OLS/MLR Assessment I – *Goodness-of-Fit*

- ***Introduction***
- ***A Quick Comparison of SLR and MLR Assessment – Not much that's new!***
- ***MLR Goodness-of-Fit: Adjusted R-squared***
- ***… and MSE (RMSE) (adding and subtracting RHS variables)***
- ***Comparing MLR Models I: Goodness-of-Fit metrics in action***

Introduction

When we looked earlier at assessment in the context of SLR models we discussed two types of measures of performance, *goodness-of-fit* and *precision/inference* metrics. The goodness-of-fit metrics told you something about how well the *predicteds* from your model fit the *actuals*, and the precision/inference metrics said something about how precisely the slope parameter was estimated. The SLR Assessment discussion focused on goodness-of-fit, as precision/inference awaits the topic of statistical inference, and the concepts of hypothesis testing, confidence intervals, and statistical significance.

To review:

OLS/SLR Assessment: Goodness-of-Fit

- Mean Squared Error (MSE): $MSE = \dfrac{SSR}{n-2}$ … an average squared residual, sort of…
- … Root Mean Squared Error (RMSE): $RMSE = \sqrt{MSE}$ … sort of an average residual, but more like a square root of an average squared residual, sort of…
- Coefficient of Determination $\left(R^2\right)$: $R^2 = 1 - \dfrac{SSR}{SST} = \dfrac{SSE}{SST} = \dfrac{S_{\hat{y}\hat{y}}}{S_{yy}} = \rho_{xy}^2 = \rho_{\hat{y}y}^2$ … proportion of the variance of the actuals *explained* by the *predicteds*, as well as the correlation (squared) between the *predicteds* and the *actuals*.

The MSE and RMSE metrics are not in standardized units, making it difficult to interpret the magnitudes. But R^2, which ranges from zero to one, is standardized to some extent, making it perhaps more useful in assessing the performance of the model:

R^2: $0 \le R^2 \le 1$ … closer to one is better…. closer to zero, not so much

… moving to OLS/MLR models and Goodness-of-Fit metrics

We continue to turn to these goodness-of-fit assessment metrics when we move to MLR models, with the formulas changing by not very much, as you'll see below. Of the metrics, however, R^2 proves to be far less useful when assessing performance of MLR models… and so we address that shortcoming with a new *Goodness-of-Fit* metric, *adjusted* R^2 (sometimes *adj* R^2, or $\bar{R}^2$).

The shortcoming of R^2 in the MLR world: R^2 gives credit to variables for *just showing up…*

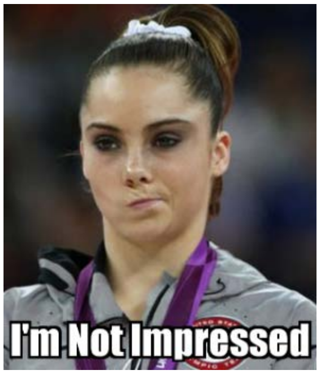
When additional explanatory variables are added to a MLR model, SSRs will typically decrease, or at worst, stay the same… but SSRs can never increase (and R^2 can never decrease) with additional explanatory variables in the model.[1]

And so in the context of the R^2 metric, RHS variables get credit for *just showing up…* irrespective of their explanatory power.

Some Intuition:

Consider a MLR model with min SSRs of SSR_0. If you have an additional RHS variable, then one option in minimizing SSRs is to just keep the coefficient of that new variable equal to zero. But when you minimize SSRs with that restriction, you are solving the previous min SSRs problem, and so the minimum SSR when restricting the new coefficient to be zero is SSR_0, the old SSR.

So with the additional explanatory variable, you can never do worse in minimizing SSRs than SSR_0, where you were before, and you can probably do better once you drop the restriction of that zero coefficient.



- If it turns out that the when minimizing SSRs for the new model, the new variable does in fact have a coefficient of zero, then SSRs will remain at SSR_0… and the new variable has added nothing (no explanatory content) to the model. R^2 is unchanged.
- Alternatively, if the new coefficient is non-zero when minimizing SSRs, then SSRs will necessarily have decreased (so long as the new variable is not perfectly collinear with the other RHS variables in the model). R^2 increases.

What usually happens: When new explanatory variables are added to a model their coefficients will typically be non-zero and R^2 will typically increase. So no one should be

[1] Assuming no changes to the dependent variable.

