複迴歸結果解釋

(取自 Wooldridge Chap 4)

這份講義取材自 Wooldridge Chap 4 的例題，請同學閱讀作者的敘述來學習:

1. 如何表達配適的模型
2. 如何陳述分析的結果

EXAMPLE 4.2 Student Performance and School Size

There is much interest in the effect of school size on student performance. (See, for example, *The New York Times Magazine*, 5/28/95.) One claim is that, everything else being equal, students at smaller schools fare better than those at larger schools. This hypothesis is assumed to be true even after accounting for differences in class sizes across schools.

The file MEAP93 contains data on 408 high schools in Michigan for the year 1993. We can use these data to test the null hypothesis that school size has no effect on standardized test scores against the alternative that size has a negative effect. Performance is measured by the percentage of students receiving a passing score on the Michigan Educational Assessment Program (MEAP) standardized tenth-grade math test (*math*10). School size is measured by student enrollment (*enroll*). The null hypothesis is H_0: $\beta_{enroll} = 0$, and the alternative is H_1: $\beta_{enroll} < 0$. For now, we will control for two other factors, average annual teacher compensation (*totcomp*) and the number of staff per one thousand students (*staff*). Teacher compensation is a measure of teacher quality, and staff size is a rough measure of how much attention students receive.

The estimated equation, with standard errors in parentheses, is

$$\begin{aligned}\widehat{math10} &= \underset{(6.113)}{2.274} + \underset{(.00010)}{.00046}\, totcomp + \underset{(.040)}{.048}\, staff - \underset{(.00022)}{.00020}\, enroll\\ n &= 408,\ R^2 = .0541.\end{aligned}$$

The coefficient on *enroll*, $-.00020$, is in accordance with the conjecture that larger schools hamper performance: higher enrollment leads to a lower percentage of students with a passing tenth-grade math score. (The coefficients on *totcomp* and *staff* also have the signs we expect.) The fact that *enroll* has an estimated coefficient different from zero could just be due to sampling error; to be convinced of an effect, we need to conduct a *t* test.

Since $n - k - 1 = 408 - 4 = 404$, we use the standard normal critical value. At the 5% level, the critical value is -1.65; the *t* statistic on *enroll* must be *less* than -1.65 to reject H_0 at the 5% level.

The *t* statistic on *enroll* is $-.00020/.00022 \approx -.91$ which is larger than -1.65: we *fail* to reject H_0 in favor of H_1 at the 5% level. In fact, the 15% critical value is -1.04, and since $-.91 > -1.04$, we fail to reject H_0 even at the 15% level. We conclude that *enroll* is not statistically significant at the 15% level.

The variable *totcomp* is statistically significant even at the 1% significance level because its *t* statistic is 4.6. On the other hand, the *t* statistic for *staff* is 1.2, and so we cannot reject H_0: $\beta_{staff} = 0$ against H_1: $\beta_{staff} > 0$ even at the 10% significance level. (The critical value is $c = 1.28$ from the standard normal distribution.)

To illustrate how changing functional form can affect our conclusions, we also estimate the model with all independent variables in logarithmic form. This allows, for example, the school size effect to diminish as school size increases. The estimated equation is

$$\begin{aligned}\widehat{math10} &= \underset{(48.70)}{-207.66} + \underset{(4.06)}{21.16}\log(totcomp) + \underset{(4.19)}{3.98}\log(staff) - \underset{(0.69)}{1.29}\log(enroll)\\ n &= 408,\ R^2 = .0654.\end{aligned}$$

The *t* statistic on log(*enroll*) is about -1.87; since this is below the 5% critical value -1.65, we reject H_0: $\beta_{\log(enroll)} = 0$ in favor of H_1: $\beta_{\log(enroll)} < 0$ at the 5% level.

