RWI : Discussion Papers

Colin Vance

Marginal Effects and Significance Testing with Heckman's Sample Selection Model

A Methodological Note

No. 39

RWI
ESSEN

Rheinisch-Westfälisches Institut für Wirtschaftsforschung

RWI : Discussion Papers
No. 39

Colin Vance

Marginal Effects and Significance Testing with Heckman's Sample Selection Model

A Methodological Note

RWI
ESSEN

Colin Vance*

Marginal Effects and Significance Testing with Heckman's Sample Selection Model: A Methodological Note

Abstract

This paper illustrates two techniques for calculating the statistical significance of the marginal effects derived from Heckman's sample selection model, an increasingly common econometric specification in political science. The discussion draws on an analysis by Sweeney (2003) of the incidence and intensity of interstate disputes. After replicating his results, the paper presents the delta method and the nonparametric bootstrap as alternative techniques for obtaining standard errors of the marginal effects, which themselves are calculated from a transformation of the model parameters. The analysis reveals two variables for which misleading inferences are drawn with respect to the precision of the estimated coefficients in the original study, suggesting that significance testing of the derived marginal effects is warranted.

1. Introduction

Empirical research in political science has increasingly used Heckman's sample selection model to analyze data sets in which censoring of the dependent variable raises concerns of selectivity bias. Recent examples include Lebovic's (2004) study of the influence of democracy on the contribution to peace keeping operations and Drury et al.'s (2005) analysis of the amount of U.S. disaster relief assistance. Both studies observe the outcome of interest, in this case foreign aid, only when it is positive, with the remainder of observations censored at zero. This raises the possibility that the sample used for estimation is nonrandom, in turn causing bias through the correlation of the error term with the explanatory variables. Heckman (1979) developed a two-stage estimator to mitigate this bias. In stage one, referred to as the selection equation, a probit model is estimated on the entire data set to capture the determinants of censoring. Stage two, referred to as the outcome equation, involves estimation of a heteroskedasticity-corrected OLS regression on the non-censored observations. To control for potential bias emerging from sample selectivity, this second stage regression appends the inverse Mills ratio calculated from the probit model as an additional regressor.

Although the Heckman model is extensively documented and can be readily estimated using most statistical packages, the interpretation of the coefficient estimates can be tricky, particularly when a variable appears in both the selection and outcome equations. In this case, the marginal effect is not given by the coefficient estimate, itself, but rather must be calculated using a non-linear function of the underlying model parameters to correct for the selectivity effect. Although several papers address how to implement this correction (Saha et al. 1997; Sigelman, Zeng 1999; Hoffmann, Kassouf 2005), they omit any discussion of how to calculate the significance of the marginal effect. Likewise, there is little discussion of this issue in the applied literature, even among the few studies that recognize the need to correct for the marginal effects (Goetz 1992; Vance, Geoghegan 2004). Instead, the standard practice seems to incorrectly rely on the standard error corresponding to the unadjusted coefficient estimate for assessing significance.

The purpose of this note is to offer practical guidance for implementing significance tests with the Heckman model using two alternative techniques: the delta method and the nonparametric bootstrap. To illustrate the techniques, Section 2 draws on a recent study by Sweeney (2003) that investigates the severity of interstate disputes. After replicating his results, Section 3 discusses how to calculate their statistical significance. The analysis of Section 4 reveals the relatively low precision of one of the coefficients deemed to be 'tightly' estimated in the original study and the high precision of another coefficient that is incorrectly discarded as insignificant, suggesting that significance testing of

the marginal effects is warranted. Section 5 concludes. The Appendix provides the code, written in Stata, for implementing the techniques.

2. An Example of Heckman's Sample Selection Model

Sweeney's analysis is primarily concerned with the effects of military capability, interest similarity and their interaction as causes of conflict severity among dyads. To test the significance of these determinants, he draws on the Correlates of War militarized interstate dispute data set, from which he derives a *severity of dispute* measure suggested by Diehl/Goertz (2000) for use as the dependent variable. This variable is censored at zero for cases in which the dispute severity is not sufficiently intense to be observable; otherwise it assumes some positive value as calculated by a weighted combination of factors, including fatalities and the level of hostility. Sweeney specifies Heckman's sample selection model to jointly analyze the incidence of observable conflict, a binary outcome, and its intensity, a continuous outcome.

