OLS/SLR Assessment I: *Goodness-of-fit*

- ***How close? Goodness-of-Fit 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***
- ***… Applications***
- ***Examples in Excel and Stata***
- ***Comparing SLR Models Using Goodness-of-Fit Metrics***

How close? Goodness-of-Fit v. Precision/Inference

1. After we have derived the OLS parameter estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$, the question always arises: 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.
2. **Quality of the Overall Model (*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 as much as we'd like to know 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, then we should feel pretty good about our estimated coefficients, even though there is always some probability that they are way off.
   a. ***GOF Metric I**: MSE (Mean Squared Error)*
      i. *MSE* is *almost sort of like* an average squared residual… I say *almost sort of like* because instead of taking an average (and dividing by n), we divide the sum of the squared residuals by n-2. (As you'll see later, that choice reflects an interest in unbiasedness.)
   b. ***GOF Metric II***: R^2 *(Coefficient of Determination)*

      i. There are two equivalent ways to think about R^2.
         1. One interpretation is that it measures the proportion of the variation in the y's (the *actuals*) *explained* by the $\hat{y}'s$ (the *predicteds*).
         2. Alternatively, it captures the magnitude of the correlation between the y's and the $\hat{y}'s$, the *actuals* and the *predicteds*.
      ii. It will turn out that $0 \le R^2 \le 1$, and so if we have R^2 close to 1 we say that goodness-of-fit is high, and if it's close to 0, goodness-of-fit is low. In contrast, it won't always be so obvious whether the MSE's are large or small in magnitude.

3. **Quality of the Individual Parameter Estimates (*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.*

   While those inferential tools won't with certainty answer the question *How Close?*, they will give us probabilistic assessments as to how close our estimated coefficients are to the true unknown parameter values: *levels of confidence* for confidence intervals and *significance levels* for hypothesis testing.

4. ***Who knew? They are related!*** It may appear at first glance that *Goodness-of-fit* and *Precision/Inference* are 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! Once we get to statistical inference, you will see that in a very formal/concrete way, the precision of parameter estimation is driven entirely by just two factors:

   - the R^2 *Goodness-of-fit* metric, and
   - the number of observations in the dataset.

   So these two apparently independent assessment metrics are not independent… not at all. ***Stay tuned!***

Before turning to R^2 we first need to introduce some ANOVA (*Analysis of Variance*) terminology and results.

Bring on the ANOVA (SST, SSE and SSR)

5. Some definitions which will be useful in deriving the *MSE/RMSE* and R^2 *goodness-of-fit* metrics:[1]

   a. **SST**: Total Sum of Squares … $\sum (y_i - \bar{y})^2$

      i. This the sum squared deviations of the actual values of the dependent variable from its mean.

      ii. Since $S_{yy} = \frac{1}{n-1}\sum (y_i - \bar{y})^2 = \frac{SST}{n-1}$, $SST = (n-1)S_{yy}$ … $(n-1)$ times the variance of the *actuals*.

[1] Everyone doesn't always use the same terminology for these concepts. In Stata regression output, SST is SS Total, SSR is SS Residual, and SSE is SS Model. And some authors flip the definitions of SSE and SSR.

b. **SSE**: Explained Sum of Squares … $= \sum(\hat{y}_i - \bar{y})^2$

i. This is the sum squared deviations of the predicted values of the dependent variable from the mean of the actual values.

ii. If there's a constant term in the model, the mean of the *actuals* is also the mean of the *predicteds*, $\bar{y} = \bar{\hat{y}}$. In this most common case: $SSE = \sum(\hat{y}_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{\hat{y}})^2$ … and so $SSE = \sum(\hat{y}_i - \bar{y})^2 = (n-1)S_{\hat{y}\hat{y}}$ … $(n-1)$ times the variance of the *predicteds*.

c. **SSR**: Residual Sum of Squares… $\sum \hat{u}_i^{\,2} = \sum(y_i - \hat{y}_i)^2$

i. This is the sum squared *residuals*, the sum of the squared differences between the *actual* and *predicted* values of the dependent variable.

ii. Since $\sum \hat{u}_i = 0$, the *residuals* by construction have mean 0, and so $S_{\hat{u}\hat{u}} = \dfrac{SSR}{n-1}$ … or put differently, $SSR = (n-1)S_{\hat{u}\hat{u}}$ … $(n-1)$ times the variance of the *residuals*.

