Maximum-likelihood estimation of the limited-dependent variable model with endogenous explanatory variable

Deon Filmer
The World Bank, US
dfilmer@worldbank.org

Michael Lokshin
The World Bank, US
mlokshin@worldbank.org

Abstract. This article describes the `probitiv` command, which implements the maximum likelihood method to estimate the limited-dependent variable model with endogenous variable.

Keywords: endogenous variables, maximum likelihood, limited-depended variables

1 Introduction

In this article we describe the implementation of the maximum likelihood algorithm to estimate the binary-dependent variable model with an endogenous explanatory variable. The econometric problem of estimating a model with a binary dependent variable with a continuous right hand side variable that may be correlated with the error term arises in a variety of settings:

- Jacoby (1997) estimates a model of take-up of a nutrition supplement program in Jamaica as a function of total household expenditures. Because of measurement error and preference heterogeneity the latter are potentially endogenous. The author estimates a two-stage model using variables that proxy household income or wealth are used as instruments for total expenditures.
- Elbadawi and Sambanis (2002) estimate whether an index of democracy has an effect on whether a country experiences an episode of civil war. The authors are concerned that reverse-causality might affect their results. They estimate a two-stage model with two-period lagged measures of democracy as instruments for the current degree of democracy.
- McGranahan (2000) estimates a model for the determinants of charitable giving in 17^{th} century England. The model for whether or not a gift was given to the poor includes the number of family members mentioned in the will—which is found to increase the probability of charitable giving. The author tests whether this is caused by the "endogeneity" that testators that are simply more philanthropic and are simply more likely to care about the poor as well as others by using instruments derived from the parish in which the testator lived.

These three examples use the approach described in Rivers and Vuong (1988) for estimating a two-stage model that yields consistent estimates.[1] Such an approach requires potentially cumbersome adjustments to derive consistent standard-errors. The `probitiv` command implements Full Information Maximum Likelihood to estimate the binary and continuous parts of the model simultaneously which will yield consistent standard-errors. Like that in Rivers and Vuong (1988) the approach relies of joint normality of the error terms in the binary and continuous equations.

2 Methods

Consider the following two equation simultaneous model (this model is similar to Model 4 in Maddala 1983 p. 120):

$$y_{1i}^{*} = \gamma y_{2i} + \beta_1 x_{1i} + \varepsilon_{1i} \qquad (1.1)$$
$$y_{2i} = \beta_2 x_{2i} + \varepsilon_{2i} \qquad (1.2)$$

where

$$y_{1i} = 1 \;\; if \;\; y_{1i}^{*} > 0$$
$$y_{1i} = 0 \;\; otherwise$$

and $\mathbf{x}_1$ and $\mathbf{x}_2$ are vectors of weakly exogenous variables. The identification conditions in this model are that disturbance terms ε_1 and ε_2 are independent, or else there is at least one variable in $\mathbf{x}_1$ that is not included in $\mathbf{x}_2$. Assume that (x1, ε_1, ε_2) are i.i.d, and ε_1 and ε_2 having, conditional on $\mathbf{x}_1$, a joint normal distribution with mean zero and positive definite covariance matrix:

$$\Omega = \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

Given the assumption with respect to the distribution of the disturbance terms, the logarithmic likelihood function for the system of equation (1.1-1.2) is:

$$\ln L = \sum_i y_{1i} w_i (\ln[F(\eta_i)] + \ln[f(\eta_i)]) + (1 - y_{1i}) w_i (\ln[1 - F(\eta_i)] + \ln[f(\eta_i)]) \qquad (2)$$

where F is a cumulative normal distribution function, f is a normal density distribution function, ρ is a correlation between ε_1 and ε_2, w_i is an optional weight for observation i and

$$\eta_i = \frac{(\beta_1 x_{1i} + \rho(y_{2i} - \beta_2 x_{2i}) / \sigma_2)}{\sqrt{1 - \rho^2}}$$

[1] Smith and Blundell (1988) derive a similar estimation procedure for the Tobit model.

In the maximum likelihood estimation, σ and ρ are not directly estimated. Directly estimated are ln σ and atanh ρ:

$$\text{atanh}\ \rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

3 The probitiv command

3.1 Syntax

`probitiv` is implemented as `d2` evaluator that calculates the over log likelihood along with its first and second derivatives. The command allows for weights, robust estimation, as well as the full set of options associated with Stata's maximum likelihood procedures. The generic syntax for the command is:

```
probitiv depvar varlist [weight] [if exp] [in range] ,
instruments(depvar_s = varlist_s)[ robust cluster(varname)
maximize_options]
```

3.2 Options

`instrument(depvar_s=varlist_s)` gives the specification of the equation for Y2. varlist_s consists of the exogenous variables in the system and includes the set of instruments that help identify the model. It is an integral part of specifying the `probitiv` model and is not optional.

`robust` specifies that the Huber/White/sandwich estimator of the variance is to be used in place of the conventional MLE variance estimator. `robust` combined with `cluster()` further allows observations which are not independent within cluster (although they must be independent between clusters). If you specify pweights, robust is implied. See [U] 23.14 Obtaining robust variance estimates.