In Chapter 2, we encountered a model where the dependent variable appeared in its original form (called *level* form), while the independent variable appeared in log form (called *level-log* model). The interpretation of the parameters is the same in the multiple regression context, except, of course, that we can give the parameters a ceteris paribus interpretation. Holding *totcomp* and *staff* fixed, we have $\Delta\widehat{math10} = -1.29[\Delta\log(enroll)]$, so that

$$\Delta\widehat{math10} \approx -(1.29/100)(\%\Delta enroll) \approx -.013(\%\Delta enroll).$$

Once again, we have used the fact that the change in log($enroll$), when multiplied by 100, is approximately the percentage change in $enroll$. Thus, if enrollment is 10% higher at a school, $\widehat{math10}$ is predicted to be $.013(10) = 0.13$ percentage points lower ($math10$ is measured as a percentage).

Which model do we prefer: the one using the level of $enroll$ or the one using log($enroll$)? In the level-level model, enrollment does not have a statistically significant effect, but in the level-log model it does. This translates into a higher R-squared for the level-log model, which means we explain more of the variation in $math10$ by using $enroll$ in logarithmic form (6.5% to 5.4%). The level-log model is preferred because it more closely captures the relationship between $math10$ and $enroll$. We will say more about using R-squared to choose functional form in Chapter 6.

EXAMPLE 4.3 Determinants of College GPA

We use the data in GPA1 to estimate a model explaining college GPA ($colGPA$), with the average number of lectures missed per week ($skipped$) as an additional explanatory variable. The estimated model is

$$\begin{aligned}\widehat{colGPA} &= \underset{(.33)}{1.39} + \underset{(.094)}{.412}\, hsGPA + \underset{(.011)}{.015}\, ACT - \underset{(.026)}{.083}\, skipped \\ n &= 141,\ R^2 = .234.\end{aligned}$$

We can easily compute t statistics to see which variables are statistically significant, using a two-sided alternative in each case. The 5% critical value is about 1.96, since the degrees of freedom $(141 - 4 = 137)$ is large enough to use the standard normal approximation. The 1% critical value is about 2.58.

The t statistic on $hsGPA$ is 4.38, which is significant at very small significance levels. Thus, we say that "$hsGPA$ is statistically significant at any *conventional* significance level." The t statistic on ACT is 1.36, which is not statistically significant at the 10% level against a two-sided alternative. The coefficient on ACT is also practically small: a 10-point increase in ACT, which is large, is predicted to increase $colGPA$ by only .15 points. Thus, the variable ACT is practically, as well as statistically, insignificant.

The coefficient on $skipped$ has a t statistic of $-.083/.026 = -3.19$, so $skipped$ is statistically significant at the 1% significance level $(3.19 > 2.58)$. This coefficient means that another lecture missed per week lowers predicted $colGPA$ by about .083. Thus, holding $hsGPA$ and ACT fixed, the predicted difference in $colGPA$ between a student who misses no lectures per week and a student who misses five lectures per week is about .42. Remember that this says nothing about specific students; rather, .42 is the estimated average across a subpopulation of students.

In this example, for each variable in the model, we could argue that a one-sided alternative is appropriate. The variables $hsGPA$ and $skipped$ are very significant using a two-tailed test and have the signs that we expect, so there is no reason to do a one-tailed test. On the other hand, against a one-sided alternative $(\beta_3 > 0)$, ACT is significant at the 10% level but not at the 5% level. This does not change the fact that the coefficient on ACT is pretty small.

EXAMPLE 4.4 Campus Crime and Enrollment

Consider a simple model relating the annual number of crimes on college campuses (*crime*) to student enrollment (*enroll*):

$$\log(crime) = \beta_0 + \beta_1 \log(enroll) + u.$$

This is a constant elasticity model, where β_1 is the elasticity of crime with respect to enrollment. It is not much use to test H_0: $\beta_1 = 0$, as we expect the total number of crimes to increase as the size of the campus increases. A more interesting hypothesis to test would be that the elasticity of crime with respect to enrollment is one: H_0: $\beta_1 = 1$. This means that a 1% increase in enrollment leads to, on average, a 1% increase in crime. A noteworthy alternative is H_1: $\beta_1 > 1$, which implies that a 1% increase in enrollment increases campus crime by *more* than 1%. If $\beta_1 > 1$, then, in a relative sense—not just an absolute sense—crime is more of a problem on larger campuses. One way to see this is to take the exponential of the equation:

