OLS/SLR Assessment I: Goodness-of-fit

SST

SSE

SSR

MSE

RMSE

R^2

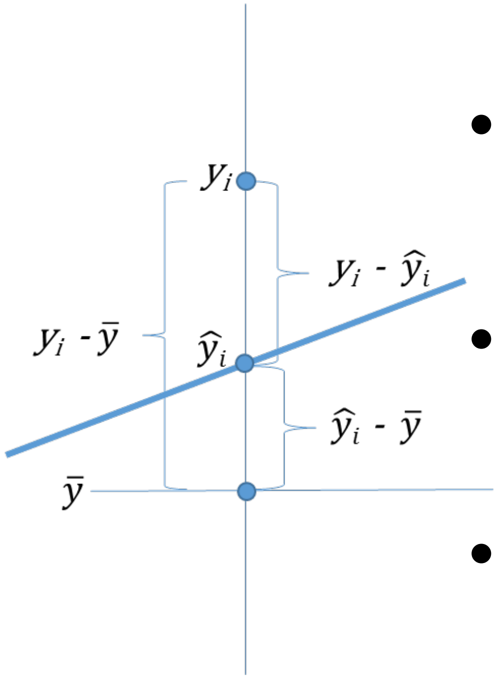
- How close? Goodness-of-Fit (GOF) v. Precision/Inference
- Bring on the ANOVA Table! (SST, SSE and SSR)
- Goodness-of-Fit (GOF) metrics
  - GOF I: Mean Squared Error (MSE)… and Root MSE (RMSE)
  - GOF II: R-squared
- *Thinking about R-squared*
  - Applications
- Comparing SLR models using Goodness-of-Fit (GOF) metrics

How did we do? Goodness-of-Fit (GOF) v. Precision/Inference

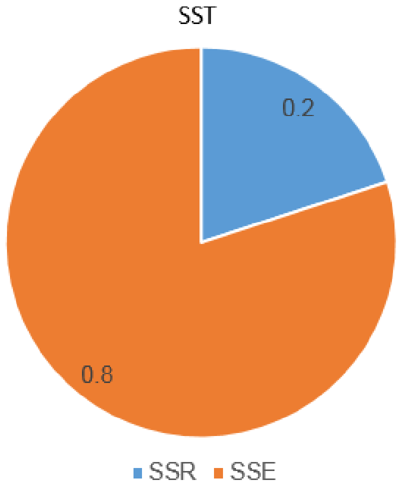
- ***Assessment***: How well did we do? How close are the estimated coefficients to the true parameters, β_0 and β_1? We'll have several answers. None will be entirely satisfactory… though they will be informative, nonetheless.
- ***Goodness-of-Fit*** **:** Goodness-of-Fit metrics tell us something about the quality of the overall model, about how well the *predicteds* fit the *actuals*. They may not tell us much about how precisely we've estimated the true parameters. But if we have a lot of data and the Goodness-of-Fit metrics look good, maybe we should feel pretty good about our estimated coefficients.
- ***Precision/Inference***: While goodness-of-fit metrics tell us something about how well our estimated model fits the data, they don't directly tell us anything about how precisely we have estimated the unknown parameters, the true $\beta's$. Later on, we will have lots to say about precision of estimation… but that discussion awaits the development of the tools of statistical inference, including *Confidence Intervals* and *Hypothesis Tests.*
- ***Who knew? They are related!*** At first glance that *Goodness-of-fit* and *Precision/Inference* look to be completely unrelated, as one looks at how well a SLR model fits the data whilst the other considers the precision of estimation of individual parameters. But quite the contrary!

Stay tuned!

Bring on the ANOVA Table! (*SSTs, SSEs and SSRs*)



- ***SST*** *(Total Sum of Squares)*
  - $SST = \sum (y_i - \bar{y})^2 = (n-1)S_{yy}$, (n-1) times the variance of the *actuals*
- ***SSE*** *(Explained Sum of Squares)*
  - $SSE = \sum (\hat{y}_i - \bar{y})^2 = (n-1)S_{\hat{y}\hat{y}}$, (n-1) times the variance of the *predicteds*
- ***SSR*** *(Residual Sum of Squares)*
  - $SSR = \sum (y_i - \hat{y}_i)^2 = \sum \hat{u}_i^{\,2} = (n-1)S_{\hat{u}\hat{u}}$, $(n-1)$ times the variance of the *residuals*
- ***SST = SSE + SSR*** (if there is a constant term in the model)
  - Put differently: $\dfrac{SST}{n-1} = \dfrac{SSE}{n-1} + \dfrac{SSR}{n-1}$, or $S_{yy} = S_{\hat{y}\hat{y}} + S_{\hat{u}\hat{u}}$



The sample variance of the actuals is the sum of the sample variances of the predicteds and of the residuals.

