The Stata Journal (2021)
21, Number 4, pp. 972–998 DOI: 10.1177/1536867X211063149

On identification and estimation of Heckman models

Jonathan Cook
Division of Research and Statistics
Oeste Corp
Washington, DC
jacook@uci.edu

Joon-Suk Lee
joonsuk.lee@outlook.com

Noah Newberger
Division of Research and Statistics
Oeste Corp
Washington, DC
noahnewberger@gmail.com

Abstract. In this article, we present commands to enable fixing the value of the correlation between the unobservables in Heckman models. These commands can solve two practical issues. First, for situations in which a valid exclusion restriction is not available, these commands enable exploring how the results could be affected by sample-selection bias. Second, stepping through values of this correlation can verify whether the global maximum of the likelihood function has been found. We provide several commands to fit these and related models with a fixed value of the correlation between the unobservables.

Keywords: st0658, heckman_fixedrho, heckman_scanrho, heckprobit_fixedrho, heckprobit_scanrho, etregress_fixedrho, etregress_scanrho, biprobit_fixedrho, biprobit_scanrho, Heckman model, sample-selection correction, endogenous treatment, bivariate probit

1 Introduction

Heckman's (1976, 1978) work on sample-selection and endogenous treatment models has been widely used in applied research. Convincing identification of these models requires an exclusion restriction (often referred to as an "instrument"), that is, a variable that affects selection or treatment but not the outcome directly. For many applications, a valid exclusion restriction is not available. Without a valid exclusion, identification of the model is only possible through the distributional assumptions placed on the model.

One approach to fitting a Heckman model without an exclusion restriction is to fix the value of the correlation between the unobservables in the selection and outcome equations (which we refer to as "*rho*") at multiple plausible values. The idea is to treat *rho* as if it were unidentified rather than to identify it based on distributional assumptions. We can then see how the results are affected by the value of *rho*. As we discuss in the next section, we can think of *rho* as the degree of sample-selection bias. This approach was taken for endogenous treatment and sample-selection models by Altonji, Elder, and Taber (2005) and Chan and Cook (2020), respectively. Altonji, Elder, and Taber also propose using the correlation in observable characteristics to bound the possible values of the correlation in unobservable characteristics, which we do not discuss in this article.

We provide commands that enable fixing the value of *rho* and which are based on Altonji, Elder, and Taber (2005) and Chan and Cook (2020). This article's commands have been used to address concerns about sample-selection bias when a valid exclusion is not available (as in Aobdia [2019]; Choudhary, Merkley, and Schipper [2019]; Downey, Bedard, and Boland [2020]; and Tran and Dinh [Forthcoming]).

Another benefit of fixing the value of *rho* is improved ease of maximizing the likelihood function. The likelihood functions for these models are known to be difficult to maximize. Zuehlke (2017) reports several instances in which authors have reported estimates that are not the global maximum of the likelihood function. The likelihood function is not globally concave, but if the value of *rho* is fixed, the likelihood is concave in the remaining parameters. Olsen (1982) suggests stepping through values of *rho* to find the global maximum of the likelihood function. Zuehlke notes that standard statistical software does not provide the option of maximizing the likelihood function in this manner. Rodemeier (2020) uses this article's commands for this purpose.

We provide an example in section 4 of how to use these commands to step through values of *rho* to maximize the likelihood function. For this example, Stata's `heckman` command does not provide the actual maximum-likelihood estimate unless the initial values are set to a neighborhood of the correct values.

The next section reviews Heckman's sample-selection model and discusses the effect of fixing *rho*. Section 3 provides the syntax for the commands that we are introducing. In this section, we also discuss related models, including endogenous treatment models. Examples are provided in section 4. Section 5 concludes.

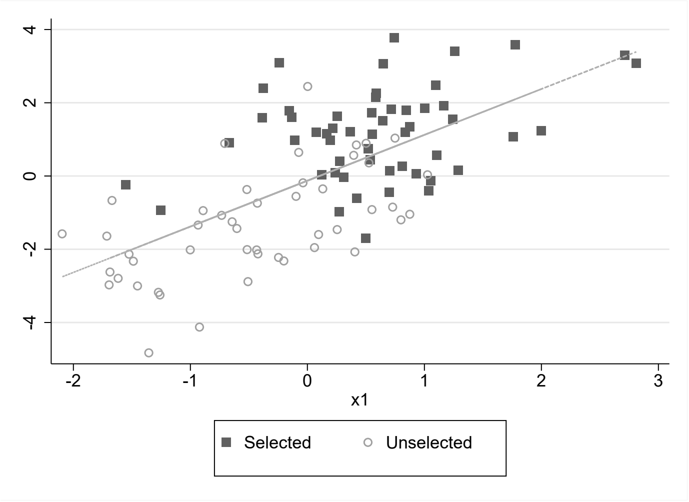
2 Heckman's sample-selection model

When the observations used for a regression are nonrandomly selected, there is a concern that selection could affect the results. In the classic example of regressing education on wage for married women, there is a concern that the decision to work may be affected by both education and unobserved factors.[1] The women with lower education who enter the workforce are those who expect higher wages. This self-selection can overstate the average wage for women with low education and bias the relationship between education and wage downward relative to the true causal effect.

