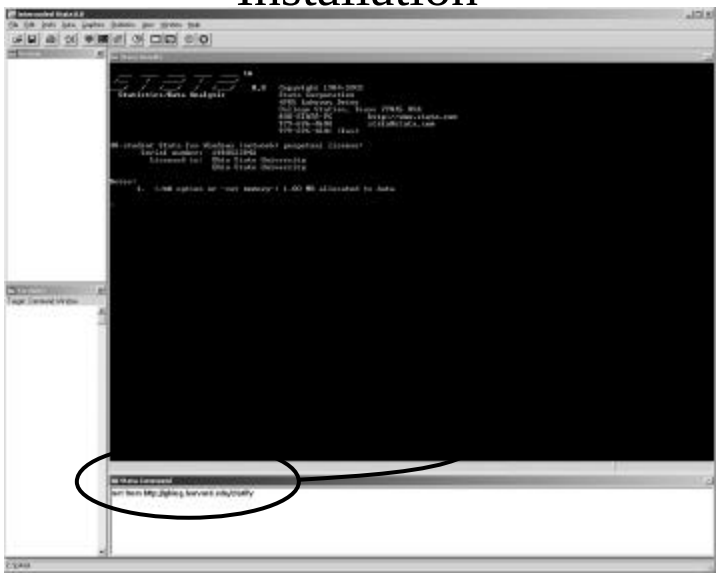
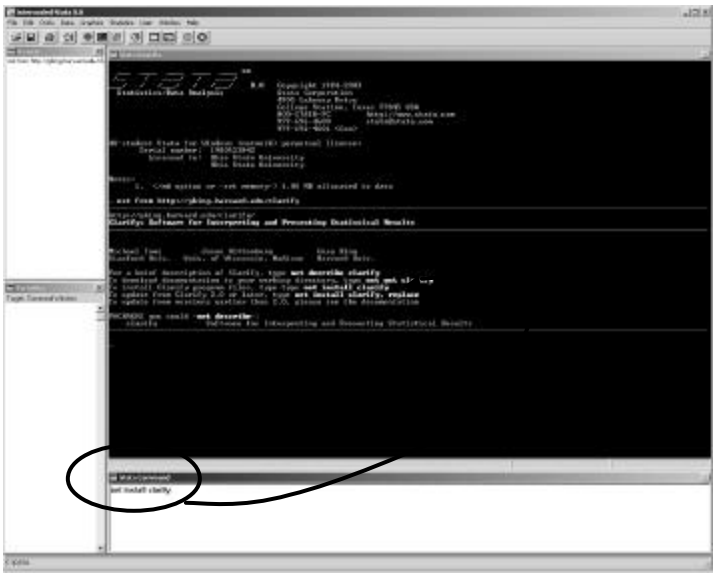
Clarify Outline