Goodness-of-Fit (GOF) metrics

GOF I: Mean Squared Error (MSE/RMSE)

- Mean Squared Error: $MSE = \frac{SSR}{n-2}$ (in squared units of the *y* variable)

- Root Mean Squared Error: $RMSE = \sqrt{MSE} = \sqrt{\frac{SSR}{n-2}}$ (in units of the *y* variable)

GOF II: R-squared

- The ***Coefficient of Determination***, is defined by: $R^2 = 1 - \frac{SSR}{SST}$

- So long as there is a constant term in the model (so the mean predicted value is the same as the mean actual value), $SSR = SST - SSE$ and:

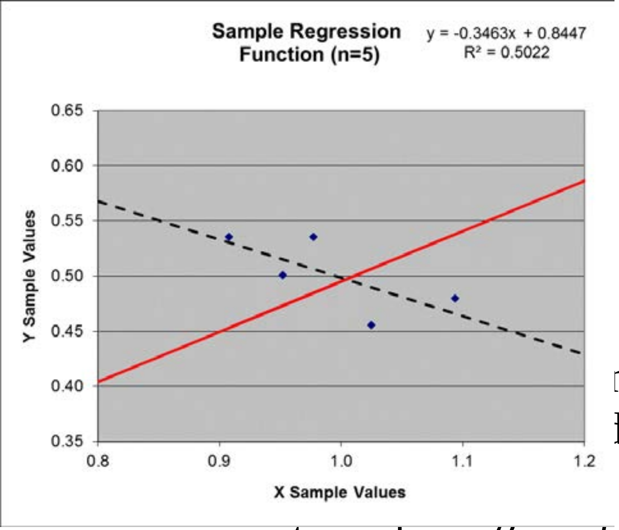
$$R^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST} = \frac{SSE/(n-1)}{SST/(n-1)} = \frac{Sample\,Var(predicted)}{Sample\,Var(actual)} = \frac{S_{\hat{y}\hat{y}}}{S_{yy}}.$$

Thinking about R-squared

- ***R*² *is bounded***: By construction, $0 \le R^2 \le 1$ (if there is a constant term in the model)… higher values mean that you've done a better job explaining the variation in the actuals. Don't get too excited if R^2 is close to 1, or too depressed if it's close to 0. Doing good econometrics is way more than just maximizing R^2.

- ***R*² *as the Ratio of Variances***. Given the results above, R-squared is the ratio of the *Sample Variance* of the *predicteds* to the *Sample Variance* of the *actuals*… the percent of the variation of the actuals *explained* by the model. This is the most common, and perhaps the most insightful, interpretation of R^2.

- ***R*² *as the correlation*² *between predicted and actuals***. R^2 is also the square of the sample correlation between the independent and dependent variables, as well as the sample correlation between the *actuals* and *predicteds*: $\rho_{xy}^2 = \rho_{\hat{y}y}^2 = R^2$.