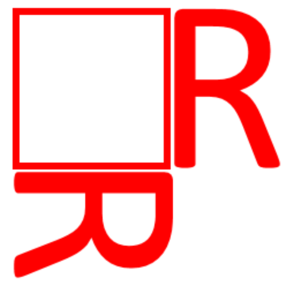
impressed if R^2 increases when new RHS variables are added to the MLR analysis… that's entirely to be expected. Certainly McKayla Maroney is not impressed! Here's an application, which illustrates the point… and tests your understanding:

***MLR Application*: C*orrelations² provide a lower bound on MLR* R^2.**

Suppose you are considering a MLR box office revenues analysis, with explanatory variables *wk1*, *wk2* and *wk3*. Here are the correlations of the variables in the model:

```
. corr rtotgross wk1 wk2 wk3
```

	rtotgr~s	wk1	wk2	wk3
rtotgross	1.0000			
wk1	0.8762	1.0000		
wk2	0.9405	0.9322	1.0000	
wk3	**0.9474**	0.8387	0.9456	1.0000



Notice that the largest correlation between a RHS variable and *rtotgross* is 0.9474 (*wk3*). Then as shown below, the R^2 in the full model must be at least $.9474^2 = .898$, and most likely will be greater. And so the correlations (squared) (in the table above) provide a lower bound on the MLR model R^2. Or put differently: you can often get a pretty good sense of R^2 in a MLR model just by looking at the correlations (squared) amongst the variables.

If you understand the previous comment about RHS variable getting R^2 credit for just showing up, you'll understand why I claim that the R^2 in the full MLR model will have an R^2 of at least $.9474^2 = .898$.

Here's why: To get to the MLR model, let's start with the SLR model in which *rtotgross* has been regressed on *wk3*. I pick wk3 because of the three RHS variables, it has the highest correlation with *rtotgross*. Since $R^2 = \rho_{xy}^2$ for SLR models, we know that the R^2 for this SLR model will be $.9474^2 = .898$. See Model (1) below…. and build to Model (3):

	(1) rtotgross	(2) rtotgross	(3) rtotgross
wk3	7.175***	5.427***	4.778***
	(260.32)	(121.38)	(59.84)
wk1		0.735***	0.540***
		(46.60)	(21.36)
wk2			0.745***
			(9.79)
_cons	0.0390	-0.540*	-0.601**
	(0.15)	(-2.36)	(-2.64)
N	7730	7730	7730
R-sq	0.898	0.920	0.921
SSR	3026710.9	2362741.1	2333803.6

OLS/MLR Assessment I – *Goodness-of-Fit*

As predicted, R^2's are increasing moving left to right, since the coefficients for the new variables are non-zero: R^2 increases from .898 in Model (1) to .920 in Model (2), and to .921 in Model (3). And also as predicted, SSRs are decreasing.

And so we can use simple pairwise correlations together with the fact that R^2 will never decrease when additional RHS variable are added to a model, to place a lower bound on R^2 for the final MLR model… or put differently, the simple correlations alone tell you that the final MLR model will have a very high R^2… close to 1.

The following table compares the various Goodness-of-Fit concepts/definitions/formulas in SLR and MLR models. (Note that I assume that there is always a constant term in the SLR and MLR models.)

A Quick Comparison of SLR and MLR Assessment – Goodness-of-fit

	SLR	**MLR**
Sum Squares	SST = SSE + SSR	SST = SSE + SSR
R^2 (Coefficient of Determination) (w/ intercept term)	$R^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST}$ $= \rho_{xy}^2 = \rho_{\hat{y}y}^2$	$R^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST}$ $= \rho_{\hat{y}y}^2$
	$R^2 = \frac{SampleVar(predicted)}{SampleVar(actual)}$	$R^2 = \frac{SampleVar(predicted)}{SampleVar(actual)}$
Degrees of freedom (*dofs*)	$dofs = n - 2$	$dofs = n - k - 1$
MSE	$MSE = \frac{SSR}{dofs} = \frac{SSR}{n-2}$	$MSE = \frac{SSR}{dofs} = \frac{SSR}{n-k-1}$
RMSE	$RMSE = \sqrt{MSE}$	$RMSE = \sqrt{MSE}$
Adjusted R^2		$\bar{R}^2 = 1 - \frac{SSR}{SST}\frac{n-1}{n-k-1}$ $= 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{MSE}{S_{yy}}$

So what's new with MLR?... Not much, really!

As you can see, there are a few differences between SLR and MLR models, but not many!

