Practice: Basic Linear-Interactive Model

- Basic Linear-Interactive Model:

$$\text{eusup} = b_0 + b_{edu}\text{edu} + b_{lftrt}\text{lftrt} + b_{edlr}\text{edu}\times\text{lftrt} + \ldots + \varepsilon$$

- Effect of *edu*? $\dfrac{\partial \text{eusup}}{\partial \text{edu}} = b_{edu} + b_{edlr}\text{lftrt}$

  » For the record, the effect of *lftrt*: $\dfrac{\partial \text{eusup}}{\partial \text{lftrt}} = b_{lftrt} + b_{edlr}\text{edu}$

- Std Error of that Effect (of *edu* on *eusupp*)?

$$\hat{V}\left(\hat{b}_{edu} + \hat{b}_{edlr}\text{lftrt}\right) = \hat{V}\left(\hat{b}_{edu}\right) + \hat{V}\left(\hat{b}_{edlr}\right)\text{lftrt}^2 + 2\cdot\hat{C}\left(\hat{b}_{edu}, \hat{b}_{edlr}\right)\text{lftrt}$$

  – Std. Err. effect of *edu*: $\hat{V}\left(\hat{b}_{lftrt} + \hat{b}_{edlr}\text{edu}\right) = \hat{V}\left(\hat{b}_{lftrt}\right) + \hat{V}\left(\hat{b}_{edlr}\right)\text{edu}^2 + 2\cdot\hat{C}\left(\hat{b}_{edu}, \hat{b}_{edlr}\right)\text{edu}$

- To do this (e.g. in Stata, **set memory 32m**; load file)
  – First: gen edlr=education*leftright
  – Then: reg eu_supp ...education leftright edlr...

Examples & Practice:[1a]

Basic Linear-Interactive Model:

Education & Leftright Purely Micro-level Model

```
. reg eu_support education leftright edlr
      Source |       SS       df       MS              Number of obs =   42680
-------------+------------------------------           F(  3, 42676) =  371.33
       Model |  7869.51854     3  2623.17285            Prob > F      =  0.0000
    Residual |  301474.935 42676  7.06427347           R-squared     =  0.0254
-------------+------------------------------           Adj R-squared =  0.0254
       Total |  309344.453 42679  7.24816545           Root MSE      =  2.6579

------------------------------------------------------------------------------
  eu_support |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   education |   .5918654   .0302344    19.58   0.000     .5326053    .6511255
   leftright |    .080127   .0147551     5.43   0.000     .0512068    .1090472
        edlr |  -.0320928   .0054447    -5.89   0.000    -.0427646    -.021421
       _cons |   5.093097   .0831665    61.24   0.000     4.930089    5.256105
------------------------------------------------------------------------------
```

- What's effect of *edu*? Of *lftrt*? How moderate each other?

$$\frac{\partial \text{eusup}}{\partial \text{edu}} = .5919 - .0321(\text{lftrt}) \qquad \frac{\partial \text{eusup}}{\partial \text{lftrt}} = .0801 - .0321(\text{edu}) \qquad \frac{\partial^2 \text{eusup}}{\partial \text{edu} \partial \text{lftrt}} = b_{edlr} = -.0321$$

  - Only for modifying effect does standard regression output tell us directly.
- What are the standard errors of these effects?
  - Only for modifying effect does standard regression output tell us directly.
- Recall: *Main Effects* refer to beyond-range values; they not direct evidence on whether effect (generally) positive

Examples & Practice: Basic Linear-Interact Model[1b]

Education & Leftright Purely Micro-level Model

- What are the standard errors of these effects? Need this:

```
.vce
Covariance matrix of coefficients of regress model
        e(V) |  education   leftright        edlr       _cons
-------------+----------------------------------------------
   education |  .00091412
   leftright |  .00037584   .00021771
        edlr | -.00014825  -.00007456   .00002965
       _cons | -.00234221  -.00110293   .00037572   .00691666
```

- I prefer spreadsheet at this point:
  - Copy regression-estimation results; Paste into spreadsheet.
  - Ditto for estimated v-cov of estimated coefficients (as text or table)
  - Finally, you'll want summary stats for vars in your model (as text best):

```
. tabstat dgovpw psupgpw npgovpw PD NPPD NPGS NPPDGS, statistics( mean max min
median iqr skewness sd ) columns(variables)
    stats |    dgovpw   psupgpw   npgovpw        PD      NPPD      NPGS    NPPDGS
--------+----------------------------------------------------------------------
     mean |  25.24783  57.38261  1.965217  .6521739  1.369565  116.3004  765.9826
      max |      45.1       80.4       4.3         1       3.8    305.52    2181.2
      min |       11       41.1         1         0         0        49         0
      p50 |      26.6       57.2       1.8         1       1.6        96     916.2
      iqr |      13.7       10.8       1.7         1       2.2     83.47      1108
 skewness |  .1587892   .821165   .907747 -.6390097  .2729543  1.218438  .1633625
       sd |  9.655468  9.103879  .9599284  .4869848  1.213723  69.04308  644.2565
--------------------------------------------------------------------------------
```

