Multiple regression

The Population Multiple Regression Model

Consider the case of two regressors:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \quad i = 1,\ldots,n$$

- Y is the *dependent variable*
- X_1, X_2 are the two *independent variables* (*regressors*)
- (Y_i, X_{1i}, X_{2i}) denote the i^{th} observation on Y, X_1, and X_2.
- β_0 = unknown population intercept
- β_1 = effect on Y of a change in X_1, holding X_2 constant
- β_2 = effect on Y of a change in X_2, holding X_1 constant
- u_i = the regression error (omitted factors)

Interpretation of coefficients in multiple regression

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \;\; i = 1,\ldots,n$$

Consider changing X_1 by ΔX_1 while holding X_2 constant:

Population regression line ***before*** the change:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Population regression line, ***after*** the change:

$$Y + \Delta Y = \beta_0 + \beta_1 (X_1 + \Delta X_1) + \beta_2 X_2$$

Before: $$Y = \beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2 X_2$$

After: $$Y + \Delta Y = \beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2 X_2$$

Difference: $$\Delta Y = \beta_1 \Delta X_1$$

So:

$$\beta_1 = \frac{\Delta Y}{\Delta X_1}, \textbf{ holding } X_2 \textbf{ constant}$$

$$\beta_2 = \frac{\Delta Y}{\Delta X_2}, \textbf{ holding } X_1 \textbf{ constant}$$

β_0 = predicted value of Y when $X_1 = X_2 = 0$.

The OLS Estimator in Multiple Regression

With two regressors, the OLS estimator solves:

$$\min_{b_0, b_1, b_2} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_{1i} + b_2 X_{2i})]^2$$

- The OLS estimator minimizes the average squared difference between the actual values of Y_i and the prediction (predicted value) based on the estimated line.
- This minimization problem is solved using calculus
- **This yields the OLS estimators of β_0 and β_1.**

Example: the California test score data

Regression of HLR against NTRAD:

$$\widehat{HLR} = 0.117 + 0.246\text{*NTRAD}$$

Now include the size of average trade (ATS):

$$\widehat{HLR} = 0.608 + 0.442\text{ *NTRAD} - 0.077\text{*ATS}$$

- What happens to the coefficient on NTRAD?
- Why? (*Note*: corr(*STR*, *PctEL*) = +++)

Multiple regression in STATA

```
. reg highlow numtrad  ats if month==1, r

Linear regression                                      Number of obs =     426
                                                       F(  2,   423) =   67.06
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3797
                                                       Root MSE      =  .21106

------------------------------------------------------------------------------
             |               Robust
     highlow |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     numtrad |   .4422475   .0473049     9.35   0.000     .3492657    .5352294
         ats |  -.0770767   .0108854    -7.08   0.000    -.0984729   -.0556805
       _cons |   .6085741   .0768587     7.92   0.000     .4575016    .7596466
------------------------------------------------------------------------------
```

More on this printout later…

Measures of Fit for Multiple Regression

Actual = predicted + residual: $Y_i = \hat{Y}_i + \hat{u}_i$

SER = std. deviation of $\hat{u}_i$ (with d.f. correction)

$RMSE$ = std. deviation of $\hat{u}_i$ (without d.f. correction)

R^2 = fraction of variance of Y explained by X

$\bar{R}^2$ = "adjusted R^2" = R^2 with a degrees-of-freedom correction that adjusts for estimation uncertainty; $\bar{R}^2 < R^2$

SER and RMSE

As in regression with a single regressor, the *SER* and the *RMSE* are measures of the spread of the *Y*'s around the regression line:

$$SER = \sqrt{\frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2}$$

R_2 and $\bar{R}^2$

The R^2 is the fraction of the variance explained – same definition as in regression with a single regressor:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS},$$

where $ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})^2$, $SSR = \sum_{i=1}^{n}\hat{u}_i^2$, $TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$.

- The R^2 always increases when you add another regressor (*why?*) – a bit of a problem for a measure of "fit"

R^2 and $\bar{R}^2$, ctd.

The $\bar{R}^2$ (the "adjusted R^2") corrects this problem by "penalizing" you for including another regressor – the $\bar{R}^2$ does not necessarily increase when you add another regressor.

$$Adjusted\ R^2:\ \ \bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)\frac{SSR}{TSS}$$

Note that $\bar{R}^2 < R^2$, however if n is large the two will be very close.

Measures of fit, ctd.

(1) $\widehat{HLR} = 0.117 + 0.426 * \text{NTRAD},$

$$R^2 = .2585$$

(2) $\widehat{HLR} = 0.608 + 0.442 * \text{NTRAD} - 0.077 * \text{ATS},$

$$R^2 = .3797,\ \bar{R}^2 = .3768$$

- *What – precisely – does this tell you about the fit of regression (2) compared with regression (1)?*
- *Why are the R^2 and the $\bar{R}^2$ so close in (2)?*

Hypothesis Tests and Confidence Intervals

Object of interest: β_1 in,

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i \;\; u_i,\; i = 1,\ldots, n$$

$\beta_1 = \Delta Y/\Delta X$, for an autonomous change in X (***causal effect***)

The Least Squares Assumptions:

1. $E(u|X = x) = 0$.
2. (X_i, Y_i), $i = 1,\ldots,n$, are i.i.d.
3. Large outliers are rare ($E(X^4) < \infty$, $E(Y^4) < \infty$.