- 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

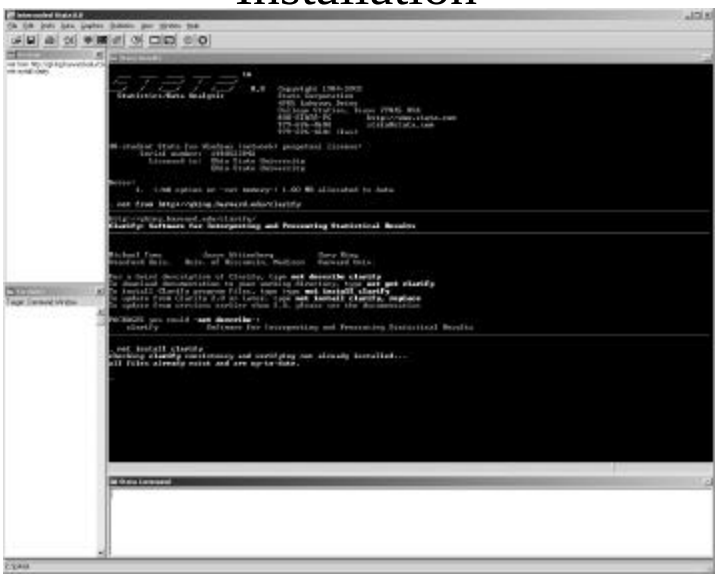
Installation



Installation



Installation



The Basic Idea of Simulation

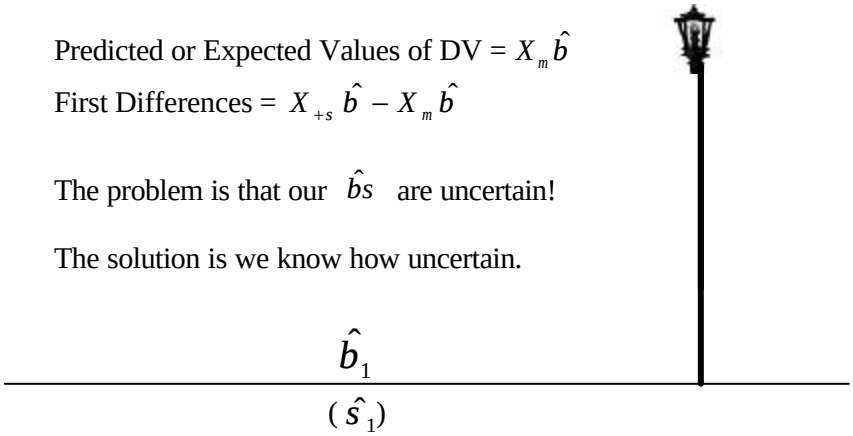
So, you estimate a model… and you want to say something *substantive* with quantities of interest:

Predicted or Expected Values of DV = $X_m \hat{b}$

First Differences = $X_{+s} \hat{b} - X_m \hat{b}$

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

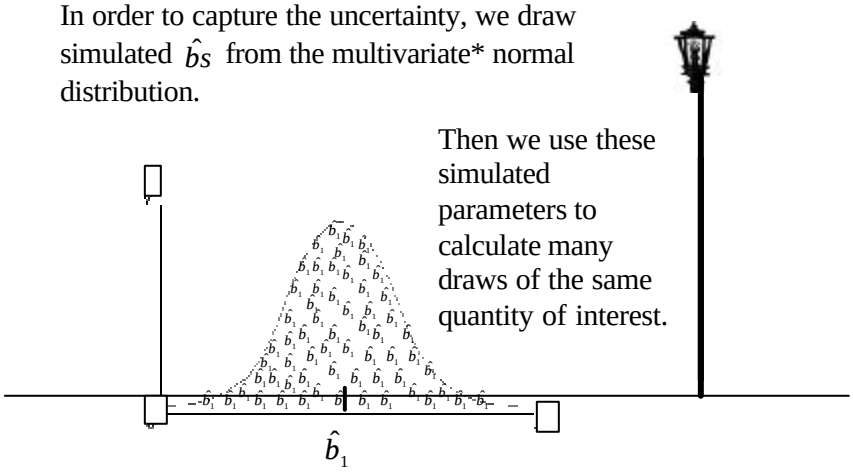
The solution is we know how uncertain.



The Basic Idea of Simulation: Parameters

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

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



Standard Deviation = $\hat{s}_1$

The Basic Idea of Simulation: Quantities of Interest

In practice… $Y_i \sim f(q_i, a)$, $q_i = g(X_i, b)$

$Y_i \sim N(m_i, s^2)$, $m_i = g(X_i, b) = b_0 + X_{i1}b_1 + X_{i2}b_2 + \cdots$

$$\hat{g} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{a}_1 \end{bmatrix} \qquad \hat{V}(\hat{g}) = \begin{bmatrix} v_{\hat{b}_{11}} & v_{\hat{b}_{12}} & \cdots & v_{\hat{b}_2\hat{a}} \\ v_{\hat{b}_{21}} & v_{\hat{b}_{22}} & \cdots & v_{\hat{b}_2\hat{a}} \\ \vdots & \vdots & & \vdots \\ v_{\hat{a}\hat{b}_1} & v_{\hat{a}\hat{b}_2} & \cdots & v_{\hat{a}} \end{bmatrix}$$