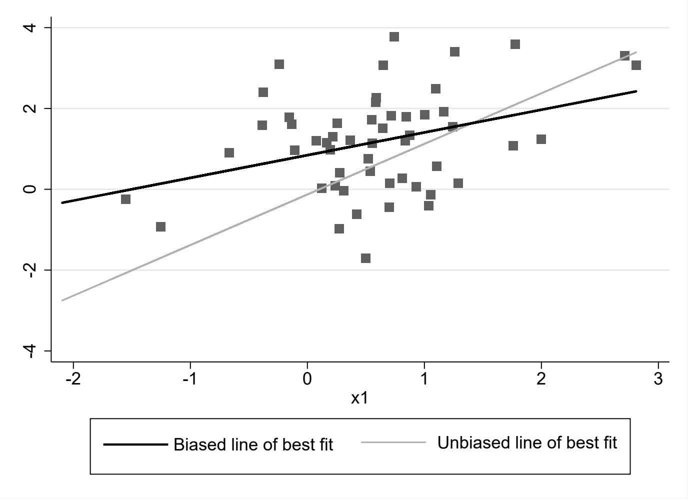
We illustrate this effect in figure 1. Figure 1a presents a scatterplot and line of best fit for all observations (both observed and unobserved). A circle denotes that an instance was unobserved. Figure 1b removes the unobserved instances and adds a new

1. It should be noted that there is concern for a similar bias among married men. Blundell and Powell (2004) find that married men with lower education are less likely to participate in the labor force.

line of best fit based on only the observed instances. The line of best fit from figure 1a is also presented for reference. Selection has decreased the slope of the line of best fit. An important point is that sample selection is a problem of bias, not of generalization to a wider population.



(a) All observations



(b) Only selected observations

Figure 1. Illustration of the effect of sample selection

The approach developed by Heckman (1976) to combat these concerns is to jointly estimate the selection and outcome processes. We begin by defining latent outcome and selection variables, y_i^* and s_i^*, as

$$y_i^* = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i \tag{1}$$

$$s_i^* = \mathbf{z}_i \boldsymbol{\gamma} + u_i \tag{2}$$

where

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\,\sigma \\ \rho\,\sigma & 1 \end{bmatrix} \right)$$

We refer to (1) as the "outcome equation" and (2) as the "selection equation". We are assuming a linear relationship between the latent variables and the regressors $\mathbf{x}_i = (1, x_{1,i}, \ldots, x_{k,i})$ and $\mathbf{z}_i = (1, z_{1,i}, \ldots, z_{r,i})$. The coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are column vectors. The correlation between ε_i and u_i, denoted as ρ, is the same *rho* that was discussed in the introduction. For the remainder of this section, we will use the Greek letter ρ rather than writing out *rho*.

We observe an indicator for selection, s_i, and outcome, y_i, defined as

$$s_i = \begin{cases} 1 & \text{if } s_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$y_i = \begin{cases} y_i^* & \text{if } s_i = 1 \\ 0 & \text{if } s_i = 0 \end{cases}$$

The value of y_i when $s_i = 0$ is arbitrary and unimportant because these values will not be used for estimation. The parameters of interest ($\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, σ, and ρ) can be estimated by maximizing the likelihood function

$$L = \prod_i \Phi\{-(\boldsymbol{\gamma}\,\mathbf{z}_i)\}^{\mathbb{1}(s_i=0)} \times \left(\Phi \left[\frac{\{\boldsymbol{\gamma}\,\mathbf{z}_i + \rho(y_i - \boldsymbol{\beta}\mathbf{x}_i)/\sigma\}}{\sqrt{(1-\rho^2)}} \right] \right)^{\mathbb{1}(s_i=1)}$$
$$\times \left[(2\pi\sigma)^{-1} \exp\{-(\boldsymbol{\beta}\,\mathbf{x}_i - y_i)^2/(2\sigma^2)\}\right]^{\mathbb{1}(s_i=1)}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions and $\mathbb{1}(A)$ is the indicator function, which takes a value of 1 if the event A is true and a value of 0 otherwise. Stata's `heckman` command performs this maximum likelihood estimation as well as the popular "two-step" estimator of this model. We do not discuss the two-step estimator here, because our approach of fixing the value of ρ is less straightforward with the two-step estimator.

2.1 Fixing the value of ρ

For the observed data, the expected value of the outcome is

$$\begin{aligned} E(y_i \,|\, \mathbf{x}_i, s_i = 1) &= \mathbf{x}_i \boldsymbol{\beta} + E(\varepsilon_i \,|\, s_i = 1) \\ &= \mathbf{x}_i \boldsymbol{\beta} + \rho\,\sigma\,\phi(\mathbf{z}_i \boldsymbol{\gamma})/\Phi(\mathbf{z}_i \boldsymbol{\gamma}) \end{aligned}$$

The term $\phi(\cdot)/\Phi(\cdot)$, known as the inverse Mills ratio, follows from the bivariate normality assumption that was placed on ε_i and u_i. An insight from Heckman (1979) is that sample-selection bias can be thought of as an omitted-variable bias, where the omitted variable is the inverse Mills ratio. Naively regressing y_i on $\mathbf{x}_i$ results in a bias of

$$\rho\,\sigma\{\mathrm{Var}(\mathbf{x})\}^{-1}\mathrm{Cov}\{\mathbf{x}, \phi(\mathbf{z}\,\boldsymbol{\gamma})/\Phi(\mathbf{z}\,\boldsymbol{\gamma})\}$$

where $\mathbf{x}$ and $\mathbf{z}$ are the usual matrices of regressors. This bias is increasing in the value of ρ. When ρ equals 0, there is no bias for the naive estimate. It is in this sense that we can think of ρ as the degree of sample-selection bias.