`cluster(varname)` specifies that the observations are independent across groups (clusters) but not necessarily within groups. `varname` specifies to which group each observation belongs; e.g., `cluster(personid)` in data with repeated observations on individuals. `cluster()` affects the estimated standard errors and variance-covariance matrix of the estimators (VCE), but not the estimated coefficients. `cluster()` can be used with `pweights` to produce estimates for unstratified cluster-sampled data. Specifying `cluster()` implies robust.

`maximize_options` control the maximization process; see maximize. With the possible exception of `iterate(0)` and `trace`, you should only have to specify them if the model is unstable.

4 Example

We will illustrate the use of the `probitiv` command by the problem of estimating the impact of the distance to the nearest primary school on the probability of school enrollment. A typical specification might be the following:

$$(\text{Enrolled})_i = \alpha + \beta \times (\text{Distance to nearest primary school})_i + \delta 1 \times (\text{Age Dummy Variables})_i + \delta 2 \times (\text{Male})_i + \varepsilon_i \quad (1)$$

Where (Enrolled) is observed as a binary variable of whether a child is enrolled in school or not. Other variables in the model might include age and gender of the child.

An objection sometimes raised for such a model is that the distance to the nearest primary school could be "endogenous" because schools might be placed where "they are most needed," that is, where the underlying probability of enrollment is low. This positive correlation between distance and ε means that schools would typically be located close to children who are less likely to enroll, which would result in an upwardly biased estimate of β.

The "Probit Instrumental Variables" approach simultaneously estimates equation (1) with an equation for the distance to the nearest school. For the sake of illustration, consider the use of a set of dummy variables for region of residence as legitimate instruments for distance to the nearest school.[2] This would lead to estimating:

$$(\text{Distance to nearest primary school})_i = \gamma 1 \times (\text{Age Dummy Variables})_i + \gamma 2 \times (\text{Male})_i + \lambda \times (\text{Region Dummy Variables})_i + \mu_i \quad (2)$$

Applying a standard Probit model in a dataset of 6 to 14 year olds for rural areas in a poor African country yields the following:

```
. probit   schin d_prim zage7-zage14 male

Iteration 0:   log likelihood = -2333.6849
Iteration 1:   log likelihood =  -1920.839
Iteration 2:   log likelihood = -1912.5393
Iteration 3:   log likelihood = -1912.4923
Iteration 4:   log likelihood = -1912.4923

Probit estimates                                  Number of obs   =       3379
                                                  LR chi2(10)     =     842.39
                                                  Prob > chi2     =     0.0000
Log likelihood = -1912.4923                       Pseudo R2       =     0.1805

------------------------------------------------------------------------------
       schin |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      d_prim |  -.0573052   .0160758    -3.56   0.000    -.0888132   -.0257973
```

[2] One might argue the validity of this instrument set on the basis that the decision to build schools is made at the regional level, but conditional on the distance to the nearest school the region of residence does not affect enrollment.

```
       zage7 |   .6036007   .1112158     5.43   0.000     .3856217    .8215796
       zage8 |   .9060991   .1058128     8.56   0.000     .6987097    1.113488
       zage9 |   1.275321   .1091719    11.68   0.000     1.061348    1.489294
      zage10 |   1.638281   .1091079    15.02   0.000     1.424433    1.852128
      zage11 |   1.910501   .1137342    16.80   0.000     1.687586    2.133415
      zage12 |    1.91603   .1096594    17.47   0.000     1.701101    2.130958
      zage13 |   2.071103   .1139791    18.17   0.000     1.847708    2.294497
      zage14 |    1.91747   .1127867    17.00   0.000     1.696412    2.138528
        male |  -.3014624   .0474146    -6.36   0.000    -.3943933   -.2085315
       _cons |  -1.075099   .0881272   -12.20   0.000    -1.247825   -.9023729
------------------------------------------------------------------------------
```

That is, the coefficient on distance to the nearest primary school is -.0573 and is significantly different from zero: in this sample a greater distance to school is associated with lower probability of enrollment. The marginal effect of a change in distance on the probability a child is enrolled in school is: dF/dX = -0.023.