we simulate parameters with M draws from the multivariate normal distribution… $\tilde{g} \sim N(\hat{g}, \hat{V})$

$$\begin{bmatrix} \tilde{b}_{11} \\ \tilde{b}_{21} \\ \vdots \\ \tilde{a}_1 \end{bmatrix} \begin{bmatrix} \tilde{b}_{12} \\ \tilde{b}_{22} \\ \vdots \\ \tilde{a}_2 \end{bmatrix} \cdots \begin{bmatrix} \tilde{b}_{1M} \\ \tilde{b}_{2M} \\ \vdots \\ \tilde{a}_M \end{bmatrix}$$

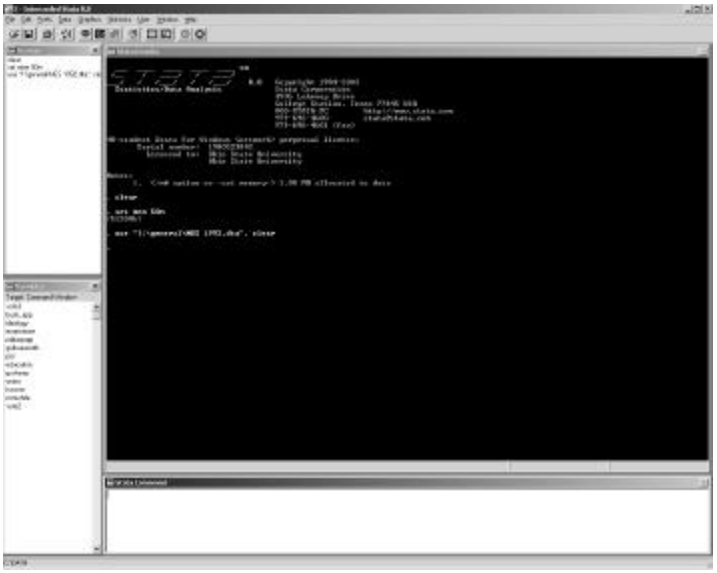
1. Choose a starting scenario, X_c.
2. Draw one value of $\tilde{g}$, and compute $\tilde{q}_c = g(X_c, \tilde{b})$.
3. Simulate the outcome $\tilde{Y}_c$, by taking a random draw from $f(\tilde{q}_c, \tilde{a})$.
4. Repeat M times to get the distribution of Y_c.

Components of Clarify

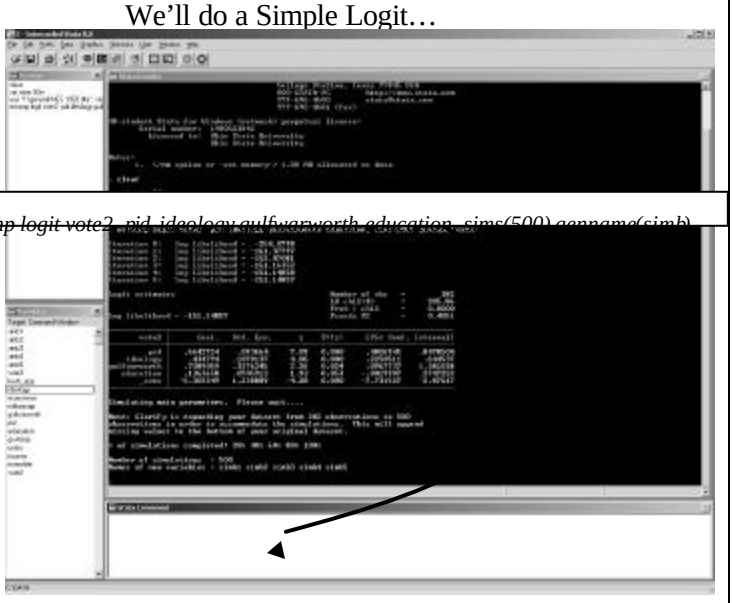
- *estsimp* – estimates the model and simulates the parameters
  - This command **must** precede your regression command
  - e.g.: *estsimp logit y x1 x2 x3 x4*
  - This will save simulated βs to your dataset!
- *setx* – sets the values for the IVs (the Xs)
  - Used after model estimation to set values 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…

