# Simultaneous Equations Models: what are they and how are they estimated

Omar M.G. Keshk

April 30, 2003

## 1 Simultaneity Or Reciprocal Causation in Political Science

Suppose that a researcher believes that two variables simultaneously determine each other. For example, a scholar of American politics may hypothesize that incumbent spending, in a campaign, is a function of challenger spending and simultaneously, challenger spending is a function of incumbent spending. A scholar of international relations may hypothesize that trade is a function of conflict and simultaneously, conflict is a function of trade. Finally, a comparativist may hypothesize that economic development is a function of democracy and simultaneously, democracy is a function of economic development. All these hypotheses share the common feature that the variables of interest are simultaneously determined. For example, challenger spending leads to changes in incumbent spending which in turn leads to changes in challenger spending which then leads to a change in incumbent spending, etc... The question becomes can such models be estimated using typical statistical procedures? If not, why not and what estimation methods can and should be used?

## 2 Background

Equations (1) & (2) present a generic two-equation model,[1]

$$y_1^* = \gamma_1 y_2^* + \beta_{\mathbf{1}}^{'}\mathbf{X_1} + \varepsilon_1 \tag{1}$$

$$y_2^* = \gamma_2 y_1^* + \beta_{\mathbf{2}}^{'}\mathbf{X_2} + \varepsilon_2 \tag{2}$$

As can be seen from the above equations, $y_1^*$ & $y_2^*$ simultaneously determine each other. Changes in $y_2^*$ will lead to changes in $y_1^*$ via (1). However, the resulting changes in (1) will immediately lead to changes in $y_2^*$ via (2). Variables that display such relationships are termed *endogenous variables.* So in the above equations, $y_1^*$ and $y_2^*$ would be termed endogenous variables. The remaining

[1] The following discussion borrows from Gujarati (2003: Ch. 18-20)

variables are termed exogenous. By itself, endogeneity is not a problem, the problem is that models containing such variables cannot be estimated by typical estimation procedures. For example, assume that $y_1^*$ & $y_2^*$ are observed as follows:

$$y_1 = y_1^*$$

$$y_2 = y_2^*$$

That is, $y_1^*$ & $y_2^*$ are fully observed.

OLS cannot be used to estimate these models, because the relationship specified by the equations violates the OLS assumption of zero covariance between the disturbance term and the independent variables. That is, in the above equations, the assumption that

$$E(\varepsilon_1|y_2^*) = \ E(\varepsilon_2|y_1^*) = 0$$

Or

$$Cov(y_2^*, \varepsilon_1) \ = \ Cov(y_1^*, \varepsilon_2) = 0$$

will be violated. To see this, note that changes in $\varepsilon_1$ will lead to changes in $y_1^*$ in equation (1). These changes in turn will lead to changes in $y_2^*$ via $y_1^*$ in equation (2). Thus, $y_2^*$ is correlated with $\varepsilon_1$, i.e., is a function of $\varepsilon_1$, indirectly. So information on $y_2^*$ will give information on $\varepsilon_1$ and so they are not mean independent. The same logic applies to the relationship between $\varepsilon_2$ and $y_1^*$.

Estimation of such models, via OLS, will lead to biased and inconsistent estimates of the coefficients. The most important part of the latter statement is the inconsistency, since no matter the sample size, the coefficients will never converge to the true population coefficients. To see this, consider a simplification of equation (1) and (2). Equation (2) is simplified to being an identity and equation (1) is simplified so there is only a constant and an endogenous variable. So now we have the following two equations:

$$y_1^* = \alpha_1 + \gamma_1 y_2^* + \varepsilon_1$$

$$y_2^* = y_1^* + x$$

where again, we assume that the $y_1^*$ & $y_2^*$ are observed in the following manner:

$$y_1 = y_1^*$$

$$y_2 = y_2^*$$

and $x$ is some variable with no impact on $y_1^*$. Now it is well known that the OLS estimate of $\gamma_1$ is obtained from the following formula:

$$\hat{\gamma}_1 = \frac{\Sigma(y_1^* - \bar{y}_1^*)(y_2^* - \bar{y}_2^*)}{\Sigma(y_2^* - \bar{y}_2^*)^2}$$

Rewriting $y_1^* - \bar{y}_1^* = \widetilde{y}_1^*$ and $y_2^* - \bar{y}_2^* = \widetilde{y}_2^*$, we have

$$\hat{\gamma}_1 = \frac{\Sigma(y_1^*)(\widetilde{y}_2^*)}{\Sigma(\widetilde{y}_2^*)^2}$$

Substituting for $y_1^*$, we get

$$\hat{\gamma}_1 = \frac{\Sigma(\alpha_1 + \gamma_1 y_2^* + \varepsilon_1)(\widetilde{y}_2^*)}{\Sigma(\widetilde{y}_2^*)^2}$$

$$\hat{\gamma}_1 = \frac{\Sigma(\alpha_1 \widetilde{y}_2^* + \gamma_1 y_2^* \widetilde{y}_2^* + \varepsilon_1 \widetilde{y}_2^*)}{\Sigma(\widetilde{y}_2^*)^2}$$

$$\hat{\gamma}_1 = \frac{\Sigma\alpha_1 \widetilde{y}_2^* + \Sigma\gamma_1 y_2^* \widetilde{y}_2^* + \Sigma\varepsilon_1 \widetilde{y}_2^*}{\Sigma(\widetilde{y}_2^*)^2}$$