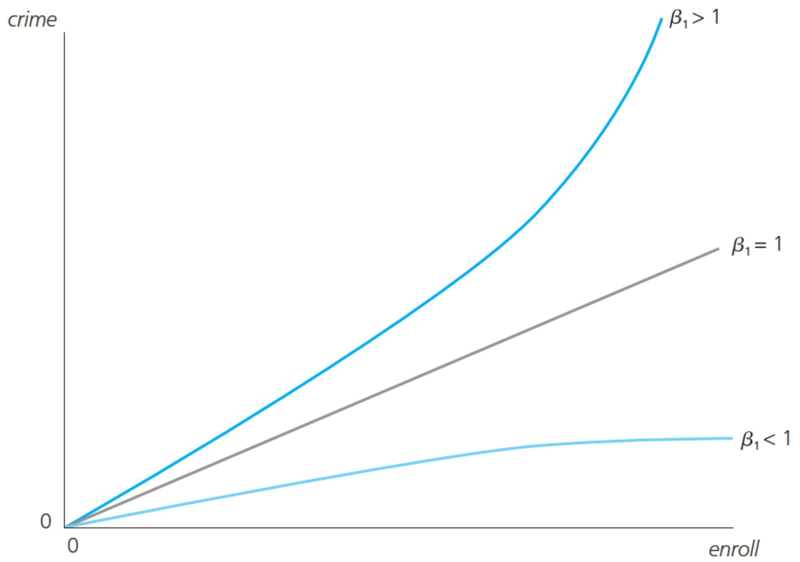
$$crime = \exp(\beta_0) enroll^{\beta_1} \exp(u).$$

(See Appendix A for properties of the natural logarithm and exponential functions.) For $\beta_0 = 0$ and $u = 0$, this equation is graphed in Figure 4.5 for $\beta_1 < 1$, $\beta_1 = 1$, and $\beta_1 > 1$.

We test $\beta_1 = 1$ against $\beta_1 > 1$ using data on 97 colleges and universities in the United States for the year 1992, contained in the data file CAMPUS. The data come from the FBI's *Uniform Crime Reports*, and the average number of campus crimes in the sample is about 394, while the average enrollment is about 16,076. The estimated equation (with estimates and standard errors rounded to two decimal places) is

$$\begin{aligned} \widehat{\log(crime)} &= \underset{(1.03)}{-6.63} + \underset{(0.11)}{1.27} \log(enroll) \\ n &= 97,\ R^2 = .585. \end{aligned} \qquad [4.14]$$

Figure 4.5 Graph of $crime = enroll^{\beta_1}$ for $\beta_1 < 1$, $\beta_1 = 1$, and $\beta_1 > 1$.



The estimated elasticity of *crime* with respect to *enroll*, 1.27, is in the direction of the alternative $\beta_1 > 1$. But is there enough evidence to conclude that $\beta_1 > 1$? We need to be careful in testing this hypothesis, especially because the statistical output of standard regression packages is much more complex than the simplified output reported in equation (4.14). Our first instinct might be to construct "the" t statistic by taking the coefficient on log(*enroll*) and dividing it by its standard error, which is the t statistic reported by a regression package. But this is the *wrong* statistic for testing H_0: $\beta_1 = 1$. The correct t statistic is obtained from (4.13): we subtract the hypothesized value, unity, from the estimate and divide the result by the standard error of $\hat{\beta}_1$: $t = (1.27 - 1)/.11 = .27/.11 \approx 2.45$. The one-sided 5% critical value for a t distribution with $97 - 2 = 95$ df is about 1.66 (using $df = 120$), so we clearly reject $\beta_1 = 1$ in favor of $\beta_1 > 1$ at the 5% level. In fact, the 1% critical value is about 2.37, and so we reject the null in favor of the alternative at even the 1% level.

