Clarify Outline

- Installation
- The Basic Idea of Simulation (and why it makes sense for substantive interpretation)
- Components of Clarify
  - estsimp, setx, simqi
- A Real, Live Example
  - Logit Model
  - Continuous IVs
  - Binary IVs
- Concluding Points
  - Models Supported
  - How to ‘Clarify’ other Models

Installation

Type: net from http://gking.harvard.edu/clarify
Installation

net install clarify

clarify will either install or tell you it is already installed.
The Basic Idea of Simulation

So, you estimate a model… and you want to say something *substantive* with quantities of interest:

Predicted or Expected Values of DV = $X_\mu \hat{\beta}$

First Differences = $X_\sigma \hat{\beta} - X_\mu \hat{\beta}$

The problem is that our $\hat{\beta}$s are uncertain!

The solution is we know how uncertain.

\[
\hat{\beta}_1 \\
(\hat{\sigma}_1)
\]

The Basic Idea of Simulation: Parameters

In order to capture the uncertainty, we draw simulated $\hat{\beta}$s from the multivariate* normal distribution.

Then we use these simulated parameters to calculate many draws of the same quantity of interest.

\[
\hat{\beta}_1 \\
\text{Standard Deviation} = \hat{\sigma}_1
\]
The Basic Idea of Simulation: Quantities of Interest

In practice...

\[ Y_i = f(\theta_i, \alpha) \]
\[ \theta_i = g(X_i, \beta) \]
\[ Y_i \sim N(\mu_i, \sigma^2) \]
\[ \mu_i = g(X_i, \beta) = \beta_0 + X_i \beta_0 + X_i \beta_2 + \ldots \]

we simulate parameters with M draws from the multivariate normal distribution…

\[ \tilde{\theta} \sim N(\tilde{\beta}, \tilde{V}) \]

1. Choose a starting scenario, \( X_c \).
2. Draw one value of \( \tilde{\theta} \), and compute \( \tilde{\theta} = g(X_c, \tilde{\beta}) \).
3. Simulate the outcome \( \tilde{Y}_c \), by taking a random draw from \( f(\tilde{\theta}, \tilde{\alpha}) \).
4. Repeat M times to get the distribution of \( Y_c \).

Components of Clarify

- **estsimp** – estimates the model and simulates the parameters
  - This command **must** precede your regression command
  - e.g.: estsimp logit y x1 x2 x3 x4
  - This will save simulated \( \beta \)s to your dataset!
- **setx** – sets the values for the IVs (the \( X \)s)
  - Used after model estimation to set values of the Xs
  - e.g.: setx x1 mean x2 p20 x3 .4 x4[16], nocwdel
  - functions = mean|median|min|max|p#|math|#|‘macro’|varname[#]
  - reset values by re-issuing the command, e.g.: setx x1 median
- **simqi** – simulates the quantities of interest
  - Automates the simulation of quantities of interest for the \( X \) values you just set.
  - e.g.: simqi, prval(1)
  - e.g.: simqi, fd(prval(1)) changex(x4 p25 p75)

There are lots of options: Explore on your own!
Onto the Machines…

1. Clear the current data
   Type: clear

2. Increase memory
   • Type: set mem 50m

3. Re-open the NES data set
   • Type: use "I: \general \Spost&Clarify\NES 1992.dta"

We’ll do a Simple Logit…

Type: .estsim logit vote2 pid ideology gulfwarworth education, sims(500) genname(simb)

Note that Clarify has added 5 new variables to our data set.
1. Type: \textit{sum simb1-simb5} \\
2. Type: \textit{setx mean} \\

Tables of First Differences \\

Type: \textit{fd(prval(1)) changex(pid –3 –2))}
Probability of Bush Vote as PID Varies

<table>
<thead>
<tr>
<th>Party ID</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Bush)</td>
<td>0.049</td>
<td>0.089</td>
<td>0.156</td>
<td>0.263</td>
<td>0.408</td>
<td>0.571</td>
<td>0.719</td>
</tr>
<tr>
<td>95% CI</td>
<td>(.025, .091)</td>
<td>(.056, .141)</td>
<td>(.111, .223)</td>
<td>(.204, .330)</td>
<td>(.332, .487)</td>
<td>(.475, .669)</td>
<td>(.613, .818)</td>
</tr>
</tbody>
</table>

And, since we know \( P(Bush) = .273 (.212, .339) \) when every variable is held at its mean, we can calculate percentage changes ourselves to increase substantive interpretability.

But a picture is worth a thousand words, so it would be nice to use Clarify to generate pictures like this:

![Graph from King et al., APS 2000 Advanced Graphing with Clarify](image)
version 8.0
set more off
defined

gen plowar=.;
gen phiwar=.;
gen eduaxis = _n + 5 in 1/12;
setx gulfwarworth 1 ideology mean pid mean;
local i = 6;
while 'i' <= 17 {
   setx education 'i';
simqi, prval(1) genpr(pi);
Pctile pi, p(2.5,97.5);
replace plowar = r(r1) if eduaxis==`i';
replace phiwar = r(r2) if eduaxis==`i';
drop pi;
local i = `i'+1;
}
gen eduaxis2 = eduaxis -.2;
setx gulfwarworth 0 ideology mean pid mean;
local i = 6;
while 'i' <= 17 {
   setx education 'i';
simqi, prval(1) genpr(pi);
Pctile pi, p(2.5,97.5);
replace plowar = r(r1) if eduaxis==`i';
replace phiwar = r(r2) if eduaxis==`i';
drop pi;
local i = `i'+1;
}
gen plonowar=.;
gen phinowar=.;
setx gulfwarworth 1 ideology mean pid mean;
gen plowar=.;
gen phiwar=.;
gen eduaxis = _n + 5 in 1/12;
setx gulfwarworth 0 ideology mean pid mean;
local i = 6;
while 'i' <= 17 {
   setx education 'i';
simqi, prval(1) genpr(pi);
Pctile pi, p(2.5,97.5);
replace plowar = r(r1) if eduaxis==`i';
replace phiwar = r(r2) if eduaxis==`i';
drop pi;
local i = `i'+1;
}
gen eduaxis2 = eduaxis -.2;
setx gulfwarworth 0 ideology mean pid mean;
gen midwar = (plowar+phiwar)/2;
gen midnowar = (plonowar+phinowar)/2;
graph twoway rspike phiwar plowar eduaxis2 || line midwar eduaxis2 || rspike phiwar plonowar plowar eduaxis2 || line midnowar eduaxis, ytitle(P(Bush) Vote) xtitle(Education)
}
Conclusion

• Models Currently Supported by Clarify
  • `regress`
  • `logit`
  • `probit`
  • `ologit`
  • `oprobit`
  • `mlogit`
  • `poisson`
  • `nbreg`
  • `sureg`
  • `weibull`

• But, you really don’t need Clarify to do this, so you can simulate quantities of interest for *any* model!
  ✓ Easy to simulate parameters because Stata saves them after estimation!
  ✓ Program the correct link function yourself!