SOFTWARE

The procedures that we recommend in “An Easy and Accurate Regression Model for Multiparty Electoral Data” have been incorporated into *Clarify: Software for Interpreting and Presenting Statistical Results*, a set of macros for use with the Stata statistics package. *Clarify* will calculate quantities of interest for a variety of statistical models, including linear regression, binary logit, binary probit, ordered logit, ordered probit, multinomial logit, poisson regression, negative binomial regression, weibull duration models, seemingly unrelated regression equations, and a growing list of others. The software and detailed documentation are available at http://gking.Harvard.edu. Here, we describe how researchers can use *Clarify* to analyze compositional data.

We can summarize the procedure in Section 2 as involving 4 basic steps:

1. Transform the vote shares (or other compositional data) into log ratios
2. Estimate a SUR and simulate the parameters
3. Choose some real or hypothetical values for the explanatory variables ($\underline{X}$’s)
4. Simulate the distribution of votes, conditional on the simulated parameters and chosen X’s.

Clarify includes a macro to achieve each of these steps.

> `tlogit`: applies the additive logistic transformation to compositional data
> `estsimp`: estimates the desired model and generates random draws of the parameters
> `setx`: sets the explanatory variables to desired values, such as means or percentiles
> `simqi`: computes desired quantities of interest, such as predicted vote shares

Each command comes with many options, but we will illustrate how to use *Clarify* by constructing a simple example. Suppose that we are studying a political system with 100 electoral districts. Each observation or row in our dataset pertains to one of those districts. In this example, we have three political parties that each garner a percentage of the vote. Their vote shares, collected in variables v1, v2, and v3, sum to 100 percent. We select party 3 as our reference party and transform the vote shares of the other two parties into log ratios with respect to party three. Thus, $Y_1 = \ln(V_1/V_3)$ and $Y_2 = \ln(V_2/V_3)$. The appropriate syntax in clarify is

```
tlogit v1 y1 v2 y2, base(y3) percent
```

which will create two new variables: y1 and y2, which are the log ratios for v1 and v2 with respect to the base variable v3.

Next, we use the estsimp sureg command to run a seemingly unrelated regression model with the log ratios y1 and y2 as our dependent variables. The syntax is

```
estsimp sureg (y1 x1 x2) (y2 x3 x4)
```

Each equation is enclosed in parentheses. Thus, the first equation states that the log ratio y1 is a linear function of the explanatory variables x1 and x2. The program will automatically add a constant term, as well, unless the user asks that it be suppressed. Likewise, the second equation states that y2 is a linear function of x3, x4, and a constant. The estsimp command will estimate the model and simulate the parameters. By default, estsimp will draw 1000 values for each parameter. In this example, the program would draw 1000 sets of betas (each set has six elements: three betas for equation 1 and three for equation two); the program would also generate 1000 simulations of S, a 2x2 matrix that governs the relationship between the errors of the two equations. *Clarify* will store these simulations in memory for subsequent use.

Third, we use the setx command to choose some hypothetical or real values for our explanatory variables. For instance, we might type

```
setx (x1 x2) mean x3 15 x4 p20
```

to set variables x1 and x2 at their respective means, x3 equal to the number 15, and x4 equal to its twentieth percentile. The setx command contains many options, which researchers can use to set the explanatory variables equal to any conceivable value.

Finally, we use the simqi command to simulate quantities of interest, such as the predicted distribution of votes. The command is

```
simqi, pv tfunc(logiti)
```

where tfunc(logiti) tells the program to apply the inverse logistic function to transform the value from log ratios into shares of the total vote.