We first discuss the role of fixing the value of ρ on identification, and then we turn to the problem of maximizing the likelihood function. For discussing identification, it is useful to contrast this model to semiparametric sample-selection models. The inconsistency of Heckman's estimator in the presence of nonnormal errors (as shown by Arabmazar and Schmidt [1982] and Robinson [1982]) inspired the creation of several semiparametric estimators. These estimators relax the bivariate normality assumption to accommodate a broader class of bivariate distributions. While Heckman's parametric estimator is identified even when $\mathbf{x}$ and $\mathbf{z}$ contain the same variables (that is, there is no exclusion restriction), these semiparametric estimators require an exclusion restriction for identification. A common finding in Monte Carlo experiments is that Heckman's estimator performs surprisingly well with nonnormal errors as long as there is a valid exclusion restriction (see, for example, Cook and Siddiqui [2020]). This finding has contributed to the widespread use of Heckman's model when there is a valid exclusion restriction.

To be clear, throughout this article, we refer to the exclusion restriction as "valid", meaning that the excluded variable or variables do not affect the outcome directly. It may be tempting to simply omit a variable from the outcome equation so that the model has an excluded variable. This approach, however, can result in estimates that are worse than those obtained by ordinary least squares (Wolfolds and Siegel 2019).[2]

The assumption that the unobservables follow a bivariate normal distribution, which is required for identification without an exclusion restriction, is generally untestable without a valid exclusion restriction. To identify the distribution of ε_i, we need to identify $\boldsymbol{\beta}$. The coefficients $\boldsymbol{\beta}$ could be estimated using a semiparametric estimator, but these estimators require a valid exclusion restriction.

Regarding finding the maximum of the likelihood function, once the value of ρ is fixed, Stata can easily find the values of the other parameters that maximize the likeli-

2. As a bit of an aside, applied researchers may wonder if Heckman models are susceptible to a problem that is analogous to the weak-instrument problem for instrumental variables. (Certo et al. [2016] find that some authors conflate exclusion restrictions for Heckman models with instruments for instrumental variables.) The short answer is "no" because Heckman models are based on structural assumptions. As long as the assumed structure is correct, which includes the assumption that the unobservables follow a bivariate normal distribution, an exclusion is not required for identification. In this case, a weak excluded variable is not a problem for identification because no exclusion is required; however, there may be multicollinearity issues if the coefficient on the excluded variable is small in magnitude.

hood function. If we step through different values of ρ, we can see where the likelihood function is maximized. This can be used to verify that the results returned by `heckman` are the true maximum and to set the initial values if Stata is not returning the global maximum.

Having discussed both a sensitivity test and a method for finding the maximum of the likelihood function, there are two points that should be made explicit. First, one should not step through values of *rho* to maximize the likelihood function when a valid exclusion restriction is not available. While exploring how the results differ for different values of ρ illustrates how sensitive the results are to the degree of sample-selection bias, this procedure cannot be used to reliably estimate the value of ρ when there is no exclusion restriction. Second, for situations in which an exclusion restriction is available, to get the correct standard errors and *p*-values, Stata's `heckman` command should be called with initial values set near the global maximum. The standard errors and *p*-values that are found with a fixed value of ρ may differ from those found without the value of ρ fixed.

3 Syntax

This section provides the syntax for the eight commands that we are introducing: `heckman_fixedrho`, `heckman_scanrho`, `heckprobit_fixedrho`, `heckprobit_scanrho`, `etregress_fixedrho`, `etregress_scanrho`, `biprobit_fixedrho`, and `biprobit_scanrho`. Each is named after the command upon which it was based with the added suffix `_fixedrho` or `_scanrho`. An important difference between these commands and their built-in counterparts is that these commands do not offer all the options that the other commands do. There are also some slight syntax differences that we highlight below.

In this section, we also provide a brief overview about the relevant models that are fit with the Stata commands `heckprobit`, `etregress`, and `biprobit`. References are provided for any reader seeking more details.

3.1 Heckman's sample-selection model

Syntax for heckman_fixedrho

We now present the syntax for the first command, `heckman_fixedrho`, which is for setting the value of *rho*.

`heckman_fixedrho` *depvar* [*indepvars*] [*if*] [*in*],
  `select(`*depvar_s* = *varlist_s* [, `offset(`*varname*`)` `noconstant`]`)`
  `rho(#)` [`vce(`*vcetype*`)` `level(#)` *maximize_options*]

Options for heckman_fixedrho

`select(`*depvar_s* `=` *varlist_s* [`, offset(`*varname*`) noconstant`]`)` specifies the selection equation. `select()` is required.

depvar_s should be coded as 0 or 1, with 0 indicating an observation not selected and 1 indicating a selected observation.

`rho(`#`)` specifies the correlation between the unobservables in the selection and outcome equations. `rho()` is required and must take a value between −1 and 1.

`vce(`*vcetype*`)` specifies the type of standard errors to be used for the estimates. *vcetype* may be `oim`, `robust`, `cluster(`*clustvar*`)`, `opg`, `bootstrap`, or `jackknife`.

