Notes: Lecture 2

NON LINEARITY

When we do linear regression, we assume that the relationship between the response variable (i.e. the dependent variable, DV) and the predictors (i.e. the independent variables, IVs) is linear. This is the assumption of linearity. If this assumption is violated, the linear regression will try to fit a straight line to data that do not follow a straight relationship. This can also be a cause of non-normally distributed residuals as well as heteroscedasticity (though not necessarily). Therefore, checking for linearity is important to make sure that the main assumptions of OLS are not violated.

Moreover, by being linear OLS is also an additive technique: we assume that the effect of an IV on the DV is the same for all values of the other IVs in the model, once they are fixed at a given value. In other words, we assume that the slope of the population regression function is constant, so that the effect on Y of a unit change in X does not depend on the value of one or more IV. In this sense, non-linearity is important also for the interpretation you can give to the results of your model.

The regression function can be nonlinear in two different cases:

1) The effect on Y of a change in X1 depends on the value of X1 itself.

For example, reducing class sizes by one student per teacher might have a greater effect on the math achievement of students if class sizes are already small than if class sizes are so large that the teacher can do little more than keeping the class under control. If so, then the score obtained in the math test (Y) is a nonlinear function of the student-teacher ratio (X1), where this function is steeper when X1 is small (e.g. logarithm function vs. just a straight line).

2) The effect on Y of a change in X1 depends on the value of another IV, say X2.

For example, the effect on test scores of reducing the student-teacher ratio will be greater in schools where the parents of students have a higher level of education than in schools where the parents of students have a lower level of education (graphically, the slope of the relationship between X1 and Y depends on X2, and it will be more or less steep according to X2).

In both cases, the population regression function is a nonlinear function of the IVs – that is, the conditional expectation of Y is a nonlinear function of one or more of the X's. Although they are nonlinear in the X's, these models are linear functions of the unknown coefficients (or parameters) of the population regression model that, taken together, are able to express the nonlinear relationship between Y and X's. Therefore, these coefficients can be estimated and tested using OLS.

Note that the term "nonlinear regression" applies to two different families of models. In the first family, the regression function is a nonlinear function of the X's but is a linear function of the unknown parameters (the Betas). In the second family, the regression function is a nonlinear function of the unknown parameters and may or may not be a nonlinear function of the X's (this is the case of logit, probit, etc.).

1) The slope of X depends on the value of X: Quadratic model

use “\...caschool.dta”

corr testscr avginc

twoway (scatter testscr avginc) (lfit testscr avginc)

reg testscr avginc

predict r, resid

kdensity r, norm

rvfplot, yline(0)

avplots

avplots, mlabel(county)

lvr2plot

lvr2plot, mlabel(county)

estat imtest, white → this test detects non-linear heteroscedasticity

estat hettest → this test detects linear heteroscedasticity

acprplot avginc, lowess lsopts(bwidth(1))

twoway (scatter testscr avginc) (lfit testscr avginc) (qfit testscr avginc) (lowess testscr avginc)

The quadratic regression function seems to fit the data better than the linear one. Imagine to draw a curve that fits the points of the scatterplot. This curve would be steeper for low values of district income, then would be flatter as district income gets higher.

gen avginc2 = avginc^2

reg testscr avginc avginc2

predict rr, resid

kdensity rr, norm

The significant coefficient for income^2 formally rejects the hypothesis that the relationship between income and test scores is linear

rvfplot, yline(0)

avplots

lvr2plot

estat imtest, white

estat hettest

acprplot avginc2, lowess lsopts(bwidth(1))

*using margins

reg testscr c.avginc##c.avginc

margins, dydx(avginc) at (avginc = (10 40 46 55))

margins, at (avginc = (5 (1) 55))

marginsplot

2) linear-log model

An alternative to use a quadratic relationship is using the natural logarithm of X. This is sometimes called a linear-log model (given that the X is logged): Y=ln(X): the logarithmic function is steeper for small than for large values of X, it is only defined for X>0, it is positive for X>1 (equal to 0 when X=1) and has slope 1/X.

scatter testscr avginc

*to show the effect of using a log scale for avginc:

scatter testscr avginc, xscale(log)

gen lnavginc = log(avginc)

reg testscr lnavginc

A 1% change in X is associated with a change in Y of 0.01*Beta. In our case, then, a 1% increase in income is associated with an increase in test scores of 0.01*36.42=0.36 points.

