STATA code for the maximum likelihood estimator used in Jana von Stein's
"Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance,"
The American Political Science Review 99 (4).

STATA code

- Please do not hesitate to contact Jana at jana.vonstein@vuw.ac.nz if you have any questions, comments, etc.
- Please also note: STATA updated a couple of commands between versions 6 and 7. If you need to use STATA 6 (or below), please e-mail me and I can send you the appropriate code.

Click to jump to:

You can also generate predicted probabilities and predicted marginal effects. There are two possible thought experiments: (1) imagine that all countries are not part of the selected group (e.g., all countries are "non-signing" types, as is done in my paper); compare their outcome equation behavior as non-signatories vs. if they had been "forced" to sign. If you use this approach and create "fake" observations to generate predicted probabilities at specific values of the independent variables, you will want to set those variables (for both the selection and the outcome equations) at the non-signatories' mean. (2) imagine all countries are part of the selected group (e.g., all countries are "signing" types); compare their outcome equation behavior as signatories vs. if they had been "forced" not to sign. If you use this approach and create "fake" observations to generate predicted probabilities at specific values of the independent variables, you will want to set those variables (for both the selection and the outcome equations) at the signatories' mean.
The decision whether to use (1) or (2) is to a large extent driven by substantive considerations (for reasons explained briefly in my paper, I found the first approach more reasonable from a policy perspective), but in any case I have provided both. You may, of course, report both, though that increases the complexity of figures.

The analyst may also wish to generate predicted probabilities and confidence intervals. Jana has Gauss code for this (though it's even less user-friendly than the STATA file, unfortunately…), which you can download from her webpage: http://www.bol.ucla.edu/~jvonstei/data-etc.html.

SOME COMMENTS/INSTRUCTIONS

(If you're not interested in too much detail on what's going on in the estimator, you might want to jump to the overview of Estimator 1).

- This code is for use in analyses in which (1) the outcome equation is observed for the group that is "selected" and the group that is not "selected"; (2) the outcome equation has a dichotomous dependent variable.
- **Estimator 1** yields five equations, all of which are estimated simultaneously:

1. **y1** is the selection equation dependent variable: for example, *sign treaty*. **ivar1** is the vector of independent variables predicting y1.

If the selection equation would normally require a hazard model (for example because the treaty commitment cannot be rescinded once it is made), the following variables need to be included in the model in addition to the substantive independent variables in order to estimate a probit model while still respecting the nature of the data:

(a) A dummy variable = 0 before signature and in the year of signature
= 1 in all years subsequent to signing the treaty.
This variable forces the probability that a signatory will sign in year *t* once it has signed to equal zero. Please see von Stein (2003) for greater detail.

(b) For each country-year, a variable marking the number of years since the country first appeared in the sample. This variable is precisely the "time until failure" variable upon which the hazard function is based.

(c) For each country-year, three temporal splines derived from (b) above.

*(b) and (c) above control for temporal dependence of the observations, and can be generated using Richard Tucker's BTSCS STATA code, available at http://www.vanderbilt.edu/~rtucker/programs/btscs/.

1a. **y1n** is –(y1). This is actually the selection equation for non-signatories, but since not signing is the opposite of signing, no additional input is needed for this.
For more detailed information on (a) through (c), please contact Jana. Her article also explains this approach in limited detail.

2. **y2** is the outcome equation dependent variable for the "selected" group: for example, *restrict the current account, given that one is a treaty signatory.* **ivar2** is the vector of independent variables predicting y2. In addition to substantive variables thought to predict the behavior of signatories, ivar2 should include whatever controls for temporal dependence one would include in a regular probit model. (For example, number of years since last restriction, etc.).
3. **y3** is the outcome equation dependent variable for the "non-selected" group: for example, *restrict the current account, given that one is* ***not*** *a treaty signatory.* **ivar3** is the vector of independent variables predicting y3. In addition to substantive variables thought to predict the behavior of non-signatories, ivar3 should include whatever controls for temporal dependence one would include in a regular probit model. (For example, number of years since last restriction, etc.).
4. **rhos** is the arctan of rho for signatories, the measure of the correlation of the error terms of the selection equation and the signatories' outcome equation.[1] The arctan allows ρ to range from $-\infty$ to $+\infty$ during statistical estimation. ρ is much more commonly normalized to range from –1 to +1 when reported in regression tables. Equation 4 (eq4), obtained from the STATA output from Estimator 1, reports the arctan of ρ^S. See the discussion of each estimator (below) for information on how to obtain the true ρ^S coefficient and p-value.
5. **rhon** is the arctan of rho for non-signatories, the measure of the correlation of the error terms of the selection equation and the signatories' outcome equation.[2] The arctan allows ρ to range from $-\infty$ to $+\infty$ during statistical estimation. ρ is much more commonly normalized to range from –1 to +1 when reported in regression tables. Equation 5 (eq5), obtained from the STATA output from Estimator 1, reports the arctan of ρ^N. See the discussion of each estimator (below) for information on how to obtain the true ρ^N coefficient and p-value.