`level(`#`)` sets the confidence level. The default is `level(95)` or as set by `set level`.

maximize_options control the maximization process. Options include `difficult`, [`no`]`log`, `trace`, `gradient`, `showstep`, `hessian`, `showtolerance`, `tolerance(`#`)`, `ltolerance(`#`)`, `nrtolerance(`#`)`, and `nonrtolerance`. These options are seldom used.

Syntax for heckman_scanrho

We now present the syntax for the next command, `heckman_scanrho`, which is for scanning through values of ρ.

`heckman_scanrho` *depvar* [*indepvars*] [*if*] [*in*]`,`
`select(`*depvar_s* `=` *varlist_s* [`, offset(`*varname*`) noconstant`]`)`
[`minrho(`#`) maxrho(`#`) step(`#`) vce(`*vcetype*`) level(`#`) nograph`
maximize_options]

Options for heckman_scanrho

`select(`*depvar_s* `=` *varlist_s* [`, offset(`*varname*`) noconstant`]`)` specifies the selection equation. `select()` is required.

depvar_s should be coded as 0 or 1, with 0 indicating an observation not selected and 1 indicating a selected observation.

`minrho(`#`)` specifies the minimum value of correlation between the unobservables in the selection and outcome equations to be considered. It must take a value between −1 and 1. Note that convergence may be difficult at values of −1 and 1. The default is `minrho(-0.9)`.

`maxrho(`#`)` specifies the maximum value of correlation between the unobservables in the selection and outcome equations to be considered. It must take a value between −1 and 1. Note that convergence may be difficult at values of −1 and 1. The default is `maxrho(0.9)`.

`step(#)` specifies the size of the step to use when scanning over values of correlation. This procedure will take a long time to run when the step size is small. The default is `step(0.01)`.

`vce(`*vcetype*`)` specifies the type of standard errors to be used for the estimates. *vcetype* may be `oim`, `robust`, `cluster(`*clustvar*`)`, `opg`, `bootstrap`, or `jackknife`.

`level(#)` sets the confidence level. The default is `level(95)` or as set by `set level`.

`nograph` suppresses the graphical output.

maximize_options control the maximization process. Options include `difficult`, `[no]log`, `trace`, `gradient`, `showstep`, `hessian`, `showtolerance`, `tolerance(#)`, `ltolerance(#)`, `nrtolerance(#)`, and `nonrtolerance`. These options are seldom used.

We now turn to discussing related models that can be thought of as extensions to the sample-selection model discussed above. Our discussion of each is brief, but we provide references for the reader wishing to gain more information.

3.2 Bivariate probit with sample selection

For an outcome that is binary instead of continuous, it is straightforward to extend the model above (as was done by Van de Ven and Van Praag [1981]). We maintain the latent variables in (1) and (2) and the selection indicator in (3):

$$\begin{aligned} y_i^* &= \mathbf{x}_i\,\boldsymbol{\beta} + \varepsilon_i \\ s_i^* &= \mathbf{z}_i\,\boldsymbol{\gamma} + u_i \\ s_i &= \begin{cases} 1 & \text{if } s_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

But now we define the observed outcome as

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \text{ and } s_i = 1 \\ 0 & \text{if } y_i^* \leq 0 \text{ and } s_i = 1 \\ 0 & \text{if } s_i = 0 \end{cases}$$

The likelihood function is

$$L = \prod_i \Phi_2(\mathbf{z}_i\,\gamma,\, \mathbf{x}_i\,\boldsymbol{\beta}; \rho)^{\mathbb{1}(y_i=1)} \times \Phi_2(\mathbf{z}_i\,\boldsymbol{\gamma},\, -\mathbf{x}_i\,\boldsymbol{\beta}; -\rho)^{\mathbb{1}(y_i=0,\, s_i=1)} \times \Phi(-\mathbf{z}_i\,\boldsymbol{\gamma})^{\mathbb{1}(s_i=0)}$$

Syntax for heckprobit_fixedrho

In Stata, the command `heckprobit` fits this model. We now present the syntax for `heckprobit_fixedrho`, which can be used to set the value of ρ in this model.

`heckprobit_fixedrho` *depvar* [*indepvars*] [*if*] [*in*] `,`
  `select(`*depvar_s* `=` *varlist_s* [`,` `offset(`*varname*`)` `noconstant`]`)`
  `rho(`#`)` [`vce(`*vcetype*`)` `level(`#`)` *maximize_options*]

Options for heckprobit_fixedrho

`select(`*depvar_s* `=` *varlist_s* [`,` `offset(`*varname*`)` `noconstant`]`)` specifies the selection equation. `select()` is required.

depvar_s should be coded as 0 or 1, with 0 indicating an observation not selected and 1 indicating a selected observation.

`rho(`#`)` specifies the correlation between the unobservables in the selection and outcome equations. `rho()` is required and must take a value between −1 and 1.

`vce(`*vcetype*`)` specifies the type of standard errors to be used for the estimates. *vcetype* may be `oim`, `robust`, `cluster(`*clustvar*`)`, `opg`, `bootstrap`, or `jackknife`.

`level(`#`)` sets the confidence level. The default is `level(95)` or as set by `set level`.

maximize_options control the maximization process. Options include `difficult`, [`no`]`log`, `trace`, `gradient`, `showstep`, `hessian`, `showtolerance`, `tolerance(`#`)`, `ltolerance(`#`)`, `nrtolerance(`#`)`, and `nonrtolerance`. These options are seldom used.