$$\hat{\gamma}_1 = \frac{\alpha_1 \Sigma\widetilde{y}_2^* + \gamma_1 \Sigma y_2^* \widetilde{y}_2^* + \Sigma\varepsilon_1 \widetilde{y}_2^*}{\Sigma(\widetilde{y}_2^*)^2}$$

$$\hat{\gamma}_1 = \frac{\alpha_1 \Sigma\widetilde{y}_2^*}{\Sigma(\widetilde{y}_2^*)^2} + \frac{\gamma_1 \Sigma y_2^* \widetilde{y}_2^*}{\Sigma(\widetilde{y}_2^*)^2} + \frac{\Sigma\varepsilon_1 \widetilde{y}_2^*}{\Sigma(\widetilde{y}_2^*)^2}$$

Noting the following:

$$\Sigma\widetilde{y}_2^* = 0$$

$$\Sigma(\widetilde{y}_2^*)^2 = \Sigma y_2^* \widetilde{y}_2^*$$

We end up with

$$\hat{\gamma}_1 = \frac{\alpha_1 * 0}{\Sigma(\widetilde{y}_2^*)^2} + \frac{\gamma_1 \Sigma(\widetilde{y}_2^*)^2}{\Sigma(\widetilde{y}_2^*)^2} + \frac{\Sigma\varepsilon_1 \widetilde{y}_2^*}{\Sigma(\widetilde{y}_2^*)^2}$$

$$\hat{\gamma}_1 = \gamma_1 + \frac{\Sigma\varepsilon_1 \widetilde{y}_2^*}{\Sigma(\widetilde{y}_2^*)^2}$$

Taking probability limits (plim), we have

$$plim(\hat{\gamma}_1) = plim(\gamma_1) + plim(\frac{\Sigma\varepsilon_1 \widetilde{y}_2^*}{\Sigma(\widetilde{y}_2^*)^2})$$

$$plim(\hat{\gamma}_1) = \gamma_1 + plim\left(\frac{\frac{\Sigma\varepsilon_1 \widetilde{y}_2^*}{N}}{\frac{\Sigma(\widetilde{y}_2^*)^2}{N}}\right) \tag{3}$$

Note that $\frac{\Sigma(\widetilde{y}_2^*)^2}{N}$ is nothing but the variance of $y_2^*$ and we can write it as $\sigma_y^2$. Now if plim $(\Sigma\varepsilon_1 \widetilde{y}_2^*)$ does not converge to zero, it is evident that the estimate $\hat{\gamma}_1$ will be biased and inconsistent. What does plim $(\Sigma\varepsilon_1 \widetilde{y}_2^*)$ converge to? To see that it is not zero, note the following, we return again to our simplified example above:

$$y_1^* = \alpha_1 + \gamma_1 y_2^* + \varepsilon_1 \tag{4}$$

$$y_2^* = y_1^* + x \tag{5}$$

Inserting (4) into (5) and solving for $y_2^*$ we get

$$y_2^* = \alpha_1 + \gamma_1 y_2^* + \varepsilon_1 + x$$

$$y_2^* - \gamma_1 y_2^* = \alpha_1 + \varepsilon_1 + x$$

$$y_2^*(1-\gamma_1) = \alpha_1 + \varepsilon_1 + x$$
$$y_2^* = \frac{\alpha_1}{(1-\gamma_1)} + \frac{\varepsilon_1}{(1-\gamma_1)} + \frac{x}{(1-\gamma_1)} \tag{6}$$

Taking expectations, we obtain

$$E(y_2^*) = E(\frac{\alpha_1}{(1-\gamma_1)}) + E(\frac{\varepsilon_1}{(1-\gamma_1)}) + E(\frac{x}{(1-\gamma_1)})$$

Noting that the $E(\varepsilon_1) = 0$ (by assumption) and $\alpha_1$, $x$ & $(1-\gamma_1)$ are constants, the above becomes

$$E(y_2^*) = \frac{\alpha_1}{(1-\gamma_1)} + \frac{0}{(1-\gamma_1)} + \frac{x}{(1-\gamma_1)}$$

$$E(y_2^*) = \frac{\alpha_1}{(1-\gamma_1)} + \frac{x}{(1-\gamma_1)}$$

Subtracting this from (6), we get

$$y_2^* - E(y_2^*) = \frac{\alpha_1}{(1-\gamma_1)} - \frac{\alpha_1}{(1-\gamma_1)} + \frac{x}{(1-\gamma_1)} - \frac{x}{(1-\gamma_1)} + \frac{\varepsilon_1}{(1-\gamma_1)}$$

$$y_2^* - E(y_2^*) = \frac{\varepsilon_1}{(1-\gamma_1)} \tag{7}$$

Now

$$\varepsilon_1 - E(\varepsilon_1) = \varepsilon$$

because $E(\varepsilon_1) = 0$. Therefore,

$$Cov(y_2^*, \varepsilon_1) = E[y_2^* - E(y_2^*)][\varepsilon_1 - E(\varepsilon_1)]$$

$$Cov(y_2^*, \varepsilon_1) = E[y_2^* - E(y_2^*)][\varepsilon_1]$$

Replacing $y_2^* - E(y_2^*)$ with (7), we get

$$Cov(y_2^*, \varepsilon_1) = E[\frac{\varepsilon_1}{(1-\gamma_1)}][\varepsilon_1]$$