As several of the variables investigated in the analysis are hypothesized to influence both outcomes, they are included as regressors in the probit and OLS stages of the model. The interpretation of the marginal effects for these variables must consequently be adjusted to correct for selectivity bias. This correction is given by the following equation (Sigelman, Zeng 1999):

$$(1) \qquad \frac{\partial E(Y \mid S^* > 0, X)}{\partial X_k} = \beta_k - \gamma_k \rho \sigma_\varepsilon \delta(-w\gamma)$$

where Y is the dependent variable, S^* is a latent variable denoting selection, β_k and γ_k are the estimated coefficients for X_k in the outcome and selector equations, ρ is the correlation coefficient between the error terms of the selector and outcome equations, σ_ε is the root mean squared error of the outcome equation, and $\delta(-w\gamma)$ is a function of the inverse Mills ratio, obtained from the linear predictions $(-w\gamma)$ of the selector equation.[1]

The application of equation (1) produces a unique value for every observation in the data. Following the recommendation of Sigelman/Zeng (1999), Sweeney averages across the observations to interpret the marginal effect. His results from Model 2 of the paper are reproduced in Table 1. To assess the sen-

---

[1] Note that this formula must be modified for the case of dummy variables. The implementation can be found in the appendix, following the discussion provided by Hoffmann/Kassouf (2005). As these authors note, equation (1) refers specifically to the *conditional* marginal effect, which in this case measures the effect of the variable for those dyads engaged in an observed dispute. They also discuss how to calculate the *unconditional* marginal effect, which would measure the effect for the entire sample of dyads.

Table 1

Sensitivities of Coefficients in Outcome Equation of Model 2

Variable	Marginal effect	Standard deviation	Minimum	Maximum
Capability ratio	165.699	0.060	165.580	166.353
DemocracyL	0.491	0.003	0.486	0.522
DependenceL	−1338.552	1.664	−1341.861	−1320.479
Common IGOs	−0.175	0.001	−0.185	−0.174
Contiguous	10.946	0.100	9.860	−1.289
Log distance	−1.486	0.018	−1.522	15.985
Source: Sweeney 2003.				

sitivity of the estimate, he reports the standard deviation as well as the minimum and maximum values of all the observations.

3. The Delta Method and Nonparametric Bootstrap

Although the approach recommended by Sigelman/Zeng (1999) can be used to gauge the variation in the estimate within the sample, it does not allow the researcher to test the hypothesis that it is statistically significant. The delta method and bootstrapping are two alternative approaches that can be availed for assessing statistical precision. These approaches are commonly applied when the aim is to compute the standard error of a non-linear function for which it is too complex to analytically compute the variance. They allow the researcher to incorporate the uncertainty associated with the multiple parameters that appear in equation (1) and thereby compute confidence intervals for the estimated marginal effects.

The delta method works by using a Taylor series to create a linear approximation of a non-linear function, after which the variance can be computed. Because the method relies on an approximation, its accuracy depends on the degree of linearity of the derivative function at the point that it is evaluated. Following the discussion in Xu/Long (2005), if the function $F(x\beta)$ is taken to be the estimator of interest, the first step of the delta method applies a first-order Taylor expansion to linearize the function evaluated at $\hat{\beta}$:

$$(2) \qquad F(x\hat{\beta}) \approx F(x\beta) + (\hat{\beta} - \beta)f(x\beta)$$

where $f(\beta)$ is the derivative of F evaluated at β. The second step involves taking the variance of both sides of the above equation to yield:

$$(3) \qquad Var[F(x\hat{\beta})] \approx Var[F(x\beta) + (\hat{\beta} - \beta)f(x\beta)] = [f(x\beta)]^2 \, Var(\hat{\beta}).$$

This expression can then be used to compute statistical significance.[2] I illustrate the method using Stata's nlcom command, which is available as of Version 8 and requires fairly rudimentary programming code.