Syntax for heckprobit_scanrho

We now present `heckprobit_scanrho`, which can be used to scan through values of ρ.

`heckprobit_scanrho` *depvar* [*indepvars*] [*if*] [*in*] `,`
  `select(`*depvar_s* `=` *varlist_s* [`,` `offset(`*varname*`)` `noconstant`]`)`
  [`minrho(`#`)` `maxrho(`#`)` `step(`#`)` `vce(`*vcetype*`)` `level(`#`)` `nograph`
  maximize_options]

Options for heckprobit_scanrho

`select(`*depvar_s* `=` *varlist_s* [`,` `offset(`*varname*`)` `noconstant`]`)` specifies the selection equation. `select()` is required.

depvar_s should be coded as 0 or 1, with 0 indicating an observation not selected and 1 indicating a selected observation.

`minrho(#)` specifies the minimum value of correlation between the unobservables in the selection and outcome equations to be considered. It must take a value between −1 and 1. Note that convergence may be difficult at values of −1 and 1. The default is `minrho(-0.9)`.

`maxrho(#)` specifies the maximum value of correlation between the unobservables in the selection and outcome equations to be considered. It must take a value between −1 and 1. Note that convergence may be difficult at values of −1 and 1. The default is `maxrho(0.9)`.

`step(#)` specifies the size of the step to use when scanning over values of correlation. This procedure will take a long time to run when the step size is small. The default is `step(0.01)`.

`vce(`*vcetype*`)` specifies the type of standard errors to be used for the estimates. *vcetype* may be `oim`, `robust`, `cluster(`*clustvar*`)`, `opg`, `bootstrap`, or `jackknife`.

`level(#)` sets the confidence level. The default is `level(95)` or as set by `set level`.

`nograph` suppresses the graphical output.

maximize_options control the maximization process. Options include `difficult`, [`no`]`log`, `trace`, `gradient`, `showstep`, `hessian`, `showtolerance`, `tolerance(#)`, `ltolerance(#)`, `nrtolerance(#)`, and `nonrtolerance`. These options are seldom used.

3.3 Endogenous binary regressors

Heckman (1978) tackles the problem of endogenous binary regressors using a similar strategy as that of sample selection. The problem is stated in terms of the observed variables:

$$y_i = \mathbf{x}_i\,\boldsymbol{\beta} + d_i\,\delta + \varepsilon_i \tag{4}$$

$$d_i = \begin{cases} 1 & \text{if } \mathbf{z}_i\,\boldsymbol{\gamma} + u_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\,\sigma \\ \rho\,\sigma & 1 \end{bmatrix}\right)$$

The likelihood function can be expressed as

$$L = \left(\Phi\left[\frac{\{\mathbf{z}_i\,\boldsymbol{\gamma} + \rho(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\}}{\sqrt{(1-\rho^2)}}\right]\right)^{\mathbb{1}(d_i=1)}$$
$$\times\left(1-\Phi\left[\frac{\{\mathbf{z}_i\,\boldsymbol{\gamma} + \rho(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\}}{\sqrt{(1-\rho^2)}}\right]\right)^{\mathbb{1}(d_i=0)}$$
$$\times\left[(2\pi\sigma)^{-1}\exp\{-(\mathbf{x}_i\boldsymbol{\beta} - y_i)^2/(2\sigma^2)\}\right]$$

This problem can also be expressed in the context of the Neyman–Rubin potential-outcomes framework.

In the potential-outcomes framework, those receiving a treatment have the outcome

$$y_i = \mathbf{x}_i\,\boldsymbol{\beta} + \delta + \epsilon_{1,i}$$

whereas those not receiving the treatment have the outcome

$$y_i = \mathbf{x}_i\,\boldsymbol{\beta} + \epsilon_{0,i}$$

The parameter δ is the average treatment effect after removing the confounding effects of treatment assignment, which Heckman (1990, 314) calls the "experimental treatment effect". In this potential-outcomes framework, the outcome and treatment can still be expressed as in (4) and (5), but the error term in (4) is defined as

$$\varepsilon_i = d_i\,\epsilon_{1,i} + (1-d_i)\,\epsilon_{0,i}$$

The variances of $\epsilon_{1,i}$ and $\epsilon_{0,i}$ and their correlations with u_i may differ. In Stata, the command `etregress` can fit this model with and without allowing for potentially different variances and correlations for the unobservables for the treated and untreated.

Syntax for etregress_fixedrho

The syntax for our command `etregress_fixedrho` is as follows:

`etregress_fixedrho` *depvar* [*indepvars*] [*if*] [*in*] `, treat(`*depvar_s* `=` *varlist_s*`)`
`rho(`*#*`)` [`poutcomes vce(`*vcetype*`) level(`*#*`)` *maximize_options*]

Options for etregress_fixedrho

`treat(`*depvar_s* `=` *varlist_s*`)` specifies the treatment equation. `treat()` is required.

depvar_s should be coded as 0 or 1, with 0 indicating an observation not selected and 1 indicating a selected observation.

`rho(`*#*`)` specifies the correlation between the unobservables in the selection and outcome equations. `rho()` is required and must take a value between −1 and 1.

poutcomes uses a potential-outcomes model with separate treatment and control group variances.

vce(*vcetype*) specifies the type of standard errors to be used for the estimates. *vcetype* may be oim, robust, cluster(*clustvar*), opg, bootstrap, or jackknife.

level(#) sets the confidence level. The default is level(95) or as set by set level.

maximize_options control the maximization process. Options include difficult, [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), and nonrtolerance. These options are seldom used.