The Sampling Distribution of $\hat{\beta}_1$:

Under the LSA's, for n large, $\hat{\beta}_1$ is approximately distributed,

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_v^2}{n\sigma_X^4}\right), \text{ where } v_i = (X_i - \mu_X)u_i$$

Hypothesis Testing and the Standard Error of $\hat{\beta}_1$

The objective is to test a hypothesis, like $\beta_1 = 0$, using data – to reach a tentative conclusion whether the (null) hypothesis is correct or incorrect.

General setup

Null hypothesis and **two-sided** alternative:

$$H_0: \ \beta_1 = \beta_{1,0} \ \text{ vs. } H_1: \ \beta_1 \neq \beta_{1,0}$$

where $\beta_{1,0}$ is the hypothesized value under the null.

Null hypothesis and **one-sided** alternative:

$$H_0: \ \beta_1 = \beta_{1,0} \ \text{ vs. } H_1: \ \beta_1 < \beta_{1,0}$$

General approach: construct t-statistic, and compute p-value (or compare to $N(0,1)$ critical value)

- In general: $$t = \frac{\text{estimate - hypothesized value}}{\text{standard error of the estimate}}$$

  where the *SE* of the estimate is the square root of the variance of the estimate.

- ***For testing the mean of Y***: $$t = \frac{\bar{Y} - \mu_{Y,0}}{s_Y / \sqrt{n}}$$

- ***For testing β₁***, $$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)},$$

  where $SE(\hat{\beta}_1)$ = the square root of an estimate of the variance of the sampling distribution of $\hat{\beta}_1$

Formula for SE($\hat{\beta}_1$)

Recall the expression for the variance of $\hat{\beta}_1$ (large n):

$$\text{var}(\hat{\beta}_1) = \frac{\text{var}[(X_i - \mu_x)u_i]}{n(\sigma_X^2)^2} = \frac{\sigma_v^2}{n\sigma_X^4}, \text{ where } v_i = (X_i - \mu_X)u_i.$$

The estimator of the variance of $\hat{\beta}_1$ replaces the unknown population values of σ_v^2 and σ_X^4 by estimators constructed from the data:

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\text{estimator of } \sigma_v^2}{(\text{estimator of } \sigma_X^2)^2} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{v}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}$$

where $\hat{v}_i = (X_i - \bar{X})\hat{u}_i$.

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{v}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}, \text{ where } \hat{v}_i = (X_i - \bar{X})\hat{u}_i.$$

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} = \text{the standard error of } \hat{\beta}_1$$

OK, this is a bit nasty, but:

- It is less complicated than it seems. The numerator estimates var(v), the denominator estimates var(X).
- Why the degrees-of-freedom adjustment $n - 2$? Because two coefficients have been estimated (β_0 and β_1).
- $SE(\hat{\beta}_1)$ is computed by regression software

Summary: To test H_0: $\beta_1 = \beta_{1,0}$ v. H_1: $\beta_1 \neq \beta_{1,0}$,

- Construct the t-statistic

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}}$$

- Reject at 5% significance level if $|t| > 1.96$
- The p-value is $p = \Pr[|t| > |t^{act}|]$ = probability in tails of normal outside $|t^{act}|$; you reject at the 5% significance level if the p-value is < 5%.
- This procedure relies on the large-n approximation; typically $n = 50$ is large enough for the approximation to be excellent.

Example: Stock volatility and number of trades data

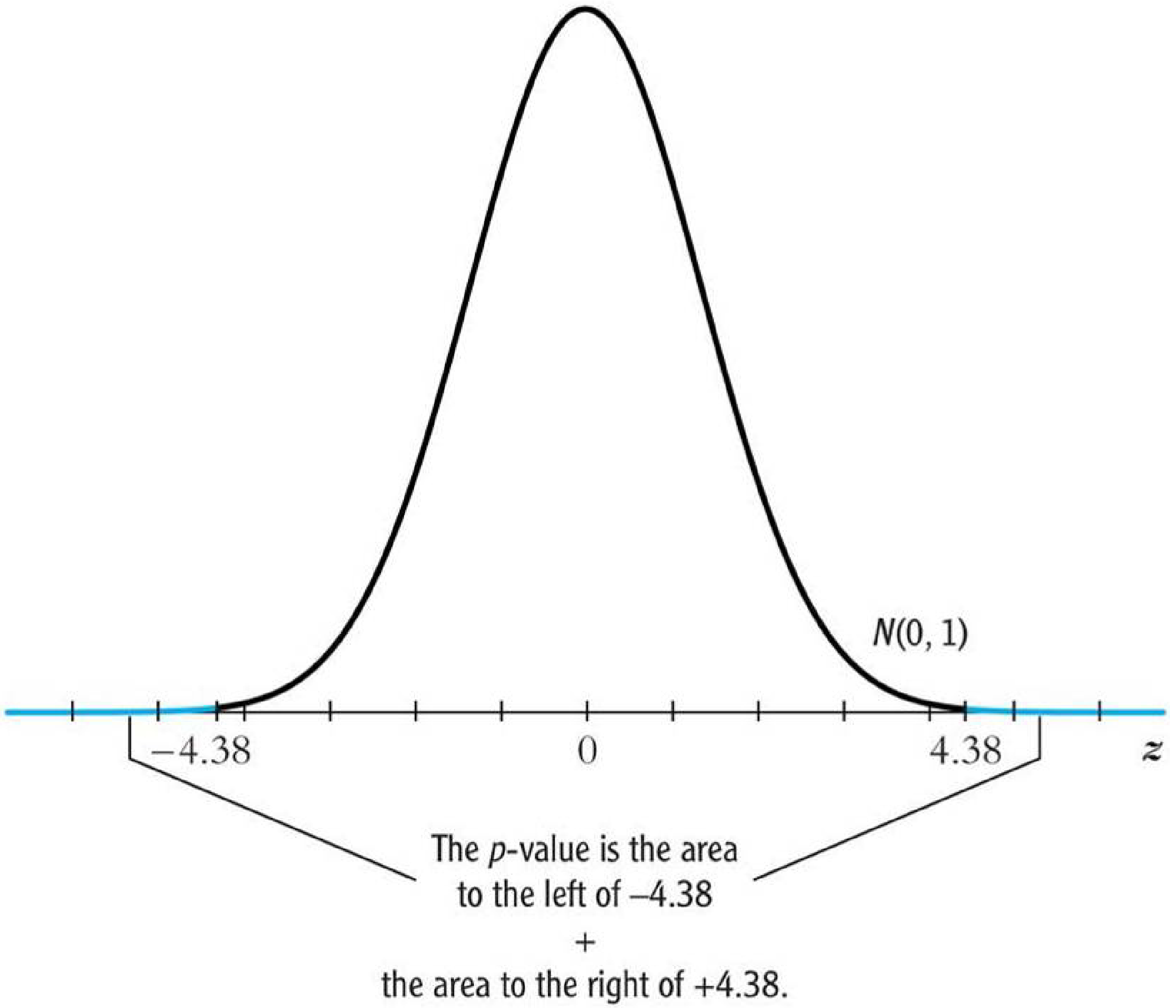
Estimated regression line: $\widehat{HLR} = 0.117 + 0.426 * \text{NTRAD}$