We should keep in mind that this analysis holds no other factors constant, so the elasticity of 1.27 is not necessarily a good estimate of ceteris paribus effect. It could be that larger enrollments are correlated with other factors that cause higher crime: larger schools might be located in higher crime areas. We could control for this by collecting data on crime rates in the local city.

EXAMPLE 4.5 Housing Prices and Air Pollution

For a sample of 506 communities in the Boston area, we estimate a model relating median housing price (*price*) in the community to various community characteristics: *nox* is the amount of nitrogen oxide in the air, in parts per million; *dist* is a weighted distance of the community from five employment centers, in miles; *rooms* is the average number of rooms in houses in the community; and *stratio* is the average student-teacher ratio of schools in the community. The population model is

$$\log(price) = \beta_0 + \beta_1\log(nox) + \beta_2\log(dist) + \beta_3 rooms + \beta_4 stratio + u.$$

Thus, β_1 is the elasticity of *price* with respect to *nox*. We wish to test H_0: $\beta_1 = -1$ against the alternative H_1: $\beta_1 \neq -1$. The t statistic for doing this test is $t = (\hat{\beta}_1 + 1)/\mathrm{se}(\hat{\beta}_1)$.

Using the data in HPRICE2, the estimated model is

$$\begin{aligned}\widehat{\log(price)} &= \underset{(0.32)}{11.08} - \underset{(.117)}{.954}\log(nox) - \underset{(.043)}{.134}\log(dist) + \underset{(.019)}{.255}\,rooms - \underset{(.006)}{.052}\,stratio\\ n &= 506,\ R^2 = .581.\end{aligned}$$

The slope estimates all have the anticipated signs. Each coefficient is statistically different from zero at very small significance levels, including the coefficient on log(*nox*). But we do not want to test that $\beta_1 = 0$. The null hypothesis of interest is H_0: $\beta_1 = -1$, with corresponding t statistic $(-.954 + 1)/.117 = .393$. There is little need to look in the t table for a critical value when the t statistic is this small: the estimated elasticity is not statistically different from -1 even at very large significance levels. Controlling for the factors we have included, there is little evidence that the elasticity is different from -1.

EXAMPLE 4.6 Participation Rates in 401(k) Plans

In Example 3.3, we used the data on 401(k) plans to estimate a model describing participation rates in terms of the firm's match rate and the age of the plan. We now include a measure of firm size, the total number of firm employees (*totemp*). The estimated equation is

$$\begin{aligned}\widehat{prate} &= \underset{(0.78)}{80.29} + \underset{(0.52)}{5.44}\, mrate + \underset{(.045)}{.269}\, age - \underset{(.00004)}{.00013}\, totemp \\ n &= 1{,}534,\ R^2 = .100.\end{aligned}$$

The smallest t statistic in absolute value is that on the variable *totemp*: $t = -.00013/.00004 = -3.25$, and this is statistically significant at very small significance levels. (The two-tailed p-value for this t statistic is about .001.) Thus, all of the variables are statistically significant at rather small significance levels.

How big, in a practical sense, is the coefficient on *totemp*? Holding *mrate* and *age* fixed, if a firm grows by 10,000 employees, the participation rate falls by $10{,}000(.00013) = 1.3$ percentage points. This is a huge increase in number of employees with only a modest effect on the participation rate. Thus, although firm size does affect the participation rate, the effect is not practically very large.

EXAMPLE 4.7 Effect of Job Training on Firm Scrap Rates

The scrap rate for a manufacturing firm is the number of defective items—products that must be discarded—out of every 100 produced. Thus, for a given number of items produced, a decrease in the scrap rate reflects higher worker productivity.