Syntax for etregress_scanrho

The syntax for etregress_scanrho follows. Note that the potential-outcomes option is not allowed.

etregress_scanrho *depvar* [*indepvars*] [*if*] [*in*], treat(*depvar_s* = *varlist_s*) [minrho(#) maxrho(#) step(#) vce(*vcetype*) level(#) nograph *maximize_options*]

Options for etregress_scanrho

treat(*depvar_s* = *varlist_s*) specifies the treatment equation. treat() is required.

depvar_s should be coded as 0 or 1, with 0 indicating an observation not selected and 1 indicating a selected observation.

minrho(#) specifies the minimum value of correlation between the unobservables in the selection and outcome equations to be considered. It must take a value between −1 and 1. Note that convergence may be difficult at values of −1 and 1. The default is minrho(-0.9).

maxrho(#) specifies the maximum value of correlation between the unobservables in the selection and outcome equations to be considered. It must take a value between −1 and 1. Note that convergence may be difficult at values of −1 and 1. The default is maxrho(0.9).

step(#) specifies the size of the step to use when scanning over values of correlation. This procedure will take a long time to run when the step size is small. The default is step(0.01).

vce(*vcetype*) specifies the type of standard errors to be used for the estimates. *vcetype* may be oim, robust, cluster(*clustvar*), opg, bootstrap, or jackknife.

level(#) sets the confidence level. The default is level(95) or as set by set level.

nograph suppresses the graphical output.

maximize_options control the maximization process. Options include difficult, [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), and nonrtolerance. These options are seldom used.

3.4 Bivariate probit

This next model (known as a recursive simultaneous-equation model) is not actually an extension of Heckman but was developed independently of Heckman's work (see Maddala and Lee [1976]). We include this model in our discussion because it bears a similarity to the aforementioned models. We begin with (4) and (5) but now interpret (4) as a latent variable:

$$y_i^* = \mathbf{x}_i \boldsymbol{\beta} + d_i \delta + \varepsilon_i$$
$$d_i = \begin{cases} 1 & \text{if } \mathbf{z}_i \boldsymbol{\gamma} + u_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

We denote the latent outcome as y_i^* rather than y_i to emphasize that it is not directly observed. The unobservables ε_i and u_i still follow bivariate normal distribution, but now the variance of ε_i is set to 1:

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

The econometrician observes d_i and the outcome

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

Syntax for biprobit_fixedrho

In Stata, the command biprobit can be used to fit this model. To maintain a syntax similar to our other commands (for example, heckman_fixedrho), which function similarly, the syntax for our command biprobit_fixedrho differs from biprobit.

biprobit_fixedrho *depvar* [*indepvars*] [*if*] [*in*], eq2(*depvar_s* = *varlist_s*)
rho(#) [vce(*vcetype*) level(#) *maximize_options*]

Options for biprobit_fixedrho

eq2(*depvar_s* = *varlist_s*) specifies the second equation. eq2() is required.

depvar_s should be coded as 0 or 1, with 0 indicating an observation not selected and 1 indicating a selected observation.

`rho(#)` specifies the correlation between the unobservables in the selection and outcome equations. `rho()` is required and must take a value between −1 and 1.

`vce(`*vcetype*`)` specifies the type of standard errors to be used for the estimates. *vcetype* may be `oim`, `robust`, `cluster(`*clustvar*`)`, `opg`, `bootstrap`, or `jackknife`.

`level(#)` sets the confidence level. The default is `level(95)` or as set by `set level`.

maximize_options control the maximization process. Options include `difficult`, [`no`]`log`, `trace`, `gradient`, `showstep`, `hessian`, `showtolerance`, `tolerance(#)`, `ltolerance(#)`, `nrtolerance(#)`, and `nonrtolerance`. These options are seldom used.

Syntax for biprobit_scanrho

Finally, we provide the syntax for `biprobit_scanrho`.

`biprobit_scanrho` *depvar* [*indepvars*] [*if*] [*in*], `eq2(`*depvar_s* `=` *varlist_s*`)` [`minrho(#)` `maxrho(#)` `step(#)` `vce(`*vcetype*`)` `level(#)` `nograph` *maximize_options*]

Options for biprobit_scanrho

`eq2(`*depvar_s* `=` *varlist_s*`)` specifies the second equation. `eq2()` is required.

depvar_s should be coded as 0 or 1, with 0 indicating an observation not selected and 1 indicating a selected observation.

`minrho(#)` specifies the minimum value of correlation between the unobservables in the selection and outcome equations to be considered. It must take a value between −1 and 1. Note that convergence may be difficult at values of −1 and 1. The default is `minrho(-0.9)`.

`maxrho(#)` specifies the maximum value of correlation between the unobservables in the selection and outcome equations to be considered. It must take a value between −1 and 1. Note that convergence may be difficult at values of −1 and 1. The default is `maxrho(0.9)`.

`step(#)` specifies the size of the step to use when scanning over values of correlation. This procedure will take a long time to run when the step size is small. The default is `step(0.01)`.