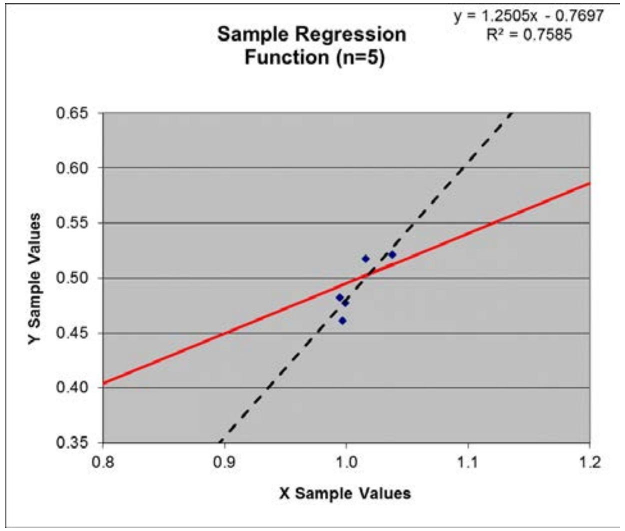


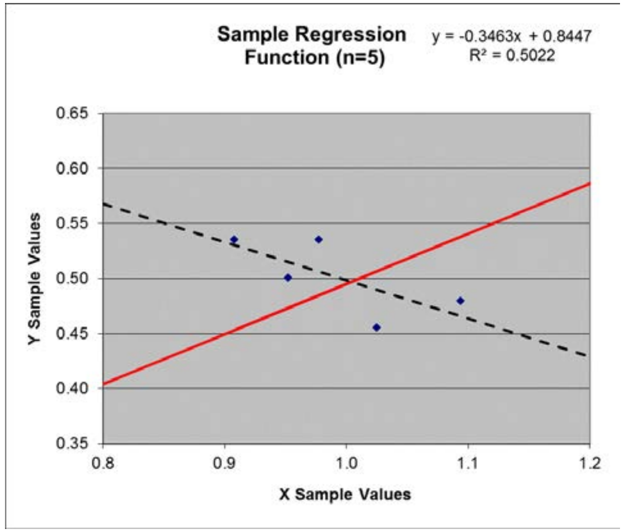
OLS/SLR estimation … more generally

metrics (R-squared and MSE/RMSE) tell you something
lel captures/explains the variation in the dependent variable,
do not tell you how well you've estimated the unknown
parameter values β_0 *and* β_1. In some cases, R-squared will be high and MSE/RMSE will be low, and your parameter estimates will be quite poor… and vice-versa.

- Some examples:
  - Suppose you have a sample of size two. With just two data points, $R^2 = 1$ and $MSE = 0$ … and in all likelihood you have miserable estimates of the unknown parameter values.
  - Here are two examples with just five observations randomly generated using a true relationship given by the solid red line…. and the dashed black line shows you the OLS estimated SLR relationship for the given dataset. In both cases, the R^2 is above.5, and the estimated relationship is all wrong. So n matters too!

nObs Matters Too!





Comparing SLR models using Goodness-of-Fit (GOF) metrics

- You can use R^2 and MSE/RMSE to compare the performance of different SLR models… but only to a limited extent. *And you must be careful!*
- If the different models all have the ***same LHS data*** (so the y's are the same in the different models… both in terms of number and in terms of values), then the SSTs and S_{yy}'s will be the same across the models, and you can compare R^2's and MSE/RMSE's. Under these conditions the R^2's and the MSE/RMSE's will move in opposite directions, since

$$R_1^2 > R_2^2 \Leftrightarrow 1 - \frac{SSR_1}{SST} > 1 - \frac{SSR_2}{SST} \Leftrightarrow SSR_1 < SSR_2 \Leftrightarrow \frac{SSR_1}{n-2} < \frac{SSR_2}{n-2} \Leftrightarrow MSE_1 < MSE_2 \ .$$

- So under these conditions, models with higher R^2's (and lower MSE/RMSE's) do a better job of fitting the data, and in that sense are preferable.
- But: If the y's are not the same across the different models, then R^2's and MSE/RMSE's are not directly comparable and accordingly, they won’t tell you much unless you make some adjustments.

OLS/SLR Assessment I - GOFs: *TakeAways*

- Goodness of Fit metrics tell you something about how well your OLS/SLR model fits the data… about the relationship between *predicteds* and *actuals*.
- SSTs, SSEs and SSRs capture the variances in the *actuals*, *predicteds* and *residuals*, respectively. SST = SSE + SSR
- Two standard GOF metrics are *(Root) Mean Squared Error* (MSE = SSR/(n-2)) and the *Coefficient of Determination* (R-sq = 1-SSR/SST = SSE/SST = varPredicteds/varActuals))
- MSE is essentially an average squared deviation of *predicteds* from *actuals*… RMSE is the square root thereof. MSE (RMSE) magnitudes tell you little about how well your model has performed, as they have no uniform scale.
- R-sq is essentially the variance of the *predicteds* relative to the variance of the *actuals*… or, the percent of the variation in the *actuals* explained by the model. $0 \leq$ R-sq ≤ 1; closer to 1 and the more of the variation in the dependent variable explained by the model. High R-sq is terrific if nObs are high as well… but maybe not so much otherwise.
- R-sq is also equal to the square of correlation between the LHS and RHS variables… as well as the square of the correlation between *predicteds* and *actuals*.
- It's OK to compares MSE's and R-sq's across models with the same LHS variable… but if there are changes to the LHS variable, such comparisons are meaningless without adjustments

onwards… to OLS/*SLR Examples*