$$Cov(y_2^*, \varepsilon_1) = E(\frac{\varepsilon_1^2}{(1-\gamma_1)})$$

$$Cov(y_2^*, \varepsilon_1) = \frac{\sigma_1^2}{(1-\gamma_1)}$$

Substituting this into the numerator of (3), and recalling that we replaced $\frac{\Sigma(\tilde{y}_2^*)^2}{N}$ with $\sigma_y^2$. we have

$$plim(\hat{\gamma}_1) = \gamma_1 + \frac{\frac{\sigma_1^2}{(1-\gamma_1)}}{\sigma_y^2}$$

$$plim(\hat{\gamma}_1) = \gamma_1 + \frac{1}{(1-\gamma_1)}\frac{\sigma_1^2}{\sigma_y^2}$$

Given that both $\sigma_1^2$ & $\sigma_y^2$ are positive, plim ($\hat{\gamma}_1$) is not only biased but this bias remains no matter the sample size and thus $\hat{\gamma}_1$ is not a consistent estimator of $\gamma_1$. What can be done in such a situation? Well, one possible solution would be to see if it is possible to separate or partition the endogenous variable into a part that is correlated with the disturbance term and a part that is not correlated with the disturbance term and then use the latter to estimate the model? As we will see, this is, in general, quite possible.

# 3 Instrumental Variable (IV) Estimation

## 3.1 What is IV estimation

Instrumental variable estimation is a broad class of estimation techniques for dealing with, among other things, correlation between independent variables and disturbance terms. In a nutshell, IV estimation requires that a researcher *find* a variable(s) or *create* a variable that is highly correlated with the endogenous variable and uncorrelated with the disturbance term. Symbolically, we need to find a variable $z$ such that

$$plim(\frac{zy^*}{N}) = \sigma_{zy} > 0$$

$$plim(\frac{z\varepsilon}{N}) = \sigma_{z\varepsilon} = 0$$

The term given to the variable $z$ is instrumental variable. What is it an instrument of or for? It is an instrument (proxy) for the endogenous variable, that is, we will use it in place of the endogenous variable. How do you *find* or *create* such an instrument. With respect to the former, this is usually very difficult and when one is found, its suitability is questioned.[2] I believe the third method lunch will go into more detail with respect to this issue and so I will say no more.

The most common avenue taken by researchers is to *create* an instrument, using the method of two-stage least squares (2SLS). Following the theory behind IV estimation, 2SLS creates an instrument that is correlated with the endogenous variable while uncorrelated with the disturbance term. That is, it separates the endogenous variable into two parts, one correlated with the disturbance term and another uncorrelated with the disturbance term. And then uses the latter, in place of the original endogenous variable to estimate the model. The procedure to create such a variable is discussed in the next section.

[2]See Bound, Joah, David A. Jaeger, and Regina M. Baker (1995). "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association* Vol. 90, N0. 430.

### 3.2 Two-Stage Least Squares (2SLS)

Two stage least squares, as the name suggests, is a two stage process. In the first stage, the endogenous variable is regressed on all the exogenous variables and the predicted value of this regression is obtained. In the second stage, the predicted values replace the original endogenous variables in the equation and estimation is carried out. Returning to our original example, we have the following two equations, that we wish to estimate (these are called the structural equations)

$$y_1^* = \gamma_1 y_2^* + \beta_{\mathbf{1}}' \mathbf{X_1} + \varepsilon_1 \tag{8}$$

$$y_2^* = \gamma_2 y_1^* + \beta_{\mathbf{2}}' \mathbf{X_2} + \varepsilon_2 \tag{9}$$

Again, assume that $y_1^*$ & $y_2^*$ are fully observed, i.e.,

$$y_1 = y_1^*$$

$$y_2 = y_2^*$$

The first stage in 2SLS is to estimate the following equations, via OLS, since $y_1^*$ & $y_2^*$ are fully observed

$$\begin{aligned} y_1^* &= \beta_{\mathbf{1}}' \mathbf{X_1} + \beta_{\mathbf{2}}' \mathbf{X_2} + \upsilon_1 \\ &= \mathbf{\Pi_1 X} + \upsilon_1 \end{aligned} \tag{10}$$

$$\begin{aligned} y_2^* &= \beta_{\mathbf{1}}' \mathbf{X_1} + \beta_{\mathbf{2}}' \mathbf{X_2} + \upsilon_2 \\ &= \mathbf{\Pi_2 X} + \upsilon_2 \end{aligned} \tag{11}$$

Where $\mathbf{X}$ is a matrix containing all the exogenous variables in (8) and (9). Note that the term given to (10) and (11) is reduced form equations and they are equations that express the endogenous variables solely in terms of the exogenous variables. From (10) and (11), we obtain

$$\hat{y}_1^* = \mathbf{\hat{\Pi}_1 X} \tag{12}$$

$$\hat{y}_2^* = \mathbf{\hat{\Pi}_2 X} \tag{13}$$

This completes the first stage of 2SLS.[3]

In the second stage, we replace $y_1^*$ & $y_2^*$ in (8) and (9) with $\hat{y}_1^*$ and $\hat{y}_2^*$, respectively, and estimate the equations (8) and (9) with OLS. That is we now estimate,

$$y_1^* = \gamma_1 \hat{y}_2^* + \beta_{\mathbf{1}}' \mathbf{X_1} + \varepsilon_1 \tag{14}$$

