHS x variable:
  - Proportionately impact the estimated slope coefficient… and have no impact on the estimated intercept. In a sense, the estimated coefficient unwinds the rescaling of x.
- Changes in the units of measurement of the LHS y variable:
  - Proportionately impact both estimated coefficients (slope and intercept).
- And if you rescale both variables:
  - The impacts on the estimated slope and intercept coefficients will be some combination of the above.
- ***Avoid the magnitude trap!*** Don't fall into the trap of thinking that the <u>sizes/magnitud</u> of estimated coefficients tell you anything useful, as they are driven in part by the units of measurement. You can make coefficients big or small just by changing units of measure.
- ***But the slope coefficient sign tells you something!*** The <u>sign</u> of the slope coefficient does not change with rescaling, and tells you the direction of the estimated effect/relationship.



So pay attention to signs… but not so much to magnitudes… unless you have specific reasons for thinking that the magnitudes are meaningful.

Meaningfulness: Beta regressions and elasticities

- How (economically) ***meaningful*** is the estimated relationship? Do you want to brag about it to the world? Or will everyone just laugh at you, and tell you that what you've estimated is trivial, and of little consequence?

- ***Beta Regressions*** – SLR using standardized variables: $x_i^* = \dfrac{x_i - \overline{x}}{S_x}$ and $y_i^* = \dfrac{y_i - \overline{y}}{S_y}$

  - Regress y^* on x^*. SRF: $\hat{y}^* = \rho_{xy} x^*$ (estimated coefficients: $\hat{\beta}_0 = 0$ and $\hat{\beta}_1 = \rho_{xy}$)

  - The estimated beta regression coefficients are invariant to units of measurement.

- ***Elasticities*** – using the SRF: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

  - ***(Point) Elasticity***: $\varepsilon = \left(\dfrac{d}{dx}\hat{y}\right)\dfrac{x}{\hat{y}} = \hat{\beta}_1 \dfrac{x}{\hat{y}}$ evaluated at $(x, \hat{y}) = (x, \hat{\beta}_0 + \hat{\beta}_1 x)$,

    somewhere along the SRF. Insensitive to units of measurement.

  - We often evaluate the elasticity ***at the means***: $\varepsilon\ (@\,\overline{x}) = \hat{\beta}_1 \dfrac{\overline{x}}{\overline{y}}$ since $\hat{y}\ @\,\overline{x}$ is $\overline{y}$

OLS/SLR Analytics: *TakeAways*

- SLR (Simple Linear Regression) Models: Focus on a linear relationship between a single explanatory (RHS) variable x and the dependent (LHS) variable y
- OLS (Ordinary Least Squares): Estimate unknown intercept and slope parameters by minimizing SSRs (Sum of the Squared Residuals)… (*residuals* are *actuals – predicteds*)
- OLS slope coefficients essentially reflect the correlation between the x's and y's (subject to a standard deviations adjustment)… also a weighted average of slopes of lines connecting data points to the sample means
- The OLS intercept coefficient ensures a zero mean of the residuals, that the SRF passes through the means, and the mean of the *predicteds* is the same as the mean of the *actuals*
- *Actuals* can be decomposed as the sum of *predicteds* and *residuals*, which are uncorrelated… and so the variance of the actuals is the sum of the variance of the *predicteds* and the variance of the *residuals*
- Estimated coefficient values are driven by units of measurement… don't assume that a large (small) coefficient means a large (small) estimated effect
- *Beta regressions* and elasticities provide unit free measures of estimated relationships

onwards… to *SLR Assessment*