We can use the scrap rate to measure the effect of worker training on productivity. Using the data in JTRAIN, but only for the year 1987 and for nonunionized firms, we obtain the following estimated equation:

$$\begin{aligned}\widehat{\log(scrap)} &= \underset{(5.69)}{12.46} - \underset{(.023)}{.029}\, hrsemp - \underset{(.453)}{.962}\, \log(sales) + \underset{(.407)}{.761}\, \log(employ) \\ n &= 29,\ R^2 = .262.\end{aligned}$$

The variable *hrsemp* is annual hours of training per employee, *sales* is annual firm sales (in dollars), and *employ* is the number of firm employees. For 1987, the average scrap rate in the sample is about 4.6 and the average of *hrsemp* is about 8.9.

The main variable of interest is *hrsemp*. One more hour of training per employee lowers log(*scrap*) by .029, which means the scrap rate is about 2.9% lower. Thus, if *hrsemp* increases by 5—each employee is trained 5 more hours per year—the scrap rate is estimated to fall by $5(2.9) = 14.5\%$. This seems like a reasonably large effect, but whether the additional training is worthwhile to the firm depends on the cost of training and the benefits from a lower scrap rate. We do not have the numbers needed to do a cost benefit analysis, but the estimated effect seems nontrivial.

What about the *statistical significance* of the training variable? The t statistic on *hrsemp* is $-.029/.023 = -1.26$, and now you probably recognize this as not being large enough in magnitude to conclude that *hrsemp* is statistically significant at the 5% level. In fact, with $29 - 4 = 25$ degrees of freedom for the one-sided alternative, H_1: $\beta_{hrsemp} < 0$, the 5% critical value is about -1.71. Thus, using a strict 5% level test, we must conclude that *hrsemp* is not statistically significant, even using a one-sided alternative.

Because the sample size is pretty small, we might be more liberal with the significance level. The 10% critical value is -1.32, and so *hrsemp* is almost significant against the one-sided alternative at the 10% level. The p-value is easily computed as $P(T_{25} < -1.26) = .110$. This may be a low enough p-value to conclude that the estimated effect of training is not just due to sampling error, but opinions would legitimately differ on whether a one-sided p-value of .11 is sufficiently small.

EXAMPLE 4.8 Model of R&D Expenditures

Economists studying industrial organization are interested in the relationship between firm size—often measured by annual sales—and spending on research and development (R&D). Typically, a constant elasticity model is used. One might also be interested in the ceteris paribus effect of the profit margin—that is, profits as a percentage of sales—on R&D spending. Using the data in RDCHEM on 32 U.S. firms in the chemical industry, we estimate the following equation (with standard errors in parentheses below the coefficients):

$$\begin{aligned}\widehat{\log(rd)} &= \underset{(.47)}{-4.38} + \underset{(.060)}{1.084}\log(sales) + \underset{(.0128)}{.0217}\, profmarg \\ n &= 32,\ R^2 = .918.\end{aligned}$$

The estimated elasticity of R&D spending with respect to firm sales is 1.084, so that, holding profit margin fixed, a 1% increase in sales is associated with a 1.084% increase in R&D spending. (Incidentally, R&D and sales are both measured in millions of dollars, but their units of measurement have no effect on the elasticity estimate.) We can construct a 95% confidence interval for the sales elasticity once we note that the estimated model has $n - k - 1 = 32 - 2 - 1 = 29$ degrees of freedom. From Table G.2, we find the 97.5$^{\text{th}}$ percentile in a t_{29} distribution: $c = 2.045$. Thus, the 95% confidence interval for $\beta_{\log(sales)}$ is $1.084 \pm .060(2.045)$, or about (.961,1.21). That zero is well outside this interval is hardly surprising: we expect R&D spending to increase with firm size. More interesting is that unity is included in the 95% confidence interval for $\beta_{\log(sales)}$, which means that we cannot reject H_0: $\beta_{\log(sales)} = 1$ against H_1: $\beta_{\log(sales)} \neq 1$ at the 5% significance level. In other words, the estimated R&D-sales elasticity is not statistically different from 1 at the 5% level. (The estimate is not practically different from 1, either.)