Ex's & Practice: Basic Linear-Interact Model[1c]
Education & Leftright Purely Micro-level Model

- Spreadsht Formula to Plot Effect Lines w/ C.I.
  - ***Col 1:*** Conditioning Var (lftrt in *d*eusup/*d*edu)
    - *1st Cell (A29):* Enter min of range to consider (smpl min)
    - *2nd Cell:* =A29+1
      - Or sub "+(max-min)/(#steps)" for +1 (choose big# to smooth)
    - Copy down until reach max value you want plot to cover.
  - ***Column Two:*** Effect (*dGovDur/dNP* here)
    - *1st Cell (B29):* Enter =$B$3+$B$5*$A29, where:
      - $B$3 is absolute reference to cell containing coefficient on *edu*
      - $B$5 is absolute reference to cell containing coefficient on *edlr*
      - $A29 is reference to cell containing value of *lftrt* for that row
        - » $ optional, but helps if want copy whole block later for other effects
    - Copy Down

Ex's & Practice: Basic Linear-Interact Model: Education & Leftright Purely Micro-level Model

- Spreadsheet Formula to Plot Effect Lines w/ C.I.
  - ***Column Three:*** standard error (or can skip to bounds)
    - *1st Cell (A29):* =($B$11+$D$13*$A29^2+2*$B$13*$A29)^0.5
      - $B$11 is absolute reference to cell containing variance of *edu* coefficient
      - $D$13*$A29^2 is absolute reference to cell containing variance of *edlr* coefficient, times the square of the value of *lftrt* for that row.
      - 2*$B$13*$A29 is absolute reference to cell containing 2 times the covariance of *edu* and *edlr* coefficients times value of *lftrt* for that row.
      - ^0.5 turns that estimated variance into a standard error.
    - Copy down.
  - ***Column Four:*** Lower bound of 95% C.I.
    - *1st Cell (B29):* Enter =+$B29-1.96*$C29, where:
      - $B29 is (column-absolute opt.) reference to cell containing effect est
      - $C29 is (column-absolute opt.) reference to cell containing std err est
      - 1.96 is critical value for 95% C.I. on large-N t-dist or std-norm dist;
        - » can sub crit.val. for diff. C.I. % or smpl-size or use sprdsht formula
    - Copy Down
  - ***Column Five:*** Upper bound (analogous, but +1.96*$C29)

Ex's & Practice: Basic Linear-Interact Model: Education & Leftright Purely Micro-level Model

- In Stata, plot dY/dX w/ c.i. from smpl min-max:
  - `egen zmin = min(z)` ; `egen zmax = max(z)` finds those sample min & max for variable z. (z=leftright in our case, i.e., GS)
  - `gen z0 = (_n-1)/(v-1)*(zmax-zmin)+zmin in 1/v` creates var counting v equal-size steps from sample min to max.
  - `gen dyhatdx=_b[x]+_b[xz]*z0` creates var of dY/dX ests (x=education)
  - Stata code tedious to get to s.e.'s & c.i. plots (bit better in matrix form)
    - First have to work in matrices for bit, then back to vars:
      - `matrix V = get(VCE)` (makes matrix of v-cov mat)
      - `matrix C= V[3,1]` (grabs 3,1 element as covar)
      - `gen column1 = 1 in 1/v` (makes a variable equal to all ones)
      - `mkmat column1, matrix(col1)` (makes vector called col1 of that var)
      - `matrix cov_x_xz = C*col1` (makes a constant vector of covar)
      - `svmat cov_x_xz, name(cov_x_xz)` (makes that vector a variable)
    - Finally, you can generate variances & std errors, which you could tabulate:
      - `gen vardyhatdx=(_se[x])^2+(z0*z0)*(_se[xz]^2)+2*z0*cov_x_xz`
      - `gen sedyhatdx=sqrt(vardyhatdx)` (makes variable equal to s.e. of effect)
      - `tabdisp z0, cellvar(dyhatdx sedyhatdx)` (makes table effects & s.e.'s)
    - Or you can generate the confidence interval bounds & plot:
      - `gen LBdyhatdx=dyhatdx-invttail(e(df_r),.05)*sedyhatdx`
      - `gen UBdyhatdx=dyhatdx+invttail(e(df_r),.05)*sedyhatdx`
      - `twoway connected dyhatdx LBdyhatdx UBdyhatdx z0`