1. The definitions of SSR, SSE, SST are the same for SLR and MLR models… as is the definition of R^2, and the fact that SST = SSE + SSR (since there is a constant term in the model).
2. In the MSE calculation, we now divide by *n-k-1*, the *degrees of freedom* (*dofs*) in the MLR model. This reflects an interest in unbiasedness … to be discussed later. This is in fact consistent with the SLR metric, since there were *n-2 dofs* in those models.
3. We have new metric for MLR models, *Adjusted* R^2, discussed in more detail below. In contrast to R^2 , and as discussed above, *Adjusted* R^2 will not give new RHS variables goodness-of-fit credit merely for *just showing up*. New RHS variables have to impress (in reducing SSRs by more than some trivial amount) for *Adjusted* R^2 to increase.

MLR Goodness-of-fit: *Adjusted R-squared*

As discussed above, when you are adding explanatory variables to a MLR model (and not changing the y's or number of observations), SSRs will always decrease (and R-sq will always increase) unless the estimated MLR coefficient for the new variable is exactly 0 (or the new variable is perfectly collinear with the other RHS variables already in the model).

So nobody should be impressed if R^2 increases when additional explanatory variables are brought into the analysis. You knew that would happen!

The question should be: By how much did R^2 increase? If R^2 increased a lot, then you should be impressed; but if it increased by not so much, then maybe you'll want to hold your applause.

$\overline{R}^2$

Adjusted R^2 is an attempt to adjust the *coefficient of determination* for this shortcoming. You'll discover that smallish decreases in SSRs will not generate a higher adj R^2; but larger decreases will… and what is *small* or *large* will depend in part on how many additional variables were added to the model..

Adjusted R-squared, *Adj* R^2, is often (and rather opaquely) defined as:

$$\bar{R}^2 = 1-(1-R^2)\left[\frac{n-1}{n-k-1}\right].$$

I don't know about you, but I find that that formula tells me nothing.

Since $R^2 = 1-\frac{SSR}{SST}$, a more easily interpreted expression for *Adj* R^2 is:

$$\bar{R}^2 = 1-\left[\frac{SSR}{SST}\right]\left[\frac{n-1}{n-k-1}\right] = 1-\left[\frac{SSR}{SST}\right]\left[\frac{n-1}{dofs}\right],$$ [2]

[2] Recall that we sometimes refer to n-k-1 as the (number of) *degrees of freedom* (*dofs*) in the model.

which looks a lot like the definition of R^2 (with a $\left[\frac{n-1}{dofs}\right]$ adjustment).

Note that since $\frac{(n-1)}{(n-k-1)} = \frac{(n-1)}{dofs} > 1$, $\left[\frac{SSR}{SST}\right]\left[\frac{n-1}{dofs}\right] > \left[\frac{SSR}{SST}\right]$ and accordingly, $\bar{R}^2 < R^2 \le 1$ for $k > 0$, with the difference inversely related to k.

And so *adjusted R^2* is always bounded above by 1.[3]

***Interpretation of Adj R^2*:** It's all about the rates of change of *SSRs* and *dofs*.

We can rewrite the previous expression for *Adjusted R^2*:

$$\bar{R}^2 = 1 - \frac{(n-1)}{SST}\left[\frac{SSR}{dofs}\right].$$

As you add explanatory variables to the model, only the terms in the square brackets (*SSR* and *dofs*) are changing, and both $(SSRs \text{ and } dofs)$ are typically declining. And so whether $\bar{R}^2$ increases or decreases will depend on the relative rates of change of *SSRs* and *dofs*:

- If the decline in SSRs is faster than the decline in *dofs*, then $\left[\frac{SSR}{dofs}\right]$ will decline and $\bar{R}^2$ will increase with the additional explanatory variables.
- But if the decline in SSRs is slower than the decline in *dofs*, then $\left[\frac{SSR}{dofs}\right]$ will increase , and $\bar{R}^2$ will decrease.

So for *Adjusted R^2* to increase, it must be the case that SSRs are dropping faster than dofs.

… and MSE (RMSE) (adding and subtracting RHS variables)

Since $\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{MSE}{S_{yy}}$, *adjusted R^2* and *MSE* will always move in opposite directions when S_{yy} is fixed. So if you are adding (or subtracting) RHS variables to (or from) a MLR model (and not impacting S_{yy}), you should expect to see $\bar{R}^2$ and *MSE* moving in exactly opposite directions.

Accordingly, the two goodness-of-fit metrics (*adjusted R^2* and *MSE/RMSE)* are effectively redundant in the sense that knowing the movements patterns of one tells you the movements of the other.