The method of bootstrapping, which is akin to Monte Carlo simulation, can be used as an alternative, albeit computationally more costly, approach to calculating confidence intervals and other measures of statistical precision. This method involves three steps:

1. Draw a large number of samples of size N – with replacement – from the original estimation sample.

2. For each of the drawn samples, reapply the estimator.

3. Derive a probability distribution of the estimator from step two and calculate the standard error.

While bootstrapping relies on relatively few assumptions, it does require that the observed sample distribution is a good estimate of the underlying population distribution (Guan 2003). Hence, the method should be use with caution when the available sample is small. An additional consideration is the number of draws to implement in step one. Convention suggests that 1000 draws is a good starting point. To assess accuracy, the process can be repeated using the same number of draws to check whether the results change meaningfully (King et al. 2000). If so, more draws are probably required, though this decision will largely depend on a subjective assessment of the appropriate precision. In this regard, the major drawback of the method is that it is time consuming. Using a Pentium 4 (3.20 GHz) computer requires several hours to produce the bootstrap statistic for a single coefficient based on 1000 draws from the present dataset.

4. Sweeney (2003) Reconsidered

The results from the application of both methods are presented in Table 2, where it is first noted that the marginal effect estimates are slightly different than those presented by Sweeney. This negligible discrepancy arises because the figures in Table 2 are point estimates which, following the suggestion of Hoffman/Kassouf (2005), are based on the evaluation of the function at the mean of the independent variables. As regards the estimates of significance, two additional points bear noting. First, the p-values calculated using the two methods are roughly in agreement, suggesting that no further validity checks

---

[2] As noted by Papke/Wooldridge (2005), one difficulty in applying the method arises from the complex calculations involved in calculating the confidence intervals for cases with multiple parameters. They describe a computational trick based on a simple substitution that can be used when software is unavailable. This trick involves estimating a linear transformation of the original model using the estimated gradient with respect to the parameters in the function.

Table 2

Significance Levels of Coefficients in Outcome Equation of Model 2

Variable	Marginal effect	P-value: Bootstrap	P-value: Delta method
Capability ratio	165.704	0.021	0.020
DemocracyL	0.492	0.113	0.113
DependenceL	–1338.409	0.000	0.000
Common IGOs	–0.175	0.176	0.165
Contiguous	10.882	0.004	0.005
Log distance	–1.485	0.400	0.397

of accuracy are necessary. Second, their calculation leads in two cases to conclusions that are qualitatively different from those drawn in the original analysis. This applies to the variable DemocracyL, which measures the democracy level of the least democratic state in the dyad. Sweeney interprets this variable as significant based on the unadjusted coefficient in Table 1 of his paper, though it is seen to be insignificant when the uncertainty associated with the selectivity effect is taken into account.[3] Interestingly, it also applies to the coefficient on the dummy variable indicating contiguous states, which Sweeney concludes is insignificant based on the unadjusted coefficient but which is seen to be highly significant based on both the bootstrap and delta method estimates. Contrasting with the conclusion of the original analysis, this estimate thus suggests that geography plays a significant role in determining the severity of conflict. Taken together, the results in Table 2 highlight the potential for drawing misleading inferences on variables included in the outcome and selector equations in the absence of significance testing.

5. Conclusion

This note has illustrated the application of bootstrapping and the delta method for evaluating the significance of the marginal effects estimated from Heckman's sample selection model. As is well-documented in the theoretical literature but rarely implemented in the applied literature, the calculation of these marginal effects must be adjusted to account for sample selection bias. Among the few papers that do implement the adjustment, it is difficult to find any that conduct significance tests of the estimates. While Sweeney explores the range of estimates to assess precision, the conventional approach seems to implicitly rely on the significance of the unadjusted coefficient obtained from the standard software output for hypothesis testing. This reliance is unwarranted. As demonstrated from the present analysis, significant or insignificant effects may be revealed from conducting bootstrapping or the delta method irrespective of the significance of the unadjusted coefficient estimates.