Regression software reports the standard errors:

$$SE(\hat{\beta}_0) = 0.0119 \qquad SE(\hat{\beta}_1) = 0.035$$

t-statistic testing $\beta_{1,0} = 0 = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} = \dfrac{0.426 - 0}{0.035} = 12.17$

- The 1% 2-sided significance level is 2.58, so we reject the null at the 1% significance level.
- Alternatively, we can compute the p-value…



The p-value based on the large-n standard normal approximation to the t-statistic is 0.00001 (10^{-5})

Confidence Intervals for β_1

Recall that a 95% confidence is, equivalently:

- The set of points that cannot be rejected at the 5% significance level;
- A set-valued function of the data (an interval that is a function of the data) that contains the true parameter value 95% of the time in repeated samples.

Because the *t*-statistic for β_1 is N(0,1) in large samples, construction of a 95% confidence for β_1 is just like the case of the sample mean:

$$95\% \text{ confidence interval for } \beta_1 = \{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\}$$

Confidence interval example: HLR and NTRAD

Estimated regression line: $\widehat{HLR} = 0.117 + 0.426*\text{NTRAD}$

$$SE(\hat{\beta}_0) = 0.0119 \qquad SE(\hat{\beta}_1) = 0.035$$

95% confidence interval for $\hat{\beta}_1$:

$$\{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\} = \{0.426 \pm 1.96 \times 0.035\}$$

$$= (0.357\ ,\ 0.495)$$

The following two statements are equivalent *(why?)*

- The 95% confidence interval does not include zero;
- The hypothesis $\beta_1 = 0$ is rejected at the 5% level

A concise (and conventional) way to report regressions:

Put standard errors in parentheses below the estimated coefficients to which they apply.

$$\widehat{HLR} = 0.117 + 0.426 * \text{NTRAD},\ R^2 = .2585,\ SER = 0.2299$$
$$(0.011) \qquad (0.035)$$

<u>This expression gives a lot of information</u>

- The estimated regression line is

$$\widehat{HLR} = 0.117 + 0.426 * \text{NTRAD}$$

- The standard error of $\hat{\beta}_0$ is 0.011
- The standard error of $\hat{\beta}_1$ is 0.035
- The R^2 is .2585; the standard error of the regression is 0.2299

OLS regression: reading STATA output

```
. reg highlow  numtrad if month==1, r

Linear regression                                      Number of obs =     426
                                                       F(  1,   424) =   81.87
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2585
                                                       Root MSE      =  .23049
```

highlow	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
numtrad	.42659	.0471464	9.05	0.000	.3339202	.5192597
_cons	.1173006	.0105374	11.13	0.000	.0965886	.1380126

so:

$t\ (\beta_1 = 0) =$ **9.05**, p-value = **0.000** (2-sided)

95% 2-sided conf. interval for β_1 is **(0.333, 0.519)**

Testing Exclusion Restrictions

- Now the null hypothesis might be something like H_0: $\beta_{k-q+1} = 0, \ldots, \beta_k = 0$
- The alternative is just H_1: H_0 is not true
- Can't just check each t statistic separately, because we want to know if the q parameters are jointly significant at a given level – it is possible for none to be individually significant at that level

Exclusion Restrictions (cont)

- To do the test we need to estimate the "restricted model" without $x_{k-q+1}, \ldots, x_k$ included, as well as the "unrestricted model" with all x's included
- Intuitively, we want to know if the change in SSR is big enough to warrant inclusion of $x_{k-q+1}, \ldots, x_k$