6. To summarize:

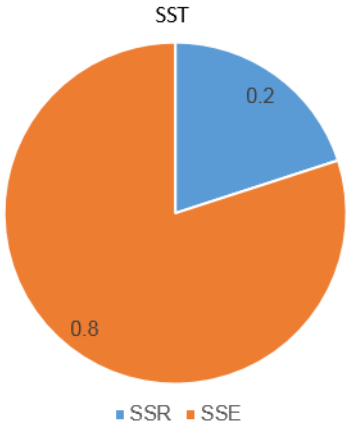
- $SST = \sum(y_i - \bar{y})^2 = (n-1)S_{yy}$, (n-1) times the variance of the *actuals*
- $SSE = \sum(\hat{y}_i - \bar{y})^2 = (n-1)S_{\hat{y}\hat{y}}$, (n-1) times the variance of the *predicteds*
- $SSR = \sum \hat{u}_i^{\,2} = \sum(y_i - \hat{y}_i)^2 = (n-1)S_{\hat{u}\hat{u}}$, (n-1) times the variance of the *residuals*

7. **Result:** [2] SST = SSE + SSR, if there is a constant term in the model.

a. … or dividing through by (n-1), we have $\dfrac{SST}{n-1} = \dfrac{SSE}{n-1} + \dfrac{SSR}{n-1}$, or $S_{yy} = S_{\hat{y}\hat{y}} + S_{\hat{u}\hat{u}}$

b. In words: The sample variance of the *actuals* is the sum of the sample variances of the *predicteds* and of the *residuals*.

c. What drives this result? Since we have a constant term in the regression, the mean of the predicted values is the



[2]Proof: The trick is to add and subtract $\hat{y}_i$ inside the expression and to then simplify:

$\sum(y_i - \bar{y})^2 = \sum(y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum(y_i - \hat{y}_i)^2 + \sum(\hat{y}_i - \bar{y})^2 + 2\sum(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$
$= SSR + SSE + 2\sum(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$. So we just need to prove that $\sum(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$
$= \sum \hat{u}_i(\hat{y}_i - \bar{y}) = 0$, the sample covariance of the predicted values and the residuals is zero.

same as the means of the actuals… or put differently: $\bar{\hat{y}} = \bar{y}$. And that fact drives the result.

8. You shouldn't be too surprised by this result. Earlier we showed that OLS effectively decomposed the y's into two uncorrelated parts, *predicteds* and *residuals*. And since $y_i = \hat{y}_i + \hat{u}_i$ and $\hat{\rho}_{\hat{y}\hat{u}} = 0$, the sample variance of the *actuals* will be the sum of the sample variances of the *predicteds* and of the *residuals*… which is exactly the result above, $S_{yy} = S_{\hat{y}\hat{y}} + S_{\hat{u}\hat{u}}$. So perhaps you saw this coming.

9. This result does not necessarily hold if there is no constant (intercept) term in the model. But do not fear! There are lots of good reasons for including a constant term in your model. In fact, general practice is to always include a constant term in your model unless you have a specific reason not to do so.

Goodness-of-Fit I: Mean Squared Error (MSE/RMSE)

10. MSE provides one measure of how close your predicted values are to the *actuals*:

    a. $MSE = \dfrac{SSR}{n-2}$.[3] measured in squared units of the dependent variable.

11. To put the metric in the same units as the y's, we take the square root of the MSE… this gives us *Root Mean Squared Error (RMSE)*, which is sometimes called the *standard error of the regression*. This metric is sort of like an average deviation of predicted from actuals… but not quite, give the specifics of the calculation and for reasons previously discussed.

    a. $RMSE = \sqrt{MSE} = \sqrt{\dfrac{SSR}{n-2}}$ … measured in units of the dependent variable.

12. Sometimes we also look at, *Mean Absolute Error (MAE)*, a goodness-of-fit metric closely related to RMSE:

    a. $MAE = \dfrac{1}{n}\sum\left|y_i - \hat{y}_i\right|$, where $\left|y_i - \hat{y}_i\right|$ is the absolute value of the i[th] residual.