[3] *Adjusted R^2* can be negative, though that rarely happens in practice.. If you see that, you have a really really really bad model! Time to find a new profession!

An important difference however is that while we don't necessarily have a good sense of when *MSE*s (*RMSE*s) are small or large, we do know that $\bar{R}^2 \leq 1$, and so we typically have an easier time evaluating magnitudes of $\bar{R}^2$.

Note however that since R^2 and $\bar{R}^2$ do not necessarily move in the same direction, *MSE*s and R^2 will not necessarily move in opposite directions. That was not the case for SLR models.

Comparing MLR Models I: Goodness-of-Fit metrics in action

To illustrate Goodness-of-Fit metrics in action, here's an example using the bodyfat dataset.

In Model (1), the *Brozek* measure of bodyfat had been regressed on *hgt* and *wgt*.

```
. esttab, r2 ar2 scalar (rmse) compress

--------------------------------------------------------------
                  (1)          (2)          (3)          (4)
               Brozek       Brozek       Brozek       Brozek
--------------------------------------------------------------
hgt            -0.650***    -0.118       -0.131       -0.138
              (-6.29)      (-1.43)      (-1.51)      (-1.55)

wgt             0.187***    -0.120***    -0.108**     -0.100*
              (14.48)      (-5.41)      (-3.18)      (-2.52)

abd                          0.880***     0.883***     0.898***
                           (15.19)      (15.13)      (12.62)

hip                                     -0.0564      -0.0723
                                        (-0.49)      (-0.58)

chest                                                -0.0348
                                                     (-0.38)

_cons           31.16***    -32.66***    -28.64**     -25.86*
               (4.51)      (-5.01)      (-2.71)      (-2.01)
--------------------------------------------------------------
N                 252          252          252          252
R-sq           0.4614       0.7210       0.7213       0.7215
adj. R-sq      0.4571       0.7177       0.7168       0.7158
rmse           5.7109       4.1184       4.1248       4.1320
--------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Note the *esttab* options: r2 $\left(R^2\right)$, ar2 $\left(\bar{R}^2\right)$, and rmse $\left(RMSE\right)$.

- In Model (2), *abd* has been added to Model (1), and *R-sq* and *adj. R-sq* both increase, while *RMSE* declines.
- In Model (3) *hip* has been added in, with *R-sq* continuing to increase as it almost always will. Now, however, *adj. R-sq* declines and *RMSE* increases. As always, *adj R-sq* and *RMSE* are moving in opposite directions.
- And in going to Model (4), with *chest* added to the model, *R-sq* continues to (slightly) increase, while *adj R-sq* again declines and *RMSE* again increases.

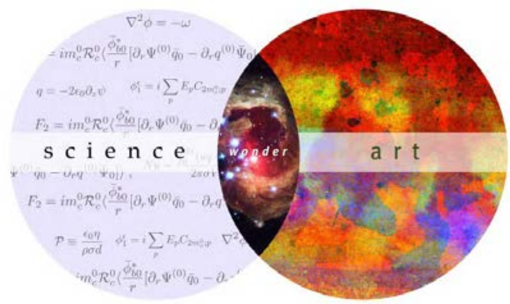
Recall that with SLR models, we could use *R-sq* to compare the performance of different models having the same dependent variable. In the MLR world, we often use *adj* R^2's to compare models, so long as the dependent variables are the same… though I'd be the last to suggest that you should only look at *adj* R^2.

Applying this criterion to the previous set of four MLR models, Model (2) is the best performer since it has the highest *adj R-sq* and the lowest *RMSE.* But all of the Models tell you something… so don't ignore the others, just because their performance stats aren't as impressive!

Art v. Science

Comparing the performance of MLR models ***is as much art as science*** … and in truth, we typically look at a number of different aspects/properties of the model. But certainly *adj R-sq* and *RMSE* are in the conversation.

We'll return to this topic later, and focus on the different criteria at play in assessing the performance of the three types of econometrics models discussed in the Introduction:



- **Forecasting models** (*less is more*; focus on out-of-sample forecasting, and don't over-fit the data)
- **Behavioral models** (*parsimony preferred*; the challenging art form)
- **Favorite coefficient models** (*more is more*; focus on the favorite coefficient… and don't worry about the other aspects of the model… other than making sure that you really have included every possible relevant explanatory variable, and accordingly that you have minimized the possibility of omitted variable impact/bias)