[1] Substantively, rhos measures the extent to which the unobservable factors that lead countries to sign affect their post-signing behavior. If one believes that the factors that lead countries to sign make them less likely to engage in the behavior predicted in y2, one should expect a negative sign on the rhos coefficient.

[2] Substantively, rhon measures the extent to which the unobservable factors that lead countries *not* to sign affect their behavior. If one believes that the factors that lead countries not to sign make them more likely to engage in the behavior predicted in y3, one should expect a negative sign on the rhon coefficient.

ESTIMATOR 1

Estimator 1 should contain everything you need to run estimator 1.

- **The primary output of interest** is the large table at the very end of the output. eq1 yields the estimates for the selection equation; eq2 yields the estimates for the outcome equation for signatories; eq3 yields the estimates for the outcome equation for non-signatories; eq4 yields the arctan of ρ for signatories; eq5 yields the arctan of ρ for non-signatories. eq4 and eq5 must be further transformed (see the next lines, as well as Estimators 3 and 4 of this codebook).
- **Other outputs of interest (which appear after the large table described above) are:**
  (1) the true (i.e., transformed) ρ^S and ρ^N coefficients
  (2) the test that the constants in equations 2 and 3 are equal (which can be generally interpreted as the test that the treaty [or policy, treatment, etc.] has an impact independent of the sources of selection.
  (3) the test that an independent variable in the outcome equation affects those in the "selected" group in the same manner as those not in the "selected" group.

PREDICTED PROBABILITIES

You can also generate predicted probabilities and predicted marginal effects. There are two possible thought experiments: (1) imagine that all countries are not part of the selected group (e.g., all countries are "non-signing" types, as is done in my paper); compare their outcome equation behavior as non-signatories vs. if they had been "forced" to sign. If you use this approach and create "fake" observations to generate predicted probabilities at specific values of the independent variables, you will want to set those variables (for both the selection and the outcome equations) at the non-signatories' mean. (2) imagine all countries are part of the selected group (e.g., all countries are "signing" types); compare their outcome equation behavior as signatories vs. if they had been "forced" not to sign. If you use this approach and create "fake" observations to generate predicted probabilities at specific values of the independent variables, you will want to set those variables (for both the selection and the outcome equations) at the signatories' mean.
The decision whether to use (1) or (2) is to a large extent driven by substantive considerations (for reasons explained briefly in my paper, I found the first approach more reasonable from a policy perspective), but in any case I have provided both. You may, of course, report both, though that increases the complexity of figures.

The analyst may also wish to generate predicted probabilities and confidence intervals. Jana has Gauss code for this (though it's even less user-friendly than the STATA file, unfortunately…), which you can download from her webpage: http://www.bol.ucla.edu/~jvonstei/data-etc.html.

ESTIMATOR 2

Estimator 2 tests the null hypothesis that ρ^S and ρ^N jointly equal zero; that is, whether selection effects exist. This is an important test, since it tells us whether (from a statistical perspective) we should even bother using a selection model.
If all goes well when you run this estimator, STATA will give you a chi-square value and the corresponding p-value at the last line of the output. This is the **probability that the joint effect of ρ^S and $\rho^N = 0$**. Hence, a statistically significant result would indicate important selection effects.

ESTIMATOR 3

Estimator 3 tests the null hypothesis that ρ^S equals zero. It is the **p-value** that one would report along with the (normalized) ρ^s coefficient in a table, etc.
If all goes well when you run this estimator, STATA will give you a chi-square value and the **corresponding p-value at the last line of the output. This is the probability** that the effect of $\boldsymbol{\rho^s} = 0$. Hence, a statistically significant result (with a negative coefficient) would suggest that the unobserved factors that lead states to sign make them considerably less likely to engage in non-compliant behavior.

ESTIMATOR 4

Estimator 4 tests the null hypothesis that ρ^N equals zero. It is the **p-value** that one would report along with the (normalized) ρ^N coefficient in a table, etc.
If all goes well when you run this estimator, STATA will give you a chi-square value and the **corresponding p-value at the last line of the output. This is the probability** that the effect of $\boldsymbol{\rho^N} = 0$. Hence, a statistically significant result (with a negative coefficient) would suggest that the unobserved factors that lead states to not sign make them considerably more likely to engage in non-compliant behavior.