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}, \text{ where}$$

r is restricted and ur is unrestricted

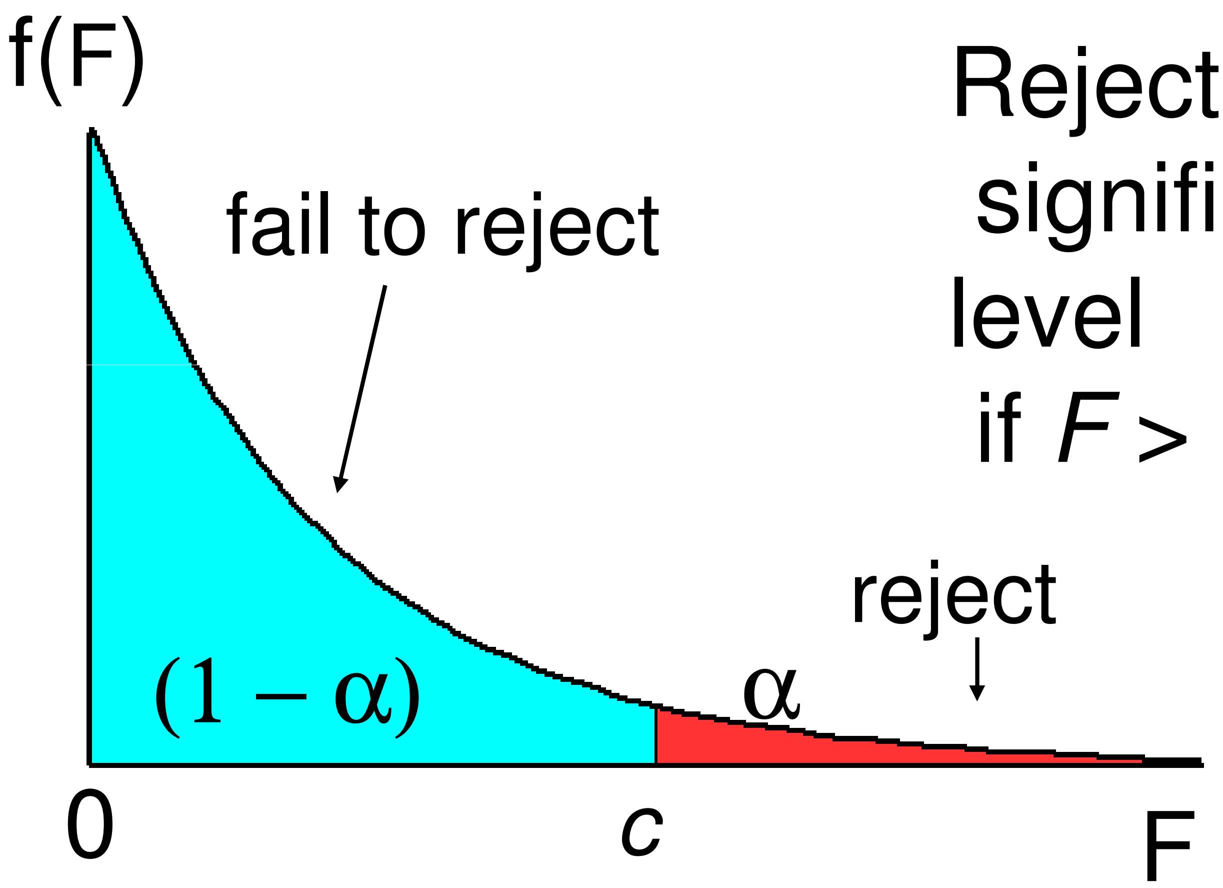
The F statistic

- The F statistic is always positive, since the SSR from the restricted model can't be less than the SSR from the unrestricted
- Essentially the F statistic is measuring the relative increase in SSR when moving from the unrestricted to restricted model
- q = number of restrictions, or $df_r - df_{ur}$
- $n - k - 1 = df_{ur}$

The *F* statistic (cont)

- To decide if the increase in SSR when we move to a restricted model is "big enough" to reject the exclusions, we need to know about the sampling distribution of our F stat
- Not surprisingly, $F \sim F_{q,n-k-1}$, where q is referred to as the numerator degrees of freedom and $n - k - 1$ as the denominator degrees of freedom

The *F* statistic (cont)



Reject H_0 at α significance level if $F > c$

The R^2 form of the F statistic

- Because the SSR's may be large and unwieldy, an alternative form of the formula is useful
- We use the fact that $SSR = SST(1 - R^2)$ for any regression, so can substitute in for SSR_u and SSR_{ur}

$$F \equiv \frac{\left(R_{ur}^2 - R_r^2\right)/q}{\left(1 - R_{ur}^2\right)/(n-k-1)}, \text{ where again}$$

r is restricted and ur is unrestricted

Overall Significance

- A special case of exclusion restrictions is to test H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$
- Since the R^2 from a model with only an intercept will be zero, the *F* statistic is simply