di log(10)

2.3025851

di log(11)

2.3978953

di log(40)

3.6888795

di log(41)

3.7135721

lincom (_b[_cons]+ _b[lnavginc]*2.3978953)-(_b[_cons]+ _b[lnavginc]*2.3025851)

lincom (_b[_cons]+ _b[lnavginc]*3.7135721)-(_b[_cons]+ _b[lnavginc]*3.6888795)

3) Log-linear model

<u>use "\...\nations.dta"</u>

graph matrix gnpcap school2 school1, half

We want to explain the gnpcap using the school1 and school2 variables → the relationship seems non-linear (especially for school2)

reg gnpcap school2 school1

acprplot school2 , lowess

acprplot school1, lowess

Two options: A) add a quadratic term for school2 (and for school1); B) transform the dependent variable (which is skewed to the right)

kdensity gnpcap, normal

sktest gnpcap

Let us try to transform gnpcap to make it more normally distributed. Potential transformations include taking the log, the square root or raising the variable to a power. Selecting the appropriate transformation is somewhat of an art.

ladder gnpcap (look for the transformation with the smallest chi-square)

gladder gnpcap

generate lggnp=log(gnpcap)

label variable lggnp "log of gnpcap"

kdensity lggnp, normal

hist lggnp, normal

sktest lggnp

graph matrix lggnp school2 school1, half

regress lggnp school2 school1

In the log-linear model, a one-unit change in X (deltaX=1) is associated with a 100*Beta% change in Y. Translated into percentages, a unit change in X is associated with a 100*b1% change in Y.

acprplot school2 , lowess

acprplot school1, lowess

less deviation from nonlinearity than before

lincom (_b[_cons]+ _b[school2]*60)-(_b[_cons]+ _b[school2]*40)

lincom (_b[_cons]+ _b[school2]*40)-(_b[_cons]+ _b[school2]*20)

In this case the relationship between the dependent variable and the independent variables (after the transformation) is linear.

INTERACTION MODELS

1) Interactions with categorical variables

use “\...nes2004.dta”

If an interaction is taking place, then OLS will not capture this effect, unless we model it.

codebook dem_therm

codebook progovmnt

codebook polknow3

recode polknow3 (0=0 "Low") (1/2=1 "Med-high"), gen(poldummy)

Let’s explore the relationship between “dem_therm” and “progovmnt” (also at different levels of political knowledge).

tab progovmnt, sum(dem_therm) nost

tab poldummy progovmnt, sum(dem_therm) nost

*additive model (dem_therm = a + b1*progovmnt + b2*poldummy)

reg dem_therm progovmnt poldummy

predict yhat1

separate yhat1, by(poldummy)

line yhat10 yhat11 progovmnt, sort legend(cols(1)) ytitle("Predicted dem_therm")

*interaction model (dem_therm = a + b1*progovmnt + b2*poldummy + b3*interaction) – the coefficient b3 tells us how much to adjust our additive estimate for each one-unit increase in political knowledge.

gen interaction = progovmnt * poldummy

reg dem_therm progovmnt poldummy interaction

predict zhat1

separate zhat1, by(poldummy)

line zhat10 zhat11 progovmnt, sort legend(cols(1)) ytitle("Predicted dem_therm")

twoway (lfit dem_therm progovmnt if(poldummy==0)) (lfit dem_therm progovmnt if(poldummy==1)) , legend(order(1 "Low pol. knowledge" 2 "Medium-High pol. knowledge") cols(1))

When poldummy=0, the regression function is a+b1*progovmnt (intercept a and slope b1). When poldummy=1, the regression function is (a+b2) + (b1+b3)*progovmnt (intercept a+b2 and slope b1+b3). The difference between the two intercepts is b2 and the difference between the two slopes is b3.