[3]In certain situations, not discussed here, it is possible to recover the coefficient estimates in of the original model from this stage alone. If this is done, then the method used is called Indirect Least Squares.

$$y_2^* = \gamma_2 \hat{y}_1^* + \beta_{\mathbf{2}}^{'} \mathbf{X_2} + \varepsilon_2 \tag{15}$$

The resulting coefficients estimates will be consistent, although, in small samples the IV estimator will be a biased estimator, as explained below. Note what 2SLS does, in (10) we created a new variable $\hat{y}_1^*$ from all the exogenous variables in both equation (10) & (11). By construction, this variable will be uncorrelated with $\varepsilon_1$ since the exogenous variables in (10) & (11) are assumed to be uncorrelated with the error terms. Thus, the new variable is uncorrelated with the error term by construction and using it in (17) does not violate any OLS assumptions.

## 3.3 Issues in 2SLS and IV Estimation in General

When applying 2SLS there are certain issues that must be considered and kept in mind.

1. **IDENTIFICATION**
Identification is generally a tedious part of IV estimation and 2SLS and I will not go into the dense detail that most books go into. However, given how 2SLS is conducted it should be clear that we have to place certain restrictions on the equations. For 2SLS, the identification issue boils down to insuring that at least one exogenous variable appearing in one equation does not appear in the other equation. Furthermore, the more such restrictions (exclusions) the better. Why? Look back to the following equations:
$$y_1^* = \gamma_1 \hat{y}_2^* + \beta_{\mathbf{1}}^{'} \mathbf{X_1} + \varepsilon_1 \tag{16}$$
$$y_2^* = \gamma_2 \hat{y}_1^* + \beta_{\mathbf{2}}^{'} \mathbf{X_2} + \varepsilon_2 \tag{17}$$
Remember that $\hat{y}_2^*$ was constructed by regressing it on all the exogenous variables. If all the exogenous variables appear in both equations, that is if the original structural equations looked like this:
$$y_1^* = \gamma_1 y_2^* + \beta_{\mathbf{1}}^{'} \mathbf{X_1} + \beta_{\mathbf{1}}^{'} \mathbf{X_2} + \varepsilon_1 \tag{18}$$
$$y_2^* = \gamma_2 y_1^* + \beta_{\mathbf{2}}^{'} \mathbf{X_1} + \beta_{\mathbf{2}}^{'} \mathbf{X_2} + \varepsilon_2 \tag{19}$$
Inserting $\hat{y}_2^*$, which is a linear combination of $\mathbf{X_1}$ and $\mathbf{X_2}$, into (18) would result in perfect collinearity and OLS estimation would not work since the matrix will not be invertible. Exclusion insures that perfect collinearity will not occur.

2. **QUALITY OF INSTRUMENT**
As stated above, the purpose of 2SLS is to *create* an instrument that is not correlated with the error term but is correlated with the endogenous variable. Now recall that this instrument stands in or acts as a proxy for the endogenous variable and as such the higher the correlation with the endogenous variable the better. Again, I believe that the third method

lunch will discuss all of this so I will not say much more, except that goodness of fit measures should be looked at after the first stage.[4]

3. **2SLS INSTRUMENTS ARE BIASED**
The literature sometimes ignores the fact that instruments from 2SLS are biased in small samples. Why? In the first stage the instruments are generated from regressions. As such they are linear combinations of the exogenous variables and the estimated coefficients in the reduced form equations. Thus, the instruments (i.e., the predicted values) are themselves a function of the error terms in the reduced form equations, which in turn are components of the error terms in the structural equations and thus, the instruments are likely to be correlated with the error terms in the structural equations. This is something that cannot be avoided and this is why the emphasis in the literature is on the consistency of using 2SLS, since as the same size increases this correlation between instruments and error terms disappear. This correlation, however, does not disappear, no matter the sample size, if the endogenous variables are used.[5]

4. **STANDARD ERRORS ARE WRONG**
Standard errors from 2SLS will be wrong and need to be corrected. To see why this is the case let us look at a single equation from the two given above. So let us look at the following equation:

$$y_1^* = \gamma_1 y_2^* + \beta_{\mathbf{1}}^{'} \mathbf{X_1} + \varepsilon_1 \tag{20}$$

This is the equation we want, however, in 2SLS, we estimate

$$y_1^* = \gamma_1 \hat{y}_2^* + \beta_{\mathbf{1}}^{'} \mathbf{X_1} + \varepsilon_1 \tag{21}$$

Thus, the error term $\varepsilon_1$ is really made up of $\varepsilon_1 + \gamma_1 \hat{v}_1$. To see this, recall that $y_2^* = \hat{y}_2^* + \hat{v}_1$, substituting this for $y_2^*$ in (20), we have

$$\begin{aligned} y_1^* &= \gamma_1(\hat{y}_2^* + \hat{v}_1) + \beta_{\mathbf{1}}^{'} \mathbf{X_1} + \varepsilon_1 \\ y_1^* &= \gamma_1 \hat{y}_2^* + \gamma_1 \hat{v}_1) + \beta_{\mathbf{1}}^{'} \mathbf{X_1} + \varepsilon_1 \\ y_1^* &= \gamma_1 \hat{y}_2^* + \beta_{\mathbf{1}}^{'} \mathbf{X_1} + (\varepsilon_1 + \gamma_1 \hat{v}_1) \\ y_1^* &= \gamma_1 \hat{y}_2^* + \beta_{\mathbf{1}}^{'} \mathbf{X_1} + \varepsilon_1^* \end{aligned} \tag{22}$$

Where $\varepsilon_1^* = \varepsilon_1 + \gamma_1 \hat{v}_1$. The form of the correction depends on nature of the endogenous variables and this will be discussed in the next section of the paper.