`vce(`*vcetype*`)` specifies the type of standard errors to be used for the estimates. *vcetype* may be `oim`, `robust`, `cluster(`*clustvar*`)`, `opg`, `bootstrap`, or `jackknife`.

`level(#)` sets the confidence level. The default is `level(95)` or as set by `set level`.

`nograph` suppresses the graphical output.

maximize_options control the maximization process. Options include difficult, [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), and nonrtolerance. These options are seldom used.

4 Examples

The help file for each command provides some examples of the syntax. In this section, we discuss the application of these commands.

▷ Identification without an exclusion restriction

Our first example considers bounding the potential effect of sample selection when we do not have a valid exclusion restriction. We use Mroz's (1987) well-known dataset of married women's wages in the 1970s. Some of the women in this dataset do not work, and thus there is no observed wage for these women. Our interest is in the effect of education on wage.

We can load this dataset by typing

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/mroz
```

Suppose that we want to regress log wage (lwage) on years of education (educ), experience (exper), and experience squared (expersq). There is a concern that unobservable variables (for example, ability) may affect both wage and the probability that a woman works. We would expect a positive relationship between the effect of ability on wage and the effect of ability on being in the labor force. This implies that we are concerned with the values of ρ that are positive.

For this regression, assume that we do not have access to a variable that affects the decision to enter the workforce but that does not affect wages directly.

Let us begin by examining the results when ρ is 0, that is, when there is no bias for linear regression:

```
. heckman_fixedrho lwage educ exper expersq, select(inlf = educ exper expersq)
> rho(0)
initial:       log likelihood = -1330.0429
alternative:   log likelihood = -1225.9652
rescale:       log likelihood = -1199.0668
rescale eq:    log likelihood = -998.57232
Iteration 0:   log likelihood = -998.57232
Iteration 1:   log likelihood = -900.69048
Iteration 2:   log likelihood = -878.80347
Iteration 3:   log likelihood = -878.76491
Iteration 4:   log likelihood = -878.76491
Heckman with a specified value of rho                   Number of obs =    753
                                                        Wald chi2(3)  =  79.60
Log likelihood = -878.76491                             Prob > chi2   = 0.0000

─────────────┬────────────────────────────────────────────────────────────────
             │ Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
─────────────┼────────────────────────────────────────────────────────────────
lwage        │
        educ │   .1074896   .0140802     7.63   0.000     .0798929    .1350864
       exper │   .0415665   .0131135     3.17   0.002     .0158645    .0672685
     expersq │  -.0008112   .0003914    -2.07   0.038    -.0015783   -.0000441
       _cons │  -.5220407   .1977017    -2.64   0.008    -.9095289   -.1345524
─────────────┼────────────────────────────────────────────────────────────────
inlf         │
        educ │   .0971238   .0221806     4.38   0.000     .0536506     .140597
       exper │   .1271342   .0178655     7.12   0.000     .0921184      .16215
     expersq │  -.0023927   .0005807    -4.12   0.000    -.0035309   -.0012546
       _cons │  -1.925493   .2887175    -6.67   0.000    -2.491369   -1.359618
─────────────┼────────────────────────────────────────────────────────────────
lnsigma      │
       _cons │  -.4105297   .0341793   -12.01   0.000    -.4775199   -.3435395
─────────────┼────────────────────────────────────────────────────────────────
       sigma │   .6632988   .0226711                        .62032    .7092555
         rho │          0
─────────────┴────────────────────────────────────────────────────────────────
```

On the other extreme, we can see the results that would be found when fixing ρ at 0.99:

```
. heckman_fixedrho lwage educ exper expersq, select(inlf = educ exper expersq)
> rho(0.99)
initial:       log likelihood = -1514.4127
alternative:   log likelihood = -1186.1468
rescale:       log likelihood = -1186.1468
rescale eq:    log likelihood = -1159.0735
Iteration 0:   log likelihood = -1159.0735
Iteration 1:   log likelihood = -1138.2448
Iteration 2:   log likelihood = -1074.3191
Iteration 3:   log likelihood = -1071.6858
Iteration 4:   log likelihood = -1071.6609
Iteration 5:   log likelihood = -1071.6609
Heckman with a specified value of rho                   Number of obs =    753
                                                        Wald chi2(3)  = 152.16
Log likelihood = -1071.6609                             Prob > chi2   = 0.0000
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
lwage						
educ	.183677	.0310964	5.91	0.000	.1227292	.2446248
exper	.2150949	.0269369	7.99	0.000	.1622995	.2678904
expersq	-.0043862	.0008355	-5.25	0.000	-.0060236	-.0027487
_cons	-3.916096	.4252316	-9.21	0.000	-4.749535	-3.082658
inlf						
educ	.1120291	.020896	5.36	0.000	.0710736	.1529846
exper	.0954691	.0170552	5.60	0.000	.0620416	.1288966
expersq	-.0019185	.0005692	-3.37	0.001	-.0030341	-.000803
_cons	-1.877612	.2719768	-6.90	0.000	-2.410677	-1.344548
lnsigma						
_cons	.5231776	.03718	14.07	0.000	.4503061	.5960491
sigma	1.687381	.0627369			1.568792	1.814934
rho	.99					

The coefficient on `educ` has increased from 0.107 to 0.184 because if the value of ρ was equal to 0.99, there would be a negative bias on `educ` when ρ is fixed at 0.