---

[3] He is, however, cautious in interpreting his finding of significance, noting that the Z-statistic obtained from his analysis of the sample variation in the marginal effect is insignificant.

References

Diehl, P. and G. Goertz (2000), *War and peace in international rivalry*. Ann Arbor: University of Michigan Press.

Drury, A.C., R.S. Olson and D.A. Van Belle (2005), The politics of humanitarian aid: U.S. foreign disaster assistance, 1964–1995. *Journal of Politics* 67: 454–473.

Goetz, St.J. (1992), A selectivity model of household food marketing behavior in sub-Saharan Africa. *American Journal of Agricultural Economics* 1992 (May): 444–452.

Guan, W. (2003), From the help desk: Bootstrapped standard errors. *Stata Journal* 3: 71–80.

Heckman, J.J. (1979), Sample selection bias as a specification error. *Econometrica* 47: 153–161.

Hoffmann, R. and A.L. Kassouf (2005), Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman's procedure. *Applied Economics* 37: 1303–1311.

King, G., M. Tomz and J. Wittenberg (2000), Making the most of statistical analyses: Improving Interpretation and presentation. *American Journal of Political Science* 44: 347–361.

Lebovic, J.H. (2004), Uniting for Peace? Democracies and United Nations Peace Operations after the Cold War. *Journal of Conflict Resolution* 48: 910–936.

Papke, L.E. and J.M. Wooldridge (2005), A computational trick for delta-method standard errors. *Economics Letters* 86: 413–417.

Saha, A., O. Capps and P.J. Byrne (1997), Calculating marginal effects in models for zero expenditures in household budgets using Heckman-type correction. *Applied Economics* 29: 1311–1316.

Sigelman, L. and L. Zeng (1999), Analyzing censored and sample-selected data with Tobit and Heckman models. *Political Analysis* 8: 167–182.

Sweeney, K.J. (2003), The severity of interstate disputes: Are dyadic capability preponderances really more pacific? *Journal of Conflict Resolution* 47: 728–750.

Xu, J. and J.S. Long (2005), Confidence intervals for predicted outcomes in regression models for categorical outcomes. Internet: www.indiana.edu/~jslsoc/stata/spostci/xulong-prvalue-23aug2005.pdf, downloaded January 2006.

Vance, C. and J. Geoghegan (2004), Modeling Semi-Subsistence and Commercial Land-Use Decisions in an Agricultural Frontier of Southern Mexico: A Switching Regression Approach. *International Regional Science Review* 27: 326–347.