[4]For an excellent exposition (although not fully from a 2SLS perspective) see Bartels, L. M.(1991) "Instrumental and Quasi-Instrumental Variables," *American Journal of Political Science*, 33, 777-800.

[5]See Gujarati (2003) p. 772, ft. 14.

# 4 Different Forms of Simultaneous Equations and How to Estimate them

So far all the examples given have dealt with fully observed variables. However, sometimes this is not the case (as political scientist are well aware). So in this section, I present the different possible equations depending on how the endogenous variables are observed and what this means for estimating them. All estimation will be in the 2SLS context, however, with certain modifications. Let us return to our generic two equation model:

$$y_1^* = \gamma_1 y_2^* + \beta_{\mathbf{1}}^{'}\mathbf{X_1} + \varepsilon_1 \tag{23}$$

$$y_2^* = \gamma_2 y_1^* + \beta_{\mathbf{2}}^{'}\mathbf{X_2} + \varepsilon_2 \tag{24}$$

The proper estimation strategy to be used depends on how $y_1^*$ & $y_2^*$ are observed.[6]

A. First, if $y_1^*$ & $y_2^*$ are observed as follows[7]

$$y_1 = y_1^*$$

and

$$y_2 = y_2^*$$

That is, both variables are fully observed, then we have the typical simultaneous equations models discussed in the statistical literature. The 2SLS procedure uses all OLS procedures. That is, in the first stage the reduced form equations are estimated using OLS and in the second stage, the modified structural equations are also estimated via OLS. So we have the following estimation steps:
Step 1: estimate

$$y_1^* = \mathbf{\Pi_1^{'}X} + \upsilon_1 \tag{25}$$

$$y_2^* = \mathbf{\Pi_2^{'}X} + \upsilon_2 \tag{26}$$

Where $\mathbf{X}$ is a matrix containing all the exogenous variables in the system of equations. Obtain, $\hat{y}_1^*$ and $\hat{y}_2^*$, then go to the next step.
Step 2: estimate

$$y_1^* = \gamma_1 \hat{y}_2^* + \beta_{\mathbf{1}}^{'}\mathbf{X_1} + \varepsilon_1 \tag{27}$$

$$y_2^* = \gamma_2 \hat{y}_1^* + \beta_{\mathbf{2}}^{'}\mathbf{X_2} + \varepsilon_2 \tag{28}$$

The estimated coefficients from this last step are biased but consistent. The final step is the correction of the standard errors and in this case, is a simple two step process in which the coefficients of each parameter in the final step are multiplied by the ratio of the standard deviation of the disturbance term in the second step estimates to the standard deviation

[6] The following discussion borrows from Maddala (1983: 242-247).
[7] This corresponds to Maddala's (1983, 243) model 1.

of the disturbance term in the original structural equation.[8] That is, in the second step we estimate and obtain

$$y_1^* = \hat{\gamma}_1 \hat{y}_2^* + \hat{\beta}_{\mathbf{1}}^{'} \mathbf{X_1} + \hat{\varepsilon}_1 \tag{29}$$

and from this obtain $\frac{\Sigma(\hat{\varepsilon}_1^*)^2}{N-k} = \hat{\sigma}^2_{\varepsilon_1^*}$. Recalling that $\varepsilon_1^* = \varepsilon_1 + \gamma_1 \hat{\upsilon}_1$. Then estimate and obtain

$$y_1^* = \hat{\gamma}_1 y_2^* + \hat{\beta}_{\mathbf{1}}^{'} \mathbf{X_1} + \hat{\varepsilon}_1$$

(notice no hat on $y_2^*$ and obtain $\frac{\Sigma(\hat{\varepsilon})^2}{N-k} = \hat{\sigma}^2_{\varepsilon_1}$. Multiply each coefficient's standard error in equation (29) by $\frac{\hat{\sigma}_{\varepsilon_1}}{\hat{\sigma}_{\varepsilon_1^*}}$. Now do the same for the other equation.
STATA has build in procedures to estimate such models.Check out the help file for *reg3*. Two caveats, however, the above estimation and correction does not take into consideration the possible correlation between error terms across equations. Methods to do this involve something called three stages least squares, I have not done any work in this vein and so I am not familiar with the math and so I will not discuss it. The command *reg3* has a built in option 2SLS and this will perform the method discussed above. Although it is not clear whether it implements the correction outline above for the standard errors. So make sure to check the documentation.