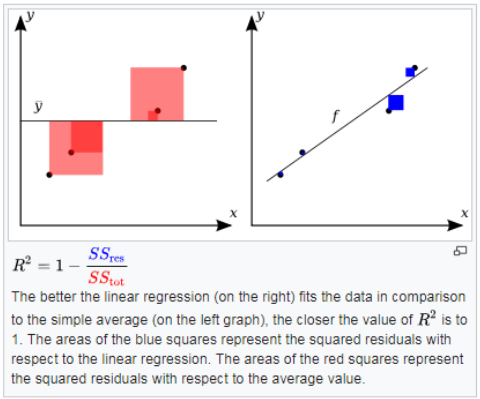
    b. MAE's are not typically included in standard regression package results… but they can usually be obtained easily.

13. One of the challenges in working with MSEs, RMSEs and MAEs is interpreting magnitudes. On their face, it's not obvious whether these metrics are small or large in magnitude. So you'll need to bring other information to bear in forming an opinion as to how well your model has fit the data. Our alternative metric, the *Coefficient of Determination* $\left(R^2\right)$, provides more readily interpreted results.

But since $\hat{y}_i - \bar{y} = \hat{\beta}_0 - \hat{\beta}_1 x_i - \left(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}\right) = \hat{\beta}_1\left(x_i - \bar{x}\right)$, $\sum \hat{u}_i(\hat{y}_i - \bar{y}) = \hat{\beta}_1 \sum \hat{u}_i\left(x_i - \bar{x}\right)$ … and this is 0, since $\sum \hat{u}_i\left(x_i - \bar{x}\right) = 0$.

[3] As you'll see later in the semester, we divide by (n-2) rather than n to achieve unbiasedness.

Goodness-of-Fit II: R-squared



14. Our second goodness-of-fit metric, the *Coefficient of Determination*, is defined by: $R^2 = 1 - \frac{SSR}{SST}$

    a. 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 so

    $$R^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}.$$

    i. If your model does not have a constant term then this last formula need not be the case. Further: If your model does not have a constant/intercept term then you should not pay too much, if any, attention to R^2.

    b. Then $R^2 = \frac{SSE}{SST} = \frac{\left[\sum(\hat{y}_i - \bar{y})^2\right]/(n-1)}{\left[\sum(y_i - \bar{y})^2\right]/(n-1)} = \frac{Sample\,Var(predicted)}{SampleVar(actual)} = \frac{S_{\hat{y}\hat{y}}}{S_{yy}}.$

    c. 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.

15. ***Interpretation I***: ***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.

16. ***Interpretation II***: ***Correlation² 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*: $\rho_{xy}^2 = \rho_{\hat{y}y}^2 = R^2$.

    a. This is an important result… so here's a quick proof:

    i. We know that $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \rho_{xy}\frac{S_y}{S_x}$.

    ii. Since $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, the sample variance of the predicted values will be defined by:

    $$S_{\hat{y}\hat{y}} = \frac{\sum(\hat{\beta}_0 + \hat{\beta}_1 x_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{x})^2}{n-1} = \hat{\beta}_1^2 S_{xx}.$$

    But then $R^2 = \frac{S_{\hat{y}\hat{y}}}{S_{yy}} = \frac{\hat{\beta}_1^2 S_{xx}}{S_{yy}} = \left[\rho_{xy}\frac{S_y}{S_x}\right]^2 \frac{S_{xx}}{S_{yy}} = \left[\rho_{xy}^2 \frac{S_{yy}}{S_{xx}}\right]\frac{S_{xx}}{S_{yy}} = \rho_{xy}^2.$

iii. … or put differently: Since $SSE = \rho_{xy}^2 \; SST$ (see following), $\rho_{xy}^2 = \dfrac{SSE}{SST} = R^2$.

$SSE = \sum \left(\hat{y}_i - \overline{y} \right)^2 = \sum \left(\hat{\beta}_0 + \hat{\beta}_1 x_i - (\hat{\beta}_0 + \hat{\beta}_1 \overline{x}) \right)^2 = \hat{\beta}_1^2 \sum (x_i - \overline{x})^2$. And so

$$SSE = \left[\rho_{xy} \frac{S_y}{S_x} \right]^2 \sum (x_i - \overline{x})^2 = \rho_{xy}^2 \frac{SST}{\sum (x_i - \overline{x})^2} \sum (x_i - \overline{x})^2 = \rho_{xy}^2 \; SST \; .$$

iv. And since $\rho_{xy} = \rho_{\hat{y}y}$ (the correlation of the x's and y's is the same as the correlation between the *predicteds* and the *actuals*), we have the desired result: $R^2 = \rho_{xy}^2 = \rho_{\hat{y}y}^2$.

b. When we move to MLR models, with multiple explanatory variables, we lose the connection between R^2 and ρ_{xy}^2 … but the connection to the correlation between *predicteds* and *actuals* will carry forward ($\rho_{\hat{y}y}^2 = R^2$ for MLR models as well).