The estimated coefficient on *profmarg* is also positive, and the 95% confidence interval for the population parameter, $\beta_{profmarg}$, is $.0217 \pm .0128(2.045)$, or about $(-.0045,.0479)$. In this case, zero is included in the 95% confidence interval, so we fail to reject H_0: $\beta_{profmarg} = 0$ against H_1: $\beta_{profmarg} \neq 0$ at the 5% level. Nevertheless, the t statistic is about 1.70, which gives a two-sided p-value of about .10, and so we would conclude that *profmarg* is statistically significant at the 10% level against the two-sided alternative, or at the 5% level against the one-sided alternative H_1: $\beta_{profmarg} > 0$. Plus, the economic size of the profit margin coefficient is not trivial: holding *sales* fixed, a one percentage point increase in *profmarg* is estimated to increase R&D spending by $100(.0217) \approx 2.2\%$. A complete analysis of this example goes beyond simply stating whether a particular value, zero in this case, is or is not in the 95% confidence interval.

EXAMPLE 4.9 **Parents' Education in a Birth Weight Equation**

As another example of computing an F statistic, consider the following model to explain child birth weight in terms of various factors:

$$bwght = \beta_0 + \beta_1 cigs + \beta_2 parity + \beta_3 faminc + \beta_4 motheduc + \beta_5 fatheduc + u, \quad [4.42]$$

where

$bwght$ = birth weight, in pounds.
$cigs$ = average number of cigarettes the mother smoked per day during pregnancy.
$parity$ = the birth order of this child.
$faminc$ = annual family income.
$motheduc$ = years of schooling for the mother.
$fatheduc$ = years of schooling for the father.

Let us test the null hypothesis that, after controlling for *cigs*, *parity*, and *faminc*, parents' education has no effect on birth weight. This is stated as H_0: $\beta_4 = 0$, $\beta_5 = 0$, and so there are $q = 2$ exclusion restrictions to be tested. There are $k + 1 = 6$ parameters in the unrestricted model (4.42); so the *df* in the unrestricted model is $n - 6$, where n is the sample size.

We will test this hypothesis using the data in BWGHT. This data set contains information on 1,388 births, but we must be careful in counting the observations used in testing the null hypothesis. It turns out that information on at least one of the variables *motheduc* and *fatheduc* is missing for 197 births in the sample; these observations cannot be included when estimating the unrestricted model. Thus, we really have $n = 1{,}191$ observations, and so there are $1{,}191 - 6 = 1{,}185$ *df* in the unrestricted model. We must be sure to use these *same* 1,191 observations when estimating the restricted model (not the full 1,388 observations that are available). Generally, when estimating the restricted model to compute an F test, we must use the same observations to estimate the unrestricted model; otherwise, the test is not valid. When there are no missing data, this will not be an issue.

The numerator *df* is 2, and the denominator *df* is 1,185; from Table G.3, the 5% critical value is $c = 3.0$. Rather than report the complete results, for brevity, we present only the R-squareds. The R-squared for the full model turns out to be $R^2_{ur} = .0387$. When *motheduc* and *fatheduc* are dropped from the regression, the R-squared falls to $R^2_r = .0364$. Thus, the F statistic is $F = [(.0387 - .0364)/(1 - .0387)](1{,}185/2) = 1.42$; since this is well below the 5% critical value, we fail to reject H_0. In other words, *motheduc* and *fatheduc* are jointly insignificant in the birth weight equation. Most statistical packages these days have built-in commands for testing multiple hypotheses after OLS estimation, and so one need not worry about making the mistake of running the two regressions on different data sets. Typically, the commands are applied after estimation of the unrestricted model, which means the smaller subset of data is used whenever there are missing values on some variables. Formulas for computing the F statistic using matrix algebra—see Appendix E—do not require estimation of the retricted model.