B. If we face a situation in which $y_1^*$ & $y_2^*$ are observed as follows:[9]

$$\begin{aligned} y_1 &= y_1^* \\ y_2 &= 1 \text{ if } y_2^* > 0 \\ y_2 &= 0 \text{ otherwise} \end{aligned}$$

That is, $y_1^*$ is fully observed and $y_2^*$ is observed as a dichotomy, then we need to modify the 2SLS process and the name given to it is: two stage probit least squares (2SPLS). [10] The modified 2SPLS process is this: First estimate the reduced form equations, which are as follows:
Step 1: estimate the following equations

$$y_1^* = \mathbf{\Pi_1^{'}} \mathbf{X} + \upsilon_1 \tag{30}$$

$$y_2^* = \mathbf{\Pi_2^{'}} \mathbf{X} + \upsilon_2 \tag{31}$$

[8] Gujarati (2003: 791).
[9] This corresponds to Maddala's (1983, 244-5) model 3.
[10] To the best of my knowledge, the term 2SPLS was given to the procedure by Alvarez and Glasgow (2000). Other terms for this procedure include Generalized Two-Stage Probit ; Two-Step Probit Estimator. I prefer 2SPLS because it provides a more complete description of steps and estimations used.

Equation (30) is estimated via OLS and the predicted value is obtained. Equation (31) is estimated via probit and here we obtain the ***linear predictor*** for use in the second stage.[11]

Step 2: estimate the following equations

$$y_1^* = \gamma_1 \hat{y}_2^* + \beta_{\mathbf{1}}^{'}\mathbf{X_1} + \varepsilon_1 \tag{32}$$

$$y_2^* = \gamma_2 \hat{y}_1^* + \beta_{\mathbf{2}}^{'}\mathbf{X_2} + \varepsilon_2 \tag{33}$$

Equation (32) is estimated using OLS and equation (33) via Probit. Again, however, the standard errors are wrong and need to be corrected. The correction in this case is a little more detailed and involve estimating the following variance covariance matrices. First define the following:

$$\alpha_1^{'} = (\gamma_1\sigma_2, \beta_1^{'}) \tag{34}$$

$$\alpha_2^{'} = (\frac{\gamma_2}{\sigma_2}, \frac{\beta_2^{'}}{\sigma_2}) \tag{35}$$

$$Cov(\upsilon_1, \upsilon_2) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix} \tag{36}$$

$$c = \sigma_1^2 - 2\gamma_1\sigma_{12} \tag{37}$$

$$d = \left(\frac{\gamma_2}{\sigma_2}\right)\sigma_1^2 - 2\left(\frac{\gamma_2}{\sigma_2}\right)\left(\frac{\sigma_{12}}{\sigma_2}\right) \tag{38}$$

$$H = (\Pi_2, J_1) \tag{39}$$

$$G = (\Pi_1, J_2) \tag{40}$$

$$V_0 = Var(\hat{\Pi}_2) \tag{41}$$

With these definitions at hand, and noting that in Probit models $\sigma_2$ is normalized to 1, the corrected variance covariance matrices for $\alpha_1$ & $\alpha_2$ can be obtained as follows:

$$V(\hat{\alpha_1}) = c(H^{'}X^{'}XH)^{-1} + (\gamma_1\sigma_2)^2(H^{'}X^{'}XH)^{-1}H^{'}X^{'}V_0X^{'}XH(H^{'}X^{'}XH)^{-1} \tag{42}$$

$$V(\hat{\alpha_2}) = (G^{'}V_0^{-1}G)^{-1} + d(G^{'}V_0^{-1}G)^{-1}G^{'}V_0^{-1}(X^{'}X)^{-1}V_0^{-1}G(G^{'}V_0^{-1}G)^{-1} \tag{43}$$

Where

- $\sigma_1^2$ is the variance of the residuals from (30)
- $V_0$ is the variance-covariance matrix of (31)
- $\mathbf{J}_1$ and $\mathbf{J}_2$ are matrices with ones and zeros such that $\mathbf{XJ_1} = \mathbf{X_1}$ and $\mathbf{XJ_2} = \mathbf{X_2}$

[11] This is option *predict name, xb* after probit in STATA.

- $\sigma_{12}$ is obtained using the formula $\frac{1}{N}\frac{\Sigma(d_t\hat{v}_1)}{\hat{f}}$
  where
  $N$ is the number of observations
  $d_t$ is the dichotomous endogenous variable,
  $\hat{v}_1$ is the residuals from (30) and
  $\hat{f}$ is (31) evaluated using the standard normal density.

Currently, STATA does not have a procedure to implement the estimation of these type models. So, for the paper with Dr. Reuveny and Dr. Pollins, I wrote such a procedure (cdsimeq) and it implements all the necessary steps to obtain consistent estimates and corrected standard errors.

The syntax is as follows:
`cdsimeq (`*continuous_endogenous_depvar* `continuous_model_exogenous_indvar(s))` `(`*dichotomous_endogenous_depvar* `dichotomous_model_exogenous_indvar(s))` [`if` *exp*] [`in` *range*] [`, NOFirst NOSecond asis INStpre ESTimates_hold` ]

The options for the cdsimeq command are as follows:

- NOFirst specifies that the displayed output from the *first stage* estimations be suppressed.
- NOSecond specifies that the displayed output from the *second stage* estimations be suppressed.
- asis is Stata's asis option, see [R] **Probit**.
- INStpre specifies that the created instruments in the first stage are not to be discarded after the program terminates. Note that if this option is specified and the program is re-run, an error will be issued saying that the variables already exist. Therefore, these variables have to be dropped or renamed before cdsimeq can be re-run.
- ESTimates_hold retains the estimation results from the OLS estimation, with corrected standard errors, in a variable called model_1 and estimation results from the Probit estimation, with corrected standard errors, in a variable called model_2.[12] Note that if this option is specified the above variables must be dropped before cdsimeq command is re-run again with the estimates_hold option.