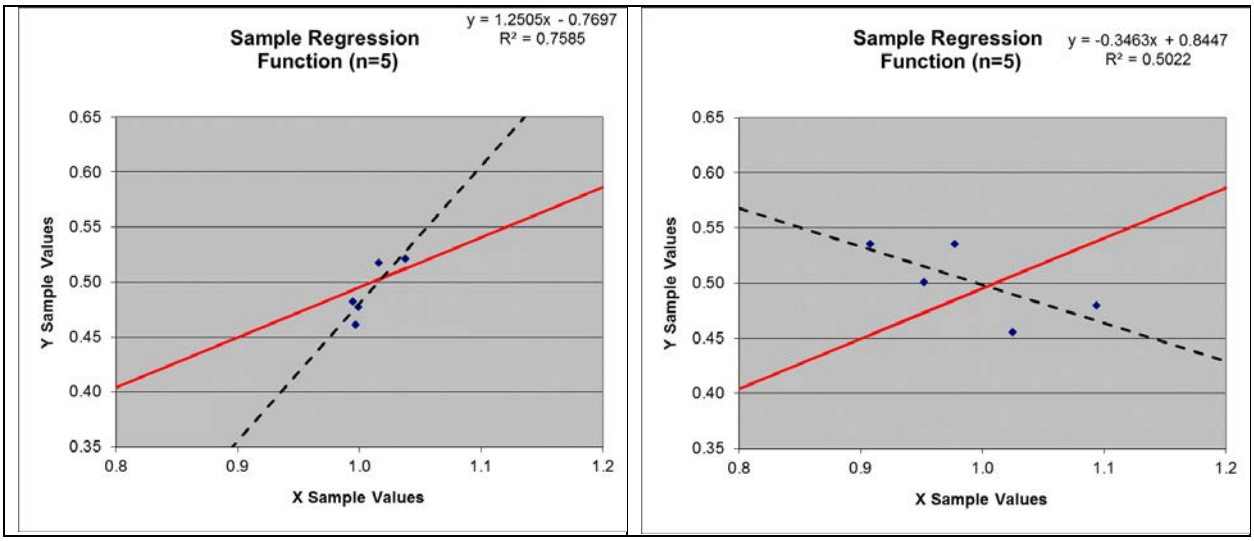
Goodness-of-Fit Applications

17. The two goodness-of-fit metrics (R-squared and MSE/RMSE) tell you something about how well your model captures/explains the variation in the dependent variable, y. They alone, however, 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.

a. ***Example***: Suppose you have a sample of size two. With just two data points, $R^2 = 1$ and $MSE = 0$ … and you have in all likelihood come up with miserable estimates of the unknown parameter values.

OLS/SLR Assessment I: ***Goodness of Fit***

18. Here are a couple 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!



19. ***nObs Matters Too!*** We will see later that the quality of the parameter estimates depends on R-squared (or MSE/RMSE) *and* the number of observations in the dataset. And so if R-squared is high and MSE/RMSE is low, *and* you have lots of data, then you have probably done a pretty good job estimating the unknown parameter values. But without much data… *Who knows?*

nObs Matters Too!

OLS/SLR Assessment I: ***Goodness of Fit***

Examples in Excel and Stata

Excel: Continuing with the bodyfat example in Excel.

Generate the predicteds, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, and residuals, $\hat{u}_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$. Generate SSRs by squaring the residuals and summing those (use the SUMSQ() function to save a step). Use the COUNT() function to count your observations, and generate $MSE = \frac{SSR}{n-2}$ and $RMSE = \sqrt{MSE}$.

To generate SSEs, demean the predicteds, and compute the sum squared of those, again using SUMSQ(). And use SUMSQ() to compute SST using the demeaned Brozek observations. Once you have all of these, you can verify that $SSR + SSE = SST$. And with SSE and SST, divide by n-1 to generate the sample variances of the explained and the actuals.

