Regression Specification Error Test (RESET test)

The next two slides are parts of pages 306 and 307 of Wooldridge's (2013) Introductory Econometrics: A Modern Approach, Fifth Edition.

Ramsey (1969) Tests for Specification Errors in Classical Linear Least-Squares Analysis. Journal of the Royal Statistical Association, Series B, 71, 350–371.

RESET as a General Test for Functional Form Misspecification

Some tests have been proposed to detect general functional form misspecification. Ramsey's (1969) **regression specification error test** (**RESET**) has proven to be useful in this regard.

The idea behind RESET is fairly simple. If the original model

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u \qquad [9.2]$$

satisfies MLR.4, then no nonlinear functions of the independent variables should be significant when added to equation (9.2). In Example 9.1, we added quadratics in the significant explanatory variables. Although this often detects functional form problems, it has the drawback of using up many degrees of freedom if there are many explanatory variables in the original model (much as the straight form of the White test for heteroskedasticity consumes degrees of freedom). Further, certain kinds of neglected nonlinearities will not be picked up by adding quadratic terms. RESET adds polynomials in the OLS fitted values to equation (9.2) to detect general kinds of functional form misspecification.

To implement RESET, we must decide how many functions of the fitted values to include in an expanded regression. There is no right answer to this question, but the squared and cubed terms have proven to be useful in most applications.

Let $\hat{y}$ denote the OLS fitted values from estimating (9.2). Consider the expanded equation

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + error. \qquad [9.3]$$

This equation seems a little odd, because functions of the fitted values from the initial estimation now appear as explanatory variables. In fact, we will not be interested in the estimated parameters from (9.3); we only use this equation to test whether (9.2) has missed important nonlinearities. The thing to remember is that $\hat{y}^2$ and $\hat{y}^3$ are just nonlinear functions of the x_j.

The null hypothesis is that (9.2) is correctly specified. Thus, RESET is the F statistic for testing $H_0: \delta_1 = 0, \delta_2 = 0$ in the expanded model (9.3). A significant F statistic suggests some sort of functional form problem. The distribution of the F statistic is approximately $F_{2,n-k-3}$ in large samples under the null hypothesis (and the Gauss-Markov assumptions). The *df* in the expanded equation (9.3) is $n - k - 1 - 2 = n - k - 3$. An *LM* version is also available (and the chi-square distribution will have two *df*). Further, the test can be made robust to heteroskedasticity using the methods discussed in Section 8.2.

EXAMPLE 9.2 HOUSING PRICE EQUATION

We estimate two models for housing prices. The first one has all variables in level form:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u. \tag{9.4}$$

The second one uses the logarithms of all variables except *bdrms*:

$$lprice = \beta_0 + \beta_1 llotsize + \beta_2 lsqrft + \beta_3 bdrms + u. \tag{9.5}$$

Using $n = 88$ houses in HPRICE1.RAW, the RESET statistic for equation (9.4) turns out to be 4.67; this is the value of an $F_{2,82}$ random variable ($n = 88$, $k = 3$), and the associated p-value is .012. This is evidence of functional form misspecification in (9.4).

The RESET statistic in (9.5) is 2.56, with p-value = .084. Thus, we do not reject (9.5) at the 5% significance level (although we would at the 10% level). On the basis of RESET, the log-log model in (9.5) is preferred.

In the previous example, we tried two models for explaining housing prices. One was rejected by RESET, while the other was not (at least at the 5% level). Often, things are not so simple. A drawback with RESET is that it provides no real direction on how to proceed if the model is rejected. Rejecting (9.4) by using RESET does not immediately suggest that (9.5) is the next step. Equation (9.5) was estimated because constant elasticity models are easy to interpret and can have nice statistical properties. In this example, it so happens that it passes the functional form test as well.

Some have argued that RESET is a very general test for model misspecification, including unobserved omitted variables and heteroskedasticity. Unfortunately, such use of RESET is largely misguided. It can be shown that RESET has no power for detecting omitted variables whenever they have expectations that are linear in the included independent variables in the model [see Wooldridge (1995) for a precise statement]. Further, if the functional form is properly specified, RESET has no power for detecting heteroskedasticity. The bottom line is that RESET is a functional form test, and nothing more.

```
Call:
lm(formula = price ~ lotsize + sqrft + bdrms)

Residuals:
     Min       1Q   Median       3Q      Max 
-120.026  -38.530   -6.555   32.323  209.376 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.177e+01  2.948e+01  -0.739  0.46221    
lotsize      2.068e-03  6.421e-04   3.220  0.00182 ** 
sqrft        1.228e-01  1.324e-02   9.275 1.66e-14 ***
bdrms        1.385e+01  9.010e+00   1.537  0.12795    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 59.83 on 84 degrees of freedom
Multiple R-squared:  0.6724,	Adjusted R-squared:  0.6607 
F-statistic: 57.46 on 3 and 84 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = lprice ~ llotsize + lsqrft + bdrms)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.68422 -0.09178 -0.01584  0.11213  0.66899 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.29704    0.65128  -1.992   0.0497 *  
llotsize     0.16797    0.03828   4.388 3.31e-05 ***
lsqrft       0.70023    0.09287   7.540 5.01e-11 ***
bdrms        0.03696    0.02753   1.342   0.1831    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1846 on 84 degrees of freedom
Multiple R-squared:  0.643,	Adjusted R-squared:  0.6302 
F-statistic: 50.42 on 3 and 84 DF,  p-value: < 2.2e-16
```

Reset test

price on lotsize,sqrft,bdrms

F statistic = 4.66820549

Critical value = 3.10789130

P-value = 0.01202171

lprice on llotsize,lsqrft,bdrms

F statistic = 2.56504156

Critical value = 3.10789130

P-value = 0.08307583