*let's try with the original variable

tab polknow3 progovmnt, sum(dem_therm) nost

gen interaction2 = progovmnt * polknow3

reg dem_therm progovmnt polknow3 interaction2

twoway (lfit dem_therm progovmnt if(polknow3==0)) (lfit dem_therm progovmnt if(polknow3==1)) (lfit dem_therm progovmnt if(polknow3==2)), legend(order(1 "Low pol.know" 2 "Medium pol.know" 3 "High pol.know") cols(1))

2) Interactions with continuous variables or interval-level variables

Y = a + b1X1 + b2X2 + b3(X1*X2)

The interaction term allows the effect of a unit change in X1 to depend on the value of X2. Indeed, if X1 changes, we get: Y + ΔY = a + b1(X1+ΔX1) + b2X2 + b3((X1+ΔX1)*X2).

Thus the effect on Y of a change in X1, holding X2 constant, is: ΔY/ΔX1=b1+b3X2, which depends on X2

ΔY/ΔX1 is the marginal effect of X1.

The marginal effect of X2 is ΔY/ΔX2 = b2+b3X1.

Putting the two effects together shows that the coefficient of b3 on the interaction term is the effect of a unit increase in X1 and X2 above and beyond the sum of the effects of a unit increase in X1 alone and a unit increase in X2 alone (this is true whether X1 and/or X2 are continuous or binary). That is, if X1 changes by ΔX1 and X2 changes by ΔX2, then the expected change in Y is:

ΔY = (b1+b3X2)*ΔX1 + (b2+b3X1)*ΔX2 + b3*ΔX1*ΔX2.

The first term is the effect from changing X1 holding X2 constant; the second term is the effect from changing X2 holding X1 constant; and the final term is the extra effect from changing both X1 and X2.

reg dem_therm income_hh kerry_therm

gen interaction3 = kerry_therm * income_hh

reg dem_therm income_hh kerry_therm interaction3

twoway (lfit dem_therm income_hh if (kerry_therm ==0)) (lfit dem_therm income_hh if (kerry_therm ==50)) (lfit dem_therm income_hh if (kerry_therm ==100)), legend(order(1 "Conservative" 2 "Indipendent" 3 "Liberal")) ytitle("") title(Conditional impact of income as kerry_therm changes)

*replicate using union_therm instead of kerry_therm

MARGINS

The margins command is a postestimation command that reports margins of responses and marginal effects (margins of derivatives of responses)

1) Margin of responses

codebook effparty45 numiss

regress effparty45 numiss

lincom _b[_cons] + _b[numiss]

margins, at(numiss=1)

marginsplot

margins, at (numiss=(0 1))

marginsplot

reg ecogr709 const45 federal45 judrev45

lincom _b[_cons] +_b[const45]*2+_b[federal45]*1+_b[judrev45]*3

margins, at(const45=2 federal45=1 judrev45=3)

marginsplot

margins, at(const45=2 federal45=5 judrev45=3)

2) Test of margins

test whether the difference between the expected values (margins) for different combination of our IVs is significant (test of equality of margins.)

regress effparty45 numiss

margins, at(numiss=(1 3))

margins, at(numiss=(1 3)) contrast(atcontrast(r._at) wald)

margins, at(numiss=(1 3)) contrast(atjoint)

marginsplot

marginsplot, yscale(range(-1 4)) ylabel(-1(.5)4) yline(0)

reg ecogr709 const45 federal45 judrev45

margins, at(const45=2 federal45=1 judrev45=3) at(const45=2 federal45=3 judrev45=3) contrast(atcontrast(r._at) wald)

margins, at(federal45=(1 3)) contrast(atcontrast(r._at) wald)

marginsplot

3) Marginal effects

Obtaining margins of derivatives of responses (marginal effects) is crucial for better understanding quadratic models and interactions.

reg dem_therm c.progovmnt##i.poldummy

margins, at(progovmnt==2 poldummy==1)

margins, dydx(progovmnt)at(poldummy=(0 1))

marginsplot, yline(0)

margins, dydx(poldummy) at(progovmnt=(0 (1) 3))

marginsplot, yline(0)

margins, dydx(poldummy) at(progovmnt=(0 (1) 3))

marginsplot, yline(0) addplot (hist progovmnt, percent yaxis(2) yscale(alt axis(2)) below color(green) fcolor(none))

***INTERACTION MODELS: GOLDEN RULES**

1. Identify a possible nonlinear relationship (from the theory!) – that is, analysts should use interaction models whenever the hypothesis they want to test is conditional in nature.

2. Specify a nonlinear function and estimate its parameters by OLS. It can be a quadratic or an interaction term. Scholars should include all constitutive terms in their interaction model specifications.