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

Nonlinear Regression Functions

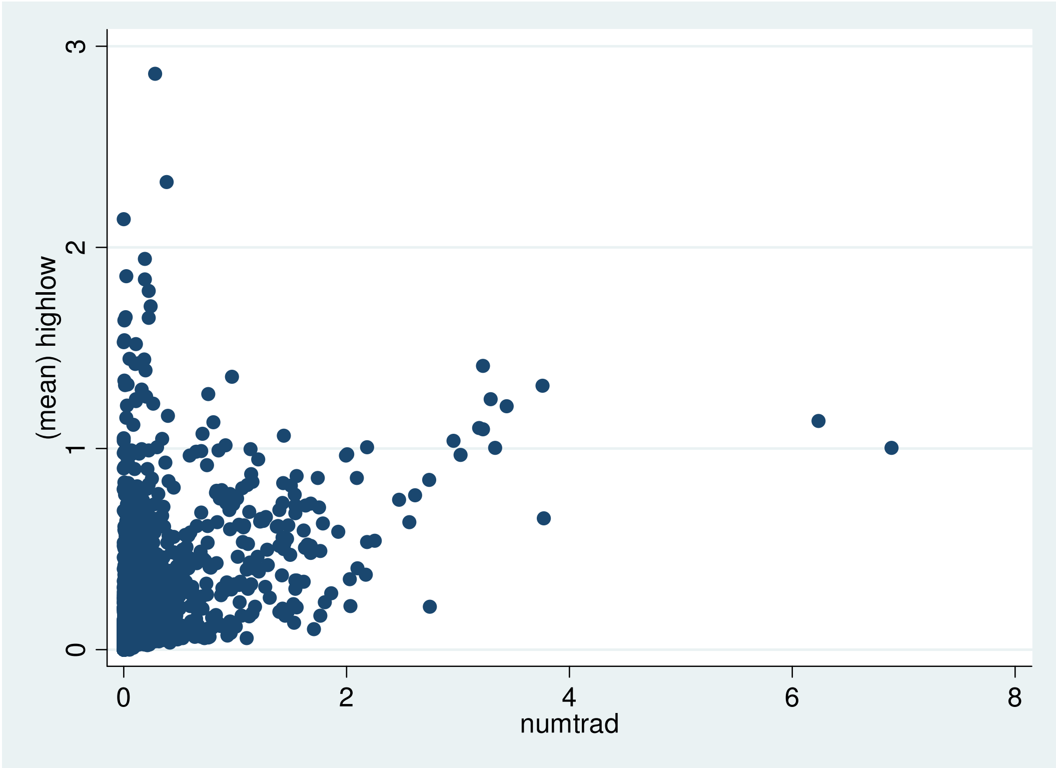
Nonlinear Regression Functions

- Everything so far has been linear in the *X*'s
- But the linear approximation is not always a good one
- The multiple regression framework can be extended to handle regression functions that are nonlinear in one or more *X*.

<u>Outline</u>

1. Nonlinear regression functions – general comments
2. Nonlinear functions of one variable
3. Nonlinear functions of two variables: interactions

Does the volatility – trades relation look linear (probably not)…



Nonlinear Regression Population Regression Functions – General Ideas

If a relation between Y and X is **nonlinear**:

- The effect on Y of a change in X depends on the value of X – that is, the marginal effect of X is not constant
- A linear regression is mis-specified – the functional form is wrong
- The estimator of the effect on Y of X is biased – it needn't even be right on average.
- The solution to this is to estimate a regression function that is nonlinear in X

The general nonlinear population regression function

$$Y_i = f(X_{1i}, X_{2i}, \ldots, X_{ki}) + u_i,\ i = 1, \ldots, n$$

Assumptions

1. $E(u_i | X_{1i}, X_{2i}, \ldots, X_{ki}) = 0$ (same); implies that f is the conditional expectation of Y given the X's.
2. $(X_{1i}, \ldots, X_{ki}, Y_i)$ are i.i.d. (same).
3. Big outliers are rare (same idea; the precise mathematical condition depends on the specific f).
4. No perfect multicollinearity (same idea; the precise statement depends on the specific f).

The Expected Effect on Y of a Change in X_1 in the Nonlinear Regression Model (8.3)

The expected change in Y, ΔY, associated with the change in X_1, ΔX_1, holding $X_2, \ldots, X_k$ constant, is the difference between the value of the population regression function before and after changing X_1, holding $X_2, \ldots, X_k$ constant. That is, the expected change in Y is the difference:

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \ldots, X_k) - f(X_1, X_2, \ldots, X_k). \qquad (8.4)$$

The estimator of this unknown population difference is the difference between the predicted values for these two cases. Let $\hat{f}(X_1, X_2, \ldots, X_k)$ be the predicted value of Y based on the estimator $\hat{f}$ of the population regression function. Then the predicted change in Y is

$$\Delta \hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, \ldots, X_k) - \hat{f}(X_1, X_2, \ldots, X_k). \qquad (8.5)$$

Nonlinear Functions of a Single Independent Variable

We'll look at two complementary approaches:

1. Polynomials in *X*

   The population regression function is approximated by a quadratic, cubic, or higher-degree polynomial

2. Logarithmic transformations
   - *Y* and/or *X* is transformed by taking its logarithm
   - this gives a "percentages" interpretation that makes sense in many applications

1. Polynomials in *X*

Approximate the population regression function by a polynomial:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \ldots + \beta_r X_i^r + u_i$$

- This is just the linear multiple regression model – except that the regressors are powers of *X*!
- Estimation, hypothesis testing, etc. proceeds as in the multiple regression model using OLS
- The coefficients are difficult to interpret, but the regression function itself is interpretable

Quadratic specification:

$$HLR_i = \beta_0 + \beta_1 * NTRAD_i + \beta_2(\mathrm{NTRAD}_i)^2 + u_i$$

Cubic specification:

$$HLR_i = \beta_0 + \beta_1 * NTRAD_i + \beta_2(\mathrm{NTRAD}_i)^2 + \beta_3(\mathrm{NTRAD}_i)^3 + u_i$$