You can now compute R^2 four ways:

1. $R^2 = 1 - \frac{SSR}{SST}$,
2. $R^2 = \frac{SSE}{SST}$,
3. $R^2 = \frac{Sample\,Var(predicted)}{SampleVar(actual)} = \frac{S_{\hat{y}\hat{y}}}{S_{yy}}$, and
4. $R^2 = \rho^2_{xy} = \rho^2_{\hat{y}y}$.

Here's what your results might look like:

Sample Variances		Sample Cov	Sample Corr	Slope estimates			
863.72	60.08	139.67	0.6132	Sxy/Sxx	0.1617	R-squared	
29.39	7.75			corr*(Sy/Sy)	0.1617	1-SSR/SST	0.37596
						SSE/SST	0.37596
Sum Squares		Sum		Intercept estimate		VarPred/VarActual	0.37596
216,794.40	15,079.02	35,057.55		Bbar-b1*wbar	(9.9952)	corr^2	0.37596
	Variance		Variance				
	60.0758		22.5861			MSE	37.640
	SSTs	count	SSEs		SSRs	RMSE	6.1351
	15,079.02	252	5,669.11		9,409.90		
wgt-wbar	Brozek-Bbar	product	Pred-muPred	Predicteds	Residuals	SSE+SSR-SST=0?	-
(24.67)	(6.34)	156.40	(3.99)	14.95	(2.35)		
(5.67)	(12.04)	68.31	(0.92)	18.02	(11.12)		
(24.92)	5.66	(141.11)	(4.03)	14.91	9.69		
5.83	(8.04)	(46.83)	0.94	19.88	(8.98)		
5.33	8.86	47.19	0.86	19.80	8.00		

I have posted **bodyfat example 2.xlsx** to illustrate.

OLS/SLR Assessment I: *Goodness of Fit*

Running the Regression in Excel

When you run the regression in Excel, you'll get the following results:

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.61316					
R Square	0.37596					
Adjusted R Square	0.37346					
Standard Error	6.13511					
Observations	252					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	5,669.11	5,669.11	150.62	2.05905E-27	
Residual	250	9,409.90	37.64			
Total	251	15,079.02				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	(9.9952)	2.3891	(4.18)	3.97276E-05	(14.7004)	(5.2899)
wgt	0.1617	0.0132	12.27	2.05905E-27	0.1358	0.1877

You can find SSE, SSR, SST, MSE, RMSE and R-squared in there … you just need to know where to look. The SS's are all in the SS column, with *Regression* for SSE, *Residual* for SSR, and *Total* for SST. MSE can be found in the MS column, row *Residual*. R squared is reported under *Regression Statistics*, and what Excel calls the *Standard Error* of the regression, we call RMSE. So the statistics are all there… you just need to know where to look.

Running the Regression in Stata

```
. reg brozek wgt

      Source |       SS           df       MS      Number of obs   =       252
-------------+----------------------------------   F(1, 250)       =    150.62
       Model |  5669.11331         1  5669.11331   Prob > F        =    0.0000
    Residual |  9409.90318       250  37.6396127   R-squared       =    0.3760
-------------+----------------------------------   Adj R-squared   =    0.3735
       Total |  15079.0165       251  60.0757629   Root MSE        =    6.1351

------------------------------------------------------------------------------
      brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |   .1617088   .0131765    12.27   0.000     .1357578    .1876598
       _cons |  -9.995151   2.389056    -4.18   0.000    -14.70039   -5.289908
------------------------------------------------------------------------------

. predict bfathat
(option xb assumed
```

Again, you can find SSE, SSR, SST, MSE, RMSE and R-squared in there … you just need to know where to look. The SS's are again in column SS, but Stata now puts the SSEs in the *Model* row. MSE's are again in column MS and row *Residual*. And R-squared and Root MSE (RMSE) are in the regression stats in the upper right corner.

We again find that SSR + SSE = SST:

```
. di 5669.11331 + 9409.90318 -15079.0165
-.00001
```

… and R-squared is indeed those correlations squared:

```
. corr Brozek bfathat wgt
(obs=252)

             |   Brozek  bfathat      wgt
-------------+---------------------------
      Brozek |   1.0000
     bfathat |   0.6132   1.0000
         wgt |   0.6132   1.0000   1.0000

. di .6132^2
.37601424
```

Comparing SLR Models Using Goodness-of-fit Metrics

For the applied econometrician, the journey is as important as the final destination. And there's plenty of science and art along the way. Each regression analysis tells you something… and leads to the next analysis. Ultimately, you typically converge on your preferred model… but there was plenty of learning along the way. And that learning definitely informed your analysis.