Appendix

The following code uses the delta method to calculate the standard error for the marginal effects of "smldmat", which is continuous, and "contigkb", which is a dummy. to calculate the marginal effect for the latter, the formula from Hoffmann/Kassouf (2005) is used. Thereafter, the bootstrap method is used to calculate the standard error for "smldmat" using 1000 resamples. This code can be used with Stata 8.2 or Stata 9.

```
#delimit;
set more off;
use "C:\SweeneyJCR.dta", clear;
heckman brl2 ln_capratio INTEREST NEWCAP2 smldmat smldep smigoabi
TERR2 ACTORS2 logdstab contigkb post1945 majmaj, robust select (disputex =
ln_capratio smldmat smldep smigoabi allies contigkb logdstab majpower peace
_spline1 _spline2 _spline3);
keep if e(sample);
sum ln_capratio;
local bln_capratio1=r(mean);
sum smldmat;
local bsmldmat1=r(mean);
sum smldep;
local bsmldep1 =r(mean);
sum smigoabi;
local bsmigoabi1 =r(mean);
sum allies;
local ballies1 =r(mean);
sum contigkb;
local bcontigkb1 =r(mean);
sum logdstab;
local blogdstab1 =r(mean);
sum majpower;
local bmajpower1 =r(mean);
sum peace;
local bpeace1 =r(mean);
sum _spline1;
local b_spline11 =r(mean);
sum _spline2;
local b_spline21 =r(mean);
sum _spline3;
local b_spline31 =r(mean);
local selxbpr1 "[#2]_b[ln_capratio]*`bln_capratio1'+[#2]_b[smldmat]*
`bsmldmat1'+[#2]_b[smldep]*`bsmldep1'+ [#2]_b[smigoabi]*`bsmigoabi1'+
[#2]_b[allies]*`ballies1'+ [#2]_b[contigkb]*`bcontigkb1'+ [#2]_b[logdstab]*
`blogdstab1'+ [#2]_b[majpower]*`bmajpower1'+ [#2]_b[peace]*`bpeace1'+
[#2]_b[_spline1]*`b_spline11'+ [#2]_b[_spline2]*`b_spline21'+ [#2]_b[_spline3]*
`b_spline31'+ [#2]_b[_cons]";
local contig1 "[#2]_b[ln_capratio]*`bln_capratio1'+[#2]_b[smldmat]*
`bsmldmat1'+[#2]_b[smldep]*`bsmldep1'+ [#2]_b[smigoabi]*`bsmigoabi1'+
[#2]_b[allies]*`ballies1'+ [#2]_b[contigkb]*1 + [#2]_b[logdstab]*`blogdstab1'+
[#2]_b[majpower]*`bmajpower1'+ [#2]_b[peace]*`bpeace1'+ [#2]_b[_spline1]*
`b_spline11'+ [#2]_b[_spline2]*`b_spline21'+ [#2]_b[_spline3]*`b_spline31'+
[#2]_b[_cons]";
local contigkb1 "normden(`contig1')/norm(`contig1')";
```

```
local contig0 "[#2]_b[ln_capratio]*`bln_capratio1'+[#2]_b[smldmat]*
`bsmldmat1'+[#2]_b[smldep]* `bsmldep1'+ [#2]_b[smigoabi]*`bsmigoabi1'+
[#2]_b[allies]*`ballies1'+ [#2]_b[contigkb]*`0 + [#2]_b[logdstab]*`blogdstab1'+
[#2]_b[majpower]*`bmajpower1'+ [#2]_b[peace]*`bpeace1'+ [#2]_b[_spline1]*
`b_spline11'+ [#2]_b[_spline2]*`b_spline21'+ [#2]_b[_spline3]*`b_spline31'+
[#2]_b[_cons]";
local contigkb0 "normden(`contig0')/norm(`contig0')";
local testpr1 "normden(`selxbpr1')/norm(`selxbpr1')";
local D1 "(`testpr1')*(`testpr1'+`selxbpr1')";
local simrho1 "(exp(2*_b[/athrho])-1)/(exp(2*_b[/athrho])+1)";
local simSEE1 "exp(_b[/lnsigma])";
```

/*Now implement the delta method. Apply formula from Sigelman/Lee
(1999) for continuous variables.*/

```
local smldmat1 _b[smldmat]-([#2]_b[smldmat] *(`simrho1')*(`simSEE1')*(`D1'));
nlcom `smldmat1';
```

/*Apply formula from Hoffmann/Kassouf (2005) for dummy varialbles*/

```
local b_contig _b[contigkb] + (`simrho1')*(`simSEE1')*(`contigkb1'-`contigkb0');
nlcom `b_contig';
```

/*Code for calculating bootstrap standard error of the variable "smldmat"*/

```
use "C:\SweeneyJCR.dta", clear;
cap program drop myse;
program myse, rclass;
```