Ex's & Practice: Basic Linear-Interact Model

Calculating Predicted-Values & Standard Errors

Procedures	Command syntax

Create the variables which set the values of the variables *x*, *z*, and *xz* (& other variables, if any) for which $\hat{y}$ will be calculated.	`egen xmin = min(x)`
	`egen xmax = max(x)`
	`gen xh = ((_n-1)/(v-1))*(xmax-xmin) in 1/v`
	`egen zh=mean(z)`
	`gen xhzh=xh*zh`
	`gen col1h=1`

Assemble the variables into a matrix, **Mh**	`mkmat xh zh xhzh col1h in 1/v, matrix(Mh)`
Create **betas**, a column vector with *k x 1* dimensions.	`matrix betas=e(b)'`
Calculate the predicted values.	`matrix yhat=Mh*betas`
Convert the vector into a variable.	`svmat yhat, name(yhat)`
Create a matrix from the estimated covariance matrix of the coefficient estimates.	`matrix V = get(VCE)`
Calculate the variance of the predicted values.	`matrix VYH=Mh*V*Mh'`
Extract the diagonal elements into a row vector.	`matrix DVYH= vecdiag(VYH)`
Transpose elements into a column vector.	`matrix VARYHAT=DVYH'`
Convert the vector into a variable.	`svmat VARYHAT, name(varyhat)`
Calculate the estimated standard error of each predicted probability.	`gen seyhat1 = sqrt(varyhat1)`
Present a table of predicted values with corresponding standard errors of the predicted values.	`tabdisp xh, cellvar(yhat1 seyhat1)`
Generate lower and upper confidence interval bounds.	`gen LByhat1=yhat1-invttail(e(df_m),.05)*seyhat1` `gen UByhat1=yhat1+invttail(e(df_m),.05)*seyhat1`
Graph the predicted probabilities and the upper and lower confidence intervals.	`twoway connected yhat1 LByhat1 UByhat1 xh`

Google

'kam franzese michigan press'

for that data, and

stata & excel resources

or go directly to

`http://www.press.umich.edu/KamFranzese/Interactions.html`

Google

'Matt Golder interaction'

or go directly to

`http://homepages.nyu.edu/~mrg217/interaction.html`

Stata

`help mfx`

`help predictnl`

Ex's & Practice: Basic Linear-Interact Model

- (Near) Equivalence of Split-Sample (e.g., Unit-by-Unit) and Full-Dummy-Interaction Models:
  - Suffices to consider just 2 grps. Generalizes to n, but (even more) tedious.
  - *Setting Up:*
    - SET MEMORY: **. set memory 32m**
    - OPEN FILE (replace w/ location of merged, subset INTUNE data file):
      - **. use "…\WS_Data_level1_merge_level2.dta", clear**
    - DECLARE PANEL IDENITIFIER: **. iis group**
  - *GENERATE INDICATOR-IndVar INTERACTIONS (if needed):*
    - **. gen ctry1_edu=education*ctry1**
    - **. gen ctry2_edu=education*ctry2**
    - **. gen ctry1_lr=leftright*ctry1**
    - **. gen ctry2_lr=leftright*ctry2**
    - **. gen ctry1_edlr=edlr*ctry1** ; **. gen ctry2_edlr=edlr*ctry2**
  - *ESTIMATE UNIT-by-UNIT and FULL-INTERACTION MODELS:*
    - **. reg eu_support education leftright edlr if ctry1==1**
    - **. reg eu_support education leftright edlr if ctry2==1**
    - **. reg eu_support ctry1 ctry2 ctry1_edu ctry1_lr ctry1_edlr ctry2_edu ctry2_lr ctry2_edlr if group<2.5 , nocons**

Ex's & Practice: Basic Linear-Interact Model

- Coefficients same, but standard errors little different.
- This because:
  - Pooled full-dummy-interaction OLS assumes one, common σ^2 for both sub-samples.
  - Separate OLS estimation allows them to differ.
- Can allow σ^2 to vary by unit in pooled full-dummy-interaction by applying panel-weighted least-squares
  - Just FWLS with weights given by (inverse of square root of the) unit-by-unit estimated variance
  - This (Panel) FWLS yields perfectly identical results:
    - ***. xtgls eu_support ctry1 ctry1_edu ctry1_lr ctry1_edlr ctry2 ctry2_edu ctry2_lr ctry2_edlr if group<2.5 , nocons panels(hetero)***