Allowing for endogenous school placement and using `probitiv` with region dummy variables as identifying instruments gives the following.

```
. probitiv schin d_prim zage7-zage14 male , instr(d_prim zage7-zage14 male
zregion2-zregion6)

Iteration 0:   log likelihood = -8181.4343
Iteration 1:   log likelihood = -8170.5543
Iteration 2:   log likelihood = -8170.3271
Iteration 3:   log likelihood = -8170.3187
Iteration 4:   log likelihood = -8170.3187

Probit IV                                         Number of obs   =       3379
                                                  Wald chi2(10)   =     894.20
Log likelihood = -8170.3187                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
schin        |
      d_prim |  -.2932207   .0770446    -3.81   0.000    -.4442254    -.142216
       zage7 |   .5572857   .1109739     5.02   0.000     .3397809    .7747904
       zage8 |    .855251   .1091819     7.83   0.000     .6412584    1.069244
       zage9 |   1.165942   .1256413     9.28   0.000     .9196896    1.412194
      zage10 |   1.523753   .1320002    11.54   0.000     1.265037    1.782468
      zage11 |   1.761336    .147004    11.98   0.000     1.473213    2.049459
      zage12 |   1.759638   .1459382    12.06   0.000     1.473604    2.045671
      zage13 |   1.909265    .152936    12.48   0.000     1.609516    2.209014
      zage14 |   1.766795   .1468191    12.03   0.000     1.479035    2.054555
        male |   -.262274   .0502879    -5.22   0.000    -.3608364   -.1637116
       _cons |  -.8699615    .126936    -6.85   0.000    -1.118752   -.6211714
-------------+----------------------------------------------------------------
d_prim       |
       zage7 |   .0310029   .1123406     0.28   0.783    -.1891808    .2511865
       zage8 |   .0631306   .1081325     0.58   0.559    -.1488052    .2750664
       zage9 |  -.0456569   .1141993    -0.40   0.689    -.2694835    .1781697
      zage10 |   .0388507   .1122615     0.35   0.729    -.1811779    .2588792
      zage11 |  -.0142678   .1152455    -0.12   0.901    -.2401449    .2116092
      zage12 |  -.0656893    .110143    -0.60   0.551    -.2815657     .150187
      zage13 |   -.037738    .113324    -0.33   0.739    -.2598491     .184373
      zage14 |  -.0364631   .1140305    -0.32   0.749    -.2599588    .1870327
        male |   .0972279   .0534201     1.82   0.069    -.0074736    .2019293
    zregion2 |   .7926999   .0866229     9.15   0.000     .6229222    .9624776
```

```
    zregion3 |   .0706183   .0814128     0.87   0.386    -.0889478    .2301845
    zregion4 |  -.0192687   .0726982    -0.27   0.791    -.1617546    .1232171
    zregion5 |   .1310172   .1017084     1.29   0.198    -.0683276     .330362
    zregion6 |   .1701614   .0901277     1.89   0.059    -.0064857    .3468085
       _cons |   .3643652   .0900332     4.05   0.000     .1879034     .540827
-------------+----------------------------------------------------------------
      /theta |   .4013036   .1453618     2.76   0.006     .1163997    .6862075
      /lnsig |   .4342984   .0121683    35.69   0.000     .4104491    .4581478
-------------+----------------------------------------------------------------
         rho |   .3810638   .1242539                      .1158769    .5955401
       sigma |    1.54388   .0187863                      1.507495    1.581143
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =     7.49   Prob > chi2 = 0.0062
------------------------------------------------------------------------------
```

The results for Equation (2) are reported in the section of the output with the "d_prim" heading. This is the analogue of the first stage regression in traditional instrumental variables regression.

The results of the model of interest are reported in the section of the output with the "schin" heading. Note that the parameter estimates cannot typically be compared to those of the simple Probit estimates. This is because Probit (and Probit IV) estimates are scaled by the standard error of the error term and this might implicitly differ between the two models. One can, however, compare marginal effects. The marginal effect implied by the `probitiv` estimates is dF/dX = -.117, almost a full order of magnitude different than that implied by the probit estimates.

The correlation coefficient between the two equations (`rho`) equals .3811. The likelihood ratio test for whether this is significantly different from zero is reported in the last line of the output. In this case the p-value is 0.0062. Since `rho` is positive and significantly different from zero the model suggests that schools are located closer to children who are less likely to be enrolled (i.e. both distance to school and ε are small). The significance of this correlation coefficient is a test of the exogeneity of the variable in question (note that the validity of the test rests on the validity of the instruments).

The variables `sigma`, `/theta`, and `/lnsig` are ancillary parameters used in the maximum likelihood procedure. `Sigma` is the square-root of the variance of the residuals of the regression part of the model and `lnsig` is its log. Theta is a transformation of the correlation between the errors from the two equations.

5 References

Elbadawi, Ibrahim and Nicholas Sambinas (2002) "How Much War Will We See? Explaining the Prevalence of Civil War" Journal of Conflict Resolution 46(3): 307-334.

Jacoby, Hanan G. (1997). "Self-Selection and the Redistributive Impact of in-Kind Transfers: An Econometric Analysis" Journal of Human Resources 32(2):233-249

Maddala, G., (1983) Limited-Dependent and Qualitative Variables in Econometric, Econometric Society Monographs No. 3, Cambridge University Press, New York

McGranahan, Leslie (2000) "Charity and the bequests motive: evidence from seventeenth century wills" *The Journal of Political Economy* 108(6):1270-1291.

Rivers, D., and H. Vuong (1988) "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models" Journal of Econometrics, Vol. 39: 347-366

Smith, R., and R. Blundell (1986) "An Exogeneity Test for a Simultaneous Equation Tobit Model with and Application to Labor Supply" Econometrica, Vol. 54(3): 679-685