Estimation of a cubic specification in STATA

```
. reg highlow numtrad  numtradessq numtradescub if month==1, r

Linear regression                                      Number of obs =     426
                                                       F(  2,   422) =       .
                                                       Prob > F      =       .
                                                       R-squared     =  0.3289
                                                       Root MSE      =  .21979

------------------------------------------------------------------------------
             |               Robust
     highlow |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     numtrad |   1.289911   .1977067     6.52   0.000     .9012987    1.678524
  numtradessq|  -9.27e-07   2.16e-07    -4.29   0.000    -1.35e-06   -5.03e-07
numtradescub |   2.07e-10   5.19e-11     3.98   0.000     1.05e-10    3.09e-10
       _cons |    .074214   .0100204     7.41   0.000      .054518      .09391
------------------------------------------------------------------------------
```

Summary: polynomial regression functions

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \ldots + \beta_r X_i^r + u_i$$

- Estimation: by OLS after defining new regressors
- Coefficients have complicated interpretations
- To interpret the estimated regression function:
  - plot predicted values as a function of *x*
  - compute predicted $\Delta Y/\Delta X$ at different values of *x*
- Hypotheses concerning degree *r* can be tested by *t*- and *F*-tests on the appropriate (blocks of) variable(s).
- Choice of degree *r*
  - plot the data; *t*- and *F*-tests, check sensitivity of estimated effects; judgment.
  - *Or use model selection criteria* (*later*)

2. Logarithmic functions of *Y* and/or *X*

- ln(*X*) = the natural logarithm of *X*
- Logarithmic transforms permit modeling relations in "percentage" terms (like elasticities), rather than linearly.

Here's why: $\ln(x+\Delta x) - \ln(x) = \ln\left(1+\frac{\Delta x}{x}\right) \cong \frac{\Delta x}{x}$

(calculus: $\frac{d\ln(x)}{dx} = \frac{1}{x}$)

Numerically:

$\ln(1.01) = .00995 \cong .01;$

$\ln(1.10) = .0953 \cong .10$ (sort of)

The three log regression specifications:

Case	Population regression function
I. linear-log	$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$
II. log-linear	$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$
III. log-log	$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$

- The interpretation of the slope coefficient differs in each case.
- The interpretation is found by applying the general "before and after" rule: "figure out the change in Y for a given change in X."

I. Linear-log population regression function

$$Y = \beta_0 + \beta_1 \ln(X) \qquad \text{(b)}$$

Now change X: $$Y + \Delta Y = \beta_0 + \beta_1 \ln(X + \Delta X) \qquad \text{(a)}$$

Subtract (a) – (b): $$\Delta Y = \beta_1 [\ln(X + \Delta X) - \ln(X)]$$

now $$\ln(X + \Delta X) - \ln(X) \cong \frac{\Delta X}{X},$$

so $$\Delta Y \cong \beta_1 \frac{\Delta X}{X}$$

or $$\beta_1 \cong \frac{\Delta Y}{\Delta X / X} \quad \text{(small } \Delta X\text{)}$$

Linear-log case, continued

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

for small ΔX,

$$\beta_1 \cong \frac{\Delta Y}{\Delta X / X}$$

Now $100 \times \frac{\Delta X}{X}$ = percentage change in X, so ***a 1% increase in X (multiplying X by 1.01) is associated with a .01β_1 change in Y.***

(1% increase in $X \Rightarrow$ .01 increase in $\ln(X)$

$\Rightarrow .01\beta_1$ increase in Y)

Example: HLR vs. ln(NTRAD)

- First defining the new regressor, LNTRAD=ln(NTRAD+1)
- The model is now linear in ln(NTRAD), so the linear-log model can be estimated by OLS:

$$\widehat{HLR} = \underset{(0.010)}{0.097} + \underset{(0.083)}{0.779}*\ln(NTRAD_i)$$

so a 1% increase in NTRAD is associated with an increase in HLR of 0.77 points in volatility

- Standard errors, confidence intervals, R^2 – all the usual tools of regression apply here.
- How does this compare to the cubic model?

II. Log-linear population regression function

$$\ln(Y) = \beta_0 + \beta_1 X \qquad \text{(b)}$$

Now change X: $\ln(Y + \Delta Y) = \beta_0 + \beta_1(X + \Delta X) \qquad \text{(a)}$

Subtract (a) – (b): $\ln(Y + \Delta Y) - \ln(Y) = \beta_1 \Delta X$

so $$\frac{\Delta Y}{Y} \cong \beta_1 \Delta X$$

or $$\beta_1 \cong \frac{\Delta Y / Y}{\Delta X} \text{ (small } \Delta X\text{)}$$

Log-linear case, continued

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

for small ΔX, $\quad \beta_1 \cong \dfrac{\Delta Y / Y}{\Delta X}$

- Now $100 \times \dfrac{\Delta Y}{Y}$ = percentage change in Y, so ***a change in X by one unit ($\Delta X = 1$) is associated with a $100\beta_1$% change in Y.***
- 1 unit increase in $X \Rightarrow \beta_1$ increase in $\ln(Y)$

  $\Rightarrow 100\beta_1$% increase in Y
- *Note*: What are the units of u_i and the SER?
  - fractional (proportional) deviations
  - for example, $SER = .2$ means…