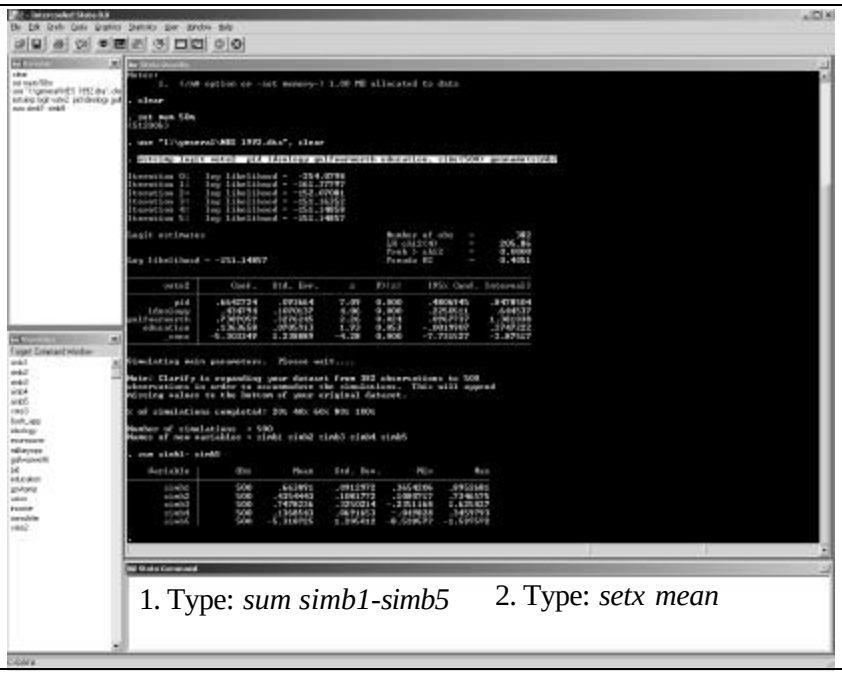


We'll do a Simple Logit…



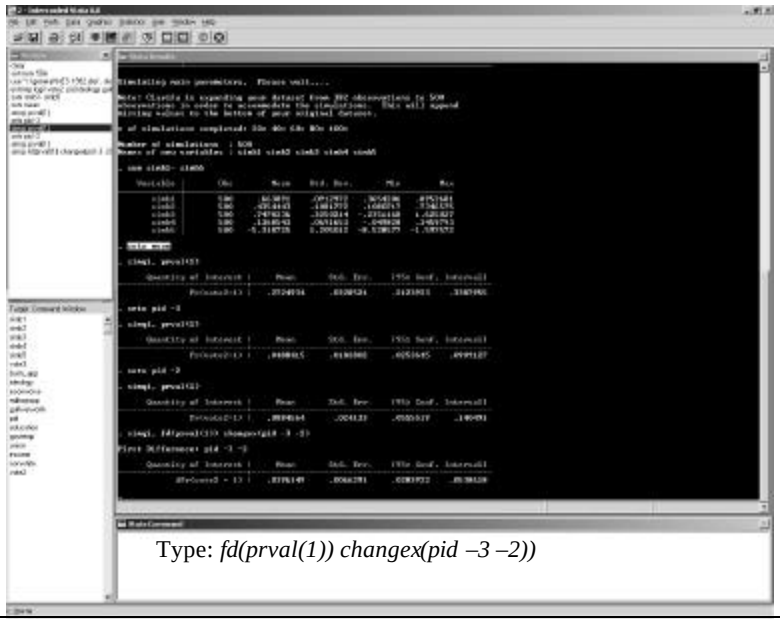
Type: *estsimp logit vote2 pid ideology gulfwarworth education sims(500) ganname(simb)*

Note that Clarify has added 5 new variables to our data set.