3. Determine whether the nonlinear model improves upon a linear model.

4. Plot the estimated nonlinear regression function.

5. Estimate the effect on Y of a unit change in X. That is, scholars should not interpret constitutive terms as if they are unconditional marginal effects.

6. Analysts should calculate substantively meaningful marginal effects and standard errors. This is relevant. If the interaction term is significant and the two constitutive terms of such interaction term are not significant, this is not a problem! That would mean that the impact of X on Y could be significant only for some values of Z for example.

7. make sure the product term is significant (to avoid cases in which the marginal effect of X varies only trivially with Z); this piece of information is not usually included in published marginal effect plots but is critical for determining whether there is empirical evidence of interaction between X and Z

8. superimpose the distribution of the Z variable

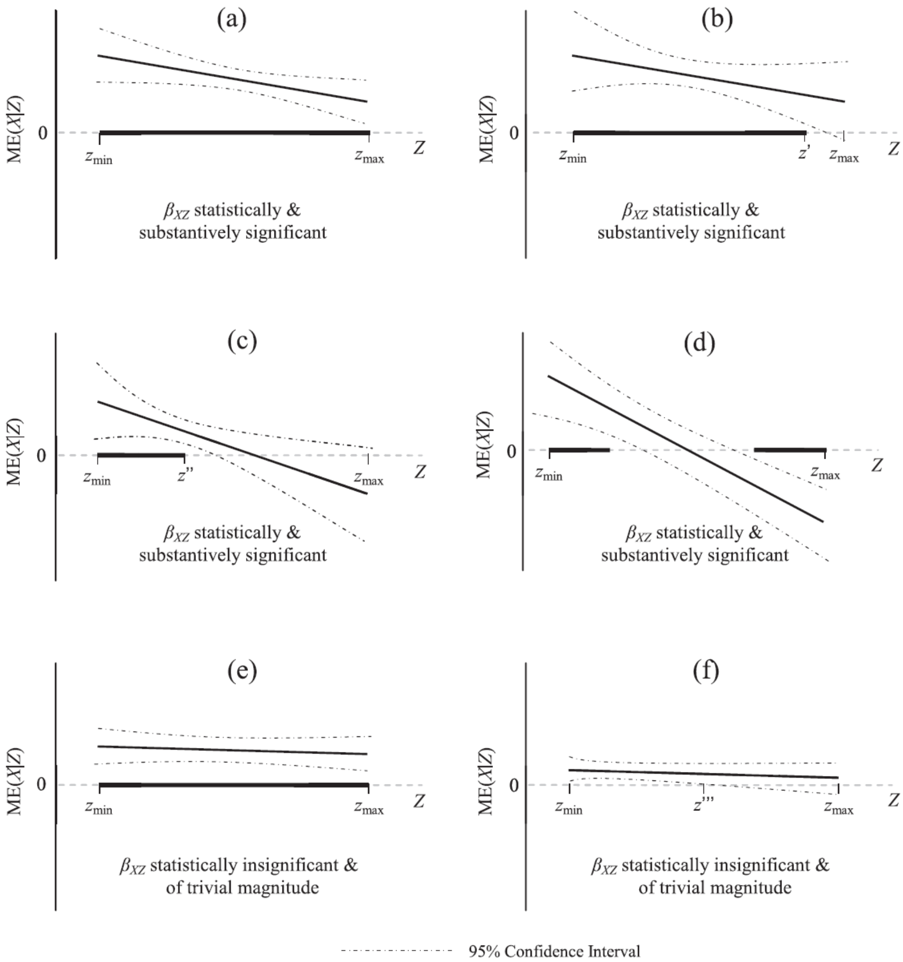
9. check the reverse interaction: Z conditional on X (not only X conditional on Z)

REFERENCES

Brambor, T., Clark, W. R. and M. Golder. 2006. Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis* 14: 63-82.

Berry, w., Golder, M. and Milton, D. 2012. Improving Tests of Theories Positing Interaction. *Journal of Politics* 74: 653-671.

Figure 3 Plots of ME(X|Z) Reflecting Several Prototypical Sets of Empirical Results



Note: The horizontal axis is bold for all values of Z at which the marginal effect of X on Y is statistically and substantively significant. z_{min} and z_{max} indicate the lowest and highest observed values of Z.