The cdsimeq command provides the following saved estimation results:

| | | | |
|---|---|---|---|
| `e(sigma_11)` | $\sigma_{11}$ | `e(sigma_12)` | $\sigma_{12}$ |
| `e(gamma_2)` | $\gamma_2$ | `e(gamma_2_sq)` | $\gamma_2^2$ |
| `e(MA_c)` | $\sigma_1^2 - 2\gamma_1\sigma_{12}$ | `e(MA_d)` | $(\frac{\gamma_2}{\sigma_2})\sigma_1^2 - 2(\frac{\gamma_2}{\sigma_2})(\frac{\sigma_{12}}{\sigma_2})$ |
| `e(F)` | (F from 1st stage) | `e(R)` | (OLS R from 1st stage) |
| `e(adj_R)` | (adjusted R from 1st stage) | `e(chi2)` | (Probit Chi2 from 1st stage) |
| `e(r2_p)` | (Probit Pseudo R from 1st stage) | | |

Finally, here is a stylized output from running the command, for illustrative purposes only:

[12]When this option is specified the created instruments are also preserved.

```
cdsimeq (continuous exog3 exog2 exog1 exog4) (dichotomous exog1 exog2 exog5 exog6 exog7)


                       NOW THE FIRST STAGE REGRESSIONS

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  7,   992) =  209.51
       Model |  617.390728     7  88.1986754           Prob > F      =  0.0000
    Residual |  417.608638   992  .420976449           R-squared     =  0.5965
-------------+------------------------------           Adj R-squared =  0.5937
       Total |  1034.99937   999   1.0360354           Root MSE      =  .64883

------------------------------------------------------------------------------
  continuous |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exog3 |   .1584685   .0218622     7.25   0.000     .1155671    .2013699
       exog2 |   -.009669   .0216656    -0.45   0.655    -.0521846    .0328466
       exog1 |   .1599552   .0212605     7.52   0.000     .1182345    .2016759
       exog4 |   .3165751   .0224563    14.10   0.000     .2725079    .3606424
       exog5 |   .4972074    .021356    23.28   0.000     .4552993    .5391156
       exog6 |  -.0780172   .0217546    -3.59   0.000    -.1207076   -.0353268
       exog7 |   .1611768    .022103     7.29   0.000     .1178028    .2045508
       _cons |   .0107516   .0206197     0.52   0.602    -.0297117     .051215
------------------------------------------------------------------------------

Iteration 0:   log likelihood = -692.49904
Iteration 1:   log likelihood = -424.29883
Iteration 2:   log likelihood = -382.05354
Iteration 3:   log likelihood = -377.16723
Iteration 4:   log likelihood = -377.07132
Iteration 5:   log likelihood = -377.07127

Probit estimates                                  Number of obs   =       1000
                                                  LR chi2(7)      =     630.86
                                                  Prob > chi2     =     0.0000
Log likelihood = -377.07127                       Pseudo R2       =     0.4555

------------------------------------------------------------------------------
 dichotomous |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exog3 |   .2134477   .0562479     3.79   0.000     .1032039    .3236916
       exog2 |   .2113067   .0537592     3.93   0.000     .1059406    .3166728
       exog1 |   .4559128    .060367     7.55   0.000     .3375958    .5742299
       exog4 |   .3903133   .0620052     6.29   0.000     .2687852    .5118413
       exog5 |   .7595488   .0646746    11.74   0.000     .6327889    .8863088
       exog6 |   .8546139   .0689585    12.39   0.000     .7194577      .98977
       exog7 |  -.1669142   .0566927    -2.94   0.003    -.2780298   -.0557986
       _cons |   .0835167   .0528104     1.58   0.114    -.0199899    .1870232
------------------------------------------------------------------------------


                 NOW THE SECOND STAGE REGRESSIONS WITH INSTRUMENTS



      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  5,   994) =  141.20
       Model |  429.827896     5  85.9655791           Prob > F      =  0.0000
```

```
    Residual |   605.17147   994  .608824416           R-squared     =  0.4153
-------------+------------------------------           Adj R-squared =  0.4124
       Total |  1034.99937   999   1.0360354           Root MSE      =  .78027

------------------------------------------------------------------------------
  continuous |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
I_dichotom~s |   .2575918   .0214505    12.01   0.000     .2154983    .2996854
       exog3 |   .0425202    .026735     1.59   0.112    -.0099435    .0949838
       exog2 |   .0118544   .0267226     0.44   0.657    -.0405848    .0642937
       exog1 |   .0077736   .0282168     0.28   0.783    -.0475978     .063145
       exog4 |   .3186363   .0283114    11.25   0.000     .2630793    .3741933
       _cons |   .0121851   .0248091     0.49   0.623    -.0364991    .0608692
------------------------------------------------------------------------------

Iteration 0:   log likelihood = -692.49904
Iteration 1:   log likelihood = -424.31527
Iteration 2:   log likelihood =  -382.0779
Iteration 3:   log likelihood = -377.20169
Iteration 4:   log likelihood = -377.10665
Iteration 5:   log likelihood = -377.10661

Probit estimates                                  Number of obs   =       1000
                                                  LR chi2(6)      =     630.78
                                                  Prob > chi2     =     0.0000
Log likelihood = -377.10661                       Pseudo R2       =     0.4554

------------------------------------------------------------------------------
 dichotomous |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
I_continuous |   1.262866   .1604171     7.87   0.000     .9484539    1.577277
       exog1 |   .2509257   .0649992     3.86   0.000     .1235297    .3783218
       exog2 |   .2260372   .0529623     4.27   0.000     .1222331    .3298413
       exog5 |   .1291197   .0958474     1.35   0.178    -.0587377    .3169771
       exog6 |   .9560943   .0721625    13.25   0.000     .8146584     1.09753
       exog7 |  -.3712822   .0674939    -5.50   0.000    -.5035678   -.2389966
       _cons |   .0707977   .0528105     1.34   0.180    -.0327091    .1743044
------------------------------------------------------------------------------


         NOW THE SECOND STAGE REGRESSIONS WITH CORRECTED STANDARD ERRORS


------------------------------------------------------------------------------
  continuous |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
I_dichotom~s |   .2575918   .1043332     2.47   0.014     .0528532    .4623305
       exog3 |   .0425202   .1291476     0.33   0.742     -.210913    .2959533
       exog2 |   .0118544   .1290542     0.09   0.927    -.2413956    .2651044
       exog1 |   .0077736   .1363699     0.06   0.955    -.2598323    .2753795
       exog4 |   .3186363   .1367953     2.33   0.020     .0501956     .587077
       _cons |   .0121851   .1198708     0.10   0.919    -.2230438    .2474139
------------------------------------------------------------------------------
------------------------------------------------------------------------------
 dichotomous |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
I_continuous |   1.262866   .7397385     1.71   0.088    -.1869952    2.712726
       exog1 |   .2509257   .3130259     0.80   0.423    -.3625938    .8644452
       exog2 |   .2260372   .2737467     0.83   0.409    -.3104964    .7625708
       exog5 |   .1291197   .4827168     0.27   0.789    -.8169878    1.075227
       exog6 |   .9560943   .2825678     3.38   0.001     .4022716    1.509917
```

```
       exog7 |  -.3712822   .3265683    -1.14   0.256    -1.011344    .2687799
       _cons |   .0707977   .2666057     0.27   0.791    -.4517399    .5933353
------------------------------------------------------------------------------
```