/*Table One, Model 2
all dyads, all years, full Interest Similarity Measure*/

```
heckman brl2 ln_capratio INTEREST NEWCAP2 smldmat smldep smigoabi
TERR2 ACTORS2 logdstab contigkb post1945 majmaj, robust select (disputex =
ln_capratio smldmat smldep smigoabi allies contigkb logdstab majpower peace
_spline1 _spline2 _spline3);
keep if e(sample);
local ln_capratio1=X[1,1];
local smldmat1 =X[1,2];
local smldep1 =X[1,3];
local smigoabi1 =X[1,4];
local allies1 =X[1,5];
local contigkb1 =X[1,6];
local logdstab1 =X[1,7];
local majpower1 =X[1,8];
local peace1 =X[1,9];
local _spline11=X[1,10];
local _spline21=X[1,11];
local _spline31=X[1,12];
local selxbpr1
"[#2]_b[ln_capratio]*`ln_capratio1'+[#2]_b[smldmat]*`smldmat1'+[#2]_b[smldep]
*`smldep1'+ [#2]_b[smigoabi]*`smigoabi1'+ [#2]_b[allies]*`allies1'+
[#2]_b[contigkb]*`contigkb1'+ [#2]_b[logdstab]*`logdstab1'+ [#2]_b[majpower]*
`majpower1'+ [#2]_b[peace]*`peace1'+ [#2]_b[_spline1]*`_spline11'+
[#2]_b[_spline2]*`_spline21'+ [#2]_b[_spline3]*`_spline31'+ [#2]_b[_cons]";
```

```
local testpr1 "normden(`selxbpr1')/norm(`selxbpr1')";
local D1 "(`testpr1')*(`testpr1'+`selxbpr1')";
local simrho1 "(exp(2*_b[/athrho])-1)/(exp(2*_b[/athrho])+1)";
local simSEE1 "exp(_b[/lnsigma])";
return scalar b_smldmat= _b[smldmat]-([#2]_b[smldmat] *(`simrho1')*
(`simSEE1')*(`D1'));
end;
clear;
use "C:\SweeneyJCR_temp.dta", clear;
heckman brl2 ln_capratio INTEREST NEWCAP2 smldmat smldep smigoabi
TERR2 ACTORS2 logdstab contigkb post1945 majmaj, robust select (disputex =
ln_capratio smldmat smldep smigoabi allies contigkb logdstab majpower peace
_spline1 _spline2 _spline3);
keep if e(sample);
sum ln_capratio;
local bln_capratio1=r(mean);
sum smldmat;
local bsmldmat1=r(mean);
sum smldep;
local bsmldep1 =r(mean);
sum smigoabi;
local bsmigoabi1 =r(mean);
sum allies;
local ballies1 =r(mean);
sum contigkb;
local bcontigkb1 =r(mean);
sum logdstab;
local blogdstab1 =r(mean);
sum majpower;
local bmajpower1 =r(mean);
sum peace;
local bpeace1 =r(mean);
sum _spline1;
local b_spline11 =r(mean);
sum _spline2;
local b_spline21 =r(mean);
sum _spline3;
local b_spline31 =r(mean);
mat X=(`bln_capratio1',`bsmldmat1',`bsmldep1',`bsmigoabi1',`ballies1',
`bcontigkb1',`blogdstab1',`bmajpower1',`bpeace1',`b_spline11',`b_spline21',
`b_spline31');
local selxbpr1 "[#2]_b[ln_capratio]*`bln_capratio1'+[#2]_b[smldmat]*
`bsmldmat1'+[#2]_b[smldep]* `bsmldep1'+ [#2]_b[smigoabi]*`bsmigoabi1'+
[#2]_b[allies]*`ballies1'+ [#2]_b[contigkb]*`bcontigkb1'+ [#2]_b[logdstab]*
`blogdstab1'+ [#2]_b[majpower]*`bmajpower1'+ [#2]_b[peace]*`bpeace1'+
[#2]_b[_spline1]*`b_spline11'+ [#2]_b[_spline2]*`b_spline21'+ [#2]_b[_spline3]*
`b_spline31'+ [#2]_b[_cons]";
local testpr1 "normden(`selxbpr1')/norm(`selxbpr1')";
local D1 "(`testpr1')*(`testpr1'+`selxbpr1')";
local simrho1 "(exp(2*_b[/athrho])-1)/(exp(2*_b[/athrho])+1)";
local simSEE1 "exp(_b[/lnsigma])";
```

/*Bootstrap*/

```
bootstrap "myse" b_smldmat =r(b_smldmat), reps(1000);
```