III. Log-log population regression function

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i \qquad \text{(b)}$$

Now change X: $$\ln(Y + \Delta Y) = \beta_0 + \beta_1 \ln(X + \Delta X) \qquad \text{(a)}$$

Subtract: $$\ln(Y + \Delta Y) - \ln(Y) = \beta_1[\ln(X + \Delta X) - \ln(X)]$$

so $$\frac{\Delta Y}{Y} \cong \beta_1 \frac{\Delta X}{X}$$

or $$\beta_1 \cong \frac{\Delta Y / Y}{\Delta X / X} \text{ (small } \Delta X\text{)}$$

Log-log case, continued

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

for small ΔX,

$$\beta_1 \cong \frac{\Delta Y / Y}{\Delta X / X}$$

Now $100 \times \frac{\Delta Y}{Y}$ = percentage change in Y, and $100 \times \frac{\Delta X}{X}$ = percentage change in X, so ***a 1% change in X is associated with a β_1% change in Y***.

- ***In the log-log specification, β_1 has the interpretation of an elasticity.***

Example: ln(HLR) vs. ln(NTRAD)

- First defining a new dependent variable, ln(HLR+1), ***and*** the new regressor, ln(NTRAD)
- The model is now a linear regression of ln(HLR) against ln(NTRAD), which can be estimated by OLS:

$$\widehat{\ln(HLR)} = \underset{(0.007)}{0.086} + \underset{(0.052)}{0.550} \times \ln(\mathrm{NTRAD}_i)$$

An 1% increase in NTRAD is associated with an increase of .55% in HLR.

Summary: Logarithmic transformations

- Three cases, differing in whether Y and/or X is transformed by taking logarithms.
- The regression is linear in the new variable(s) $\ln(Y)$ and/or $\ln(X)$, and the coefficients can be estimated by OLS.
- Hypothesis tests and confidence intervals are now implemented and interpreted "as usual."
- The interpretation of β_1 differs from case to case.
- Choice of specification should be guided by judgment (which interpretation makes the most sense in your application?), tests, and plotting predicted values

Interactions Between Independent Variables

- Perhaps more trading has different impact on volatility depending on trade size
- That is, $\dfrac{\Delta(HLR)}{\Delta(NTRAD)}$ might depend on ATS
- More generally, $\dfrac{\Delta Y}{\Delta X_1}$ might depend on X_2
- How to model such "interactions" between X_1 and X_2?
- We first consider binary X's, then continuous X's

(a) Interactions between two binary variables

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i$$

- D_{1i}, D_{2i} are binary
- β_1 is the effect of changing D_1=0 to D_1=1. In this specification, *this effect doesn't depend on the value of* D_2.
- To allow the effect of changing D_1 to depend on D_2, include the "interaction term" $D_{1i} \times D_{2i}$ as a regressor:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i$$

Interpreting the coefficients

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3(D_{1i}\times D_{2i}) + u_i$$

General rule: compare the various cases

$$E(Y_i|D_{1i}=0, D_{2i}=d_2) = \beta_0 + \beta_2 d_2 \quad \text{(b)}$$

$$E(Y_i|D_{1i}=1, D_{2i}=d_2) = \beta_0 + \beta_1 + \beta_2 d_2 + \beta_3 d_2 \quad \text{(a)}$$

subtract (a) – (b):

$$E(Y_i|D_{1i}=1, D_{2i}=d_2) - E(Y_i|D_{1i}=0, D_{2i}=d_2) = \beta_1 + \beta_3 d_2$$

- The effect of D_1 depends on d_2 (what we wanted)
- β_3 = increment to the effect of D_1, when $D_2 = 1$

Let

$$\text{HTRADE} = \begin{cases} 1 \text{ if NTRAD} \geq 25 \\ 0 \text{ if NTRAD} < 25 \end{cases} \quad \text{and} \quad \text{HATS} = \begin{cases} 1 \text{ if ATS} \geq 6.4 \\ 0 \text{ if ATS} < 6.4 \end{cases}$$

$$\underset{}{\widehat{HLR}} = \underset{(0.016)}{0.120} + \underset{(0.033)}{0.199}\text{*HTRAD} - \underset{(0.016)}{0.108}\text{*HATS} - \underset{(0.045)}{0.016}\text{*(HTRAD*}\textit{HATS}\text{)}$$

- "Effect" of HTRAD when HATS = 0 is 0.199
- "Effect" of HTRAD when HATS = 1 is 0.199 – 0.016 =
- But note that this interaction isn't statistically significant: t = -0.016/0.045

(b) Interactions between continuous and binary variables

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + u_i$$

- D_i is binary, X is continuous
- As specified above, the effect on Y of X (holding constant D) = β_2, which does not depend on D
- To allow the effect of X to depend on D, include the "interaction term" $D_i \times X_i$ as a regressor:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + u_i$$

Binary-continuous interactions: the two regression lines

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + u_i$$

Observations with $D_i = 0$ (the "$D = 0$" group):