As part of the learning process, econometricians are always comparing results across models, and making decisions about how to move forward. We'll have a lot more to say about that later, but given that we are in the midst of *Goodness-of-Fit* metrics, why not say a few words about how to use those metrics to compare models?

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.

Here are some examples using the bodyfat dataset.

Example 1: Predicting Brozek with four different SLR Models

Here are the results from four SLR models, where *Brozek* is the common LHS variable and *hgt*, *wgt*, *abd*, and *BMI* are the candidate RHS variables

	(1) Brozek	(2) Brozek	(3) Brozek	(4) Brozek
hgt	-0.189			
	(-1.41)			
wgt		0.162***		
		(12.27)		
abd			0.585***	
			(22.13)	
BMI				1.547***
				(16.79)
_cons	32.17***	-9.995***	-35.20***	-20.41***
	(3.44)	(-4.18)	(-14.29)	(-8.62)
N	252	252	252	252
R-sq	0.008	0.376	0.662	0.530
rss	14,959.3	9,409.9	5,094.9	7,087.5
rmse	7.735	6.135	4.514	5.324

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

The syntax for the *esttab* output was:

. esttab, r2 scalar(rss rmse) compress

The options in the esttab statement:

- *r2*: displays R^2
- *rss*: displays SSRs
- *rmse*: displays RMSE

and *compress* compresses the output so it is not as wide and fits better on the page.

Notice that R^2 increases as you go from *hgt* (0.008), to *wgt* (0.376), to *abd* (0.662) and then decreases with *BMI* (0.530). And as advertised, RMSE moves in exactly the opposite direction.

Looking across the four models, *abd* (waist size) has most explanatory power (highest R^2's and lowest MSE/RMSE's), *BMI* is in second place, *wgt* is a bit behind BMI and *hgt* trails the field by a hefty margin.

Example 2: Taking ln's and mixing and matching

In this example take ln's of Brozek and abd and run four models, mixing and matching. In Models (1) and (2) Brozek is first regressed on abd, and then on lnabd; in Models (3) and (4) this is repeated with lnBrozek now the dependent variable.

Here are the results:

```
. esttab, r2 scalar(rss rmse) compress

------------------------------------||--------------------------
                (1)          (2)    ||       (3)          (4)
             Brozek       Brozek    ||  lnBrozek     lnBrozek
------------------------------------||--------------------------
abd           0.585***               ||    0.0337***
             (22.13)                 ||   (17.85)

lnabd                      56.11*** ||                  3.299***
                          (22.83)   ||                (18.99)

_cons        -35.20***    -234.8*** ||    -0.280       -12.07***
            (-14.29)     (-21.12)   ||   (-1.59)     (-15.36)
------------------------------------||--------------------------
N               252          252    ||       251          251
R-sq          0.662        0.676    ||     0.561        0.591
rss          5094.9       4888.8    ||     25.48        23.73
rmse          4.514        4.422    ||     0.320        0.309
------------------------------------||--------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

(Note that ||'s have been manually added to the table… you'll learn why below.)

It is tempting to say that Model (2) is the best because it has the highest R^2 … or maybe you think that Model (4) is the best because it has the lowest MSE/RMSE. Perhaps the different recommendations should be your first clue that R^2's and MSE/RMSE's might not under these circumstances tell you the best model.

The R^2's and MSE/RMSE's in Models (1) and (2) are comparable to one another (since they have the same LHS variable)… and the R^2's and MSE/RMSE's in Models (3) and (4) are also comparable to one another (they also have the same LHS variable). But you cannot, without additional computations, compare the first two Models to the last two Models, because they have different LHS variables.

So Model (2) performs better than (1), and (4) does better than (3)… but don't you dare try to compare (2) and (4) without additional computations. And besides, if you tried to do that, you'd pick (2) on the basis of higher R^2's … or maybe (4) on the basis of lower MSE/RMSE's. Comparability across models with differing LHS variables is clearly an issue.

Now you see why the ||'s are in the results table!