1. Type: *sum simb1-simb5* 2. Type: *setx mean*

Tables of First Differences



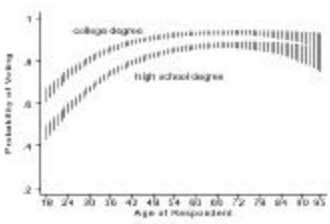
Type: *fd(prval(1)) changex(pid –3 –2))*

Probability of Bush Vote as PID Varies

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

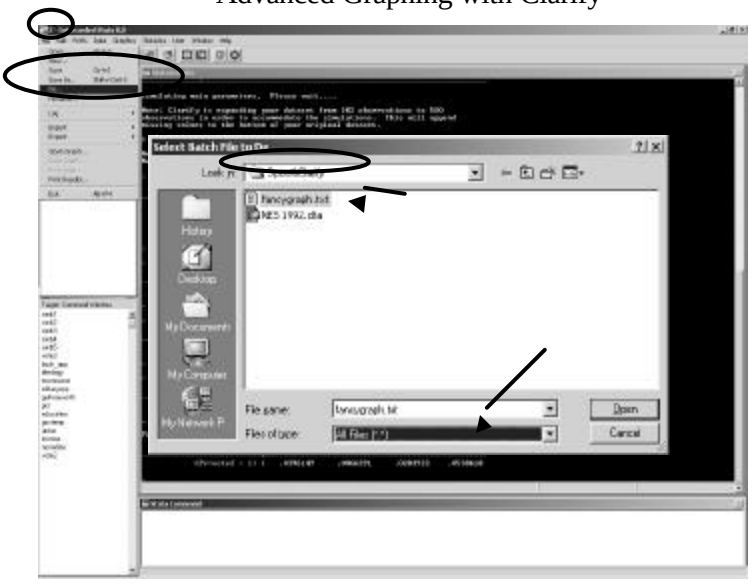
And, since we know P(Bush) is .273(.212, .339) when every variable is held at its mean, we can calculate percentage changes ourselves to increase substantive interpretability.

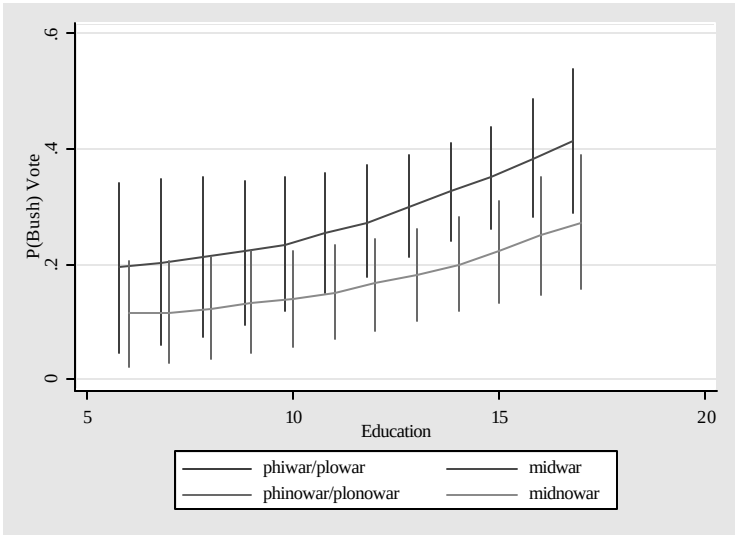
But a picture is worth a thousand words, so it would be nice to use Clarify to generate pictures like this:



From King et al., *AJPS* 2000

Advanced Graphing with Clarify






```
version 8.0
set more off
# delimit;
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 plonowar=.;
gen phinowar=.;

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 plonowar = r(r1) if eduaxis==`i';
        replace phinowar = r(r2) if eduaxis==`i';
        drop pi;
        local i = `i'+1;
};
```

```
gen eduaxis2 = eduaxis -.2;
sort eduaxis;
gen midwar = (plowar+phiwar)/2;
gen midnowar = (plonowar+phinowar)/2;
graph twoway rspike phiwar plowar eduaxis2 || line midwar eduaxis2 || rspike
phinowar plonowar eduaxis || 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!