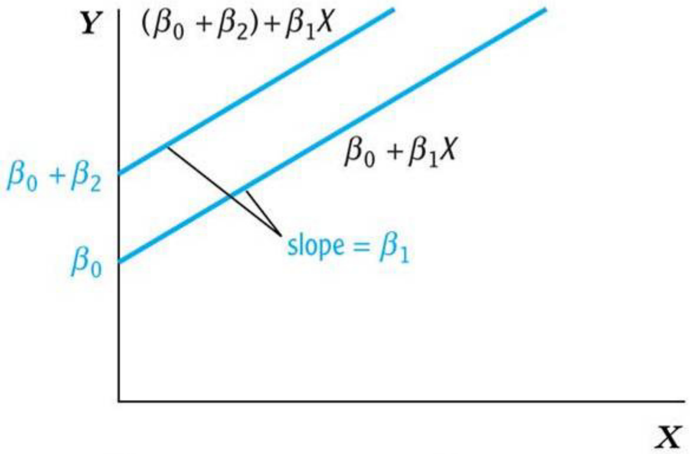
$$Y_i = \beta_0 + \beta_2 X_i + u_i$$ ***The D=0 regression line***

Observations with $D_i = 1$ (the "$D = 1$" group):

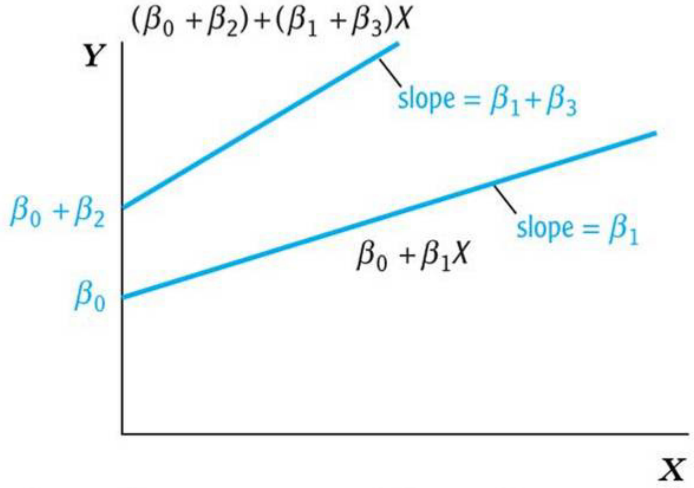
$$Y_i = \beta_0 + \beta_1 + \beta_2 X_i + \beta_3 X_i + u_i$$

$$= (\beta_0 + \beta_1) + (\beta_2 + \beta_3) X_i + u_i$$ ***The D=1 regression line***

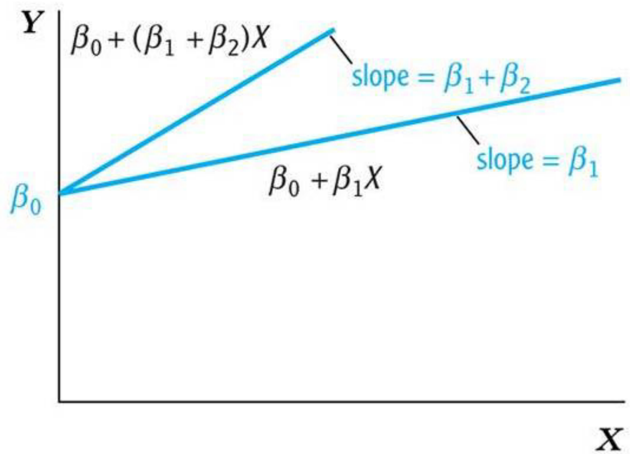
Binary-continuous interactions, ctd.



(a) Different intercepts, same slope



(b) Different intercepts, different slopes



(c) Same intercept, different slopes

Interpreting the coefficients

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + u_i$$

General rule: compare the various cases

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 (D \times X) \qquad \text{(b)}$$

Now change X:

$$Y + \Delta Y = \beta_0 + \beta_1 D + \beta_2 (X + \Delta X) + \beta_3 [D \times (X + \Delta X)] \qquad \text{(a)}$$

subtract (a) – (b):

$$\Delta Y = \beta_2 \Delta X + \beta_3 D \Delta X \text{ or } \frac{\Delta Y}{\Delta X} = \beta_2 + \beta_3 D$$

- The effect of X depends on D (what we wanted)
- β_3 = increment to the effect of X, when $D = 1$

(c) Interactions between two continuous variables

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

- X_1, X_2 are continuous
- As specified, the effect of X_1 doesn't depend on X_2
- As specified, the effect of X_2 doesn't depend on X_1
- To allow the effect of X_1 to depend on X_2, include the "interaction term" $X_{1i} \times X_{2i}$ as a regressor:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i$$

Interpreting the coefficients:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i$$

General rule: compare the various cases

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) \qquad \text{(b)}$$

Now change X_1:

$$Y + \Delta Y = \beta_0 + \beta_1 (X_1 + \Delta X_1) + \beta_2 X_2 + \beta_3 [(X_1 + \Delta X_1) \times X_2] \qquad \text{(a)}$$

subtract (a) – (b):

$$\Delta Y = \beta_1 \Delta X_1 + \beta_3 X_2 \Delta X_1 \ \text{ or } \ \frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2$$

- The effect of X_1 depends on X_2 (what we wanted)
- β_3 = increment to the effect of X_1 from a unit change in X_2

Summary: Nonlinear Regression Functions

- Using functions of the independent variables such as $\ln(X)$ or $X_1 \times X_2$, allows recasting a large family of nonlinear regression functions as multiple regression.
- Estimation and inference proceed in the same way as in the linear multiple regression model.
- Interpretation of the coefficients is model-specific, but the general rule is to compute effects by comparing different cases (different value of the original X's)
- Many nonlinear specifications are possible, so you must use judgment:
  - What nonlinear effect you want to analyze?
  - What makes sense in your application?