C. If we face a situation in which $y_1^*$ & $y_2^*$ are observed as follows:[13]

$$\begin{aligned} y_1 &= y_1^* \\ y_2 &= y_2^* \text{ if } y_2^* > 0 \\ y_2 &= 0 \text{ otherwise} \end{aligned}$$

Then we have what Amemiya (1979) calls a simultaneous equation tobit model. Estimation of such models is similar to estimation of 2SPLS, except that instead of probit regressions, tobit regressions are used in the appropriate stages. Again, standard errors are wrong and the correction is a bit detailed. Readers interested in estimating such a model can cannibalize cdsimeq to estimate such a model, since the estimation procedures for both models are very similar. Amemiya (1979) and Maddala (1983) discuss the estimation of these type of models.

D. If we face a situation in which $y_1^*$ & $y_2^*$ are observed as follows:[14]

$$\begin{aligned} y_1 &= 1 \text{ if } y_1^* > 0 \\ y_1 &= 0 \text{ otherwise} \\ y_2 &= 1 \text{ if } y_2^* > 0 \\ y_2 &= 0 \text{ otherwise} \end{aligned} \tag{44}$$

Then we a simultaneous probability model. Maddala (1983) discusses the estimation and correction of the standard errors. This discussion will not be reproduced here.

Finally, there are two other possibilities, Maddala(1983) discusses their estimation, however, he says that the derivation of the covariance matrices is to complicated and he does not discuss them.

# 5 Interpretation

How do we interpret the results and more specificially the coefficients of the instrumental variables? Stay tuned to the discussion during the method lunch.

# 6 Conclusion

Simultaneous relationships are probably a lot more common than is presented in the political science literature. In our work, we should consider such possibilities. I am not arguing that everything simultaneously determines each other

[13] This corresponds to Maddala's (1983, 243–4) model 2.
[14] This corresponds to Maddala's (1983, 246–7) model 6.

(although this could be argued), instead, in our research we should consider such possibilities and at least eliminate them before proceeding with our typical single equation estimation methods.

# References


[1] Alvarez, R. Michael and Garrett Glasgow. 2000. " Two Stage Estimation of Nonrecursive Choice Models," *Political Analysis,* 8(2): 147–165.

[2] Amemiya, Takeshi. 1978. " The Estimation of a Simultaneous Equation Generalized Probit Model," *Econometrica,* 46: 1193–1205.

[3] ——-. 1979. " The Estimation of a Simultaneous–Equation Tobit Model," *International Economic Review,* 20: 169–181.

[4] Gujarati Damodar N. 2003. *Basic Econometrics.* 4th ed. New York: McGraw–Hill, Inc.

[5] Keshk, Omar, Rafael Reuveny, and Brian Pollins. 2002. " Simultaneity Between Bilateral Trade and Militarized Interstate Disputes," Unpublished Manscript.

[6] Maddala, G. S. 1983. *Limited–Dependent and Qualitative Variables in Econometrics.* Cambridge: University Press.