Seeing the possible values of the coefficient for different values of ρ provides bounds on the true value of the coefficient. In figure 2, we plot the estimated coefficient on `educ` as we vary ρ from 0 to 0.99.

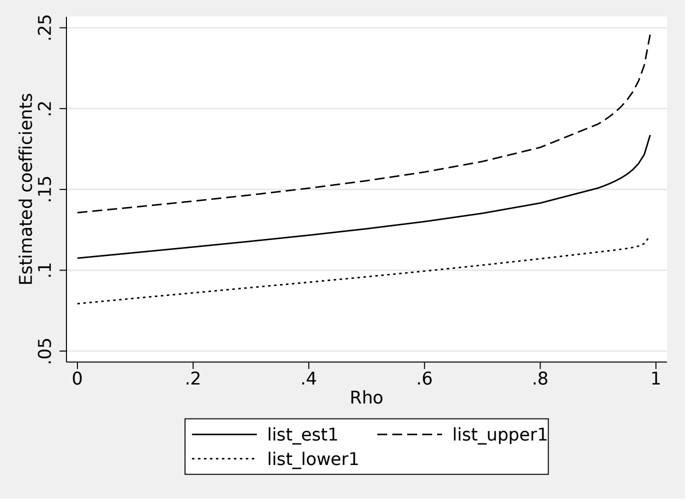


Figure 2. The estimated coefficient for various values of ρ

◁

▷ Finding maximum likelihood estimates

Our next example is a situation in which we have a valid exclusion and want to verify that we found the (true) maximum likelihood estimate. We use the specification from Zuehlke (2017), for which an author had reported a local rather than a global maximum of the likelihood function.

We begin with the Mroz dataset used in the previous example. We now use the data on family income (`faminc`) and the number of children (the sum of `kidslt6` and `kidsge6`) for our excluded variables. Unlike the previous example, our dependent variable is now raw wage instead of log wage.

Set up by loading the dataset and creating two new variables:

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/mroz, clear
. generate agesq = age^2
. generate child = kidslt6 + kidsge6
```

Call `heckman`:

```
. heckman wage educ exper expersq city,
> select(inlf = age agesq faminc child educ)
Iteration 0:   log likelihood =  -1579.565
Iteration 1:   log likelihood = -1579.4992
Iteration 2:   log likelihood = -1579.4984
Iteration 3:   log likelihood = -1579.4984
Heckman selection model                         Number of obs     =        753
(regression model with sample selection)              Selected    =        428
                                                      Nonselected =        325

                                                Wald chi2(4)      =      43.81
Log likelihood = -1579.498                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
wage         |
        educ |   .4679386   .0766016     6.11   0.000     .3178023    .6180749
       exper |   .0291692   .0620275     0.47   0.638    -.0924025    .1507409
     expersq |  -.0001513   .0018553    -0.08   0.935    -.0037876     .003485
        city |   .4467801   .3160013     1.41   0.157     -.172571    1.066131
       _cons |  -2.233281   1.330283    -1.68   0.093    -4.840587    .3740248
-------------+----------------------------------------------------------------
inlf         |
         age |    .164907   .0648387     2.54   0.011     .0378255    .2919884
       agesq |  -.0021891   .0007541    -2.90   0.004    -.0036671    -.000711
      faminc |   4.58e-06   4.53e-06     1.01   0.311    -4.29e-06    .0000135
       child |  -.1506999   .0382955    -3.94   0.000    -.2257576   -.0756422
        educ |   .0906133    .023413     3.87   0.000     .0447246     .136502
       _cons |  -3.708868   1.398727    -2.65   0.008    -6.450321   -.9674141
-------------+----------------------------------------------------------------
     /athrho |  -.0709249   .2065123    -0.34   0.731    -.4756817    .3338318
    /lnsigma |   1.130613   .0352108    32.11   0.000     1.061601    1.199625
-------------+----------------------------------------------------------------
         rho |  -.0708063    .205477                     -.4427786    .3219596
       sigma |   3.097555   .1090674                      2.890996    3.318872
      lambda |  -.2193263   .6383778                     -1.470524    1.031871
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0): chi2(1) = 0.08         Prob > chi2 = 0.7747
```

It is strange that Stata has reported a negative value of ρ; our intuition tells us that this should be positive.

Next we use `heckman_scanrho`, which will plot values of the likelihood function for each value of ρ:

```
. heckman_scanrho wage educ exper expersq city,
> select(inlf = age agesq faminc child educ)
Performing estimation for each value of rho...
initial:       log likelihood = -6697.0169
alternative:   log likelihood = -2918.6299
rescale:       log likelihood = -2029.7693
rescale eq:    log likelihood = -1650.8254
Iteration 0:   log likelihood = -1650.8254
Iteration 1:   log likelihood = -1525.1016
Iteration 2:   log likelihood = -1518.6474
Iteration 3:   log likelihood = -1518.5761
Iteration 4:   log likelihood = -1518.5761
Heckman with a specified value of rho                 Number of obs =    753
                                                      Wald chi2(4)  =  97.30
Log likelihood = -1518.5761                           Prob > chi2   = 0.0000

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
wage         |
        educ |   .6481271    .069892     9.27   0.000     .5111413    .7851129
       exper |   .0911848   .0489938     1.86   0.063    -.0048412    .1872108
     expersq |   -.001739   .0014851    -1.17   0.242    -.0046497    .0011716
        city |   .1995114   .2570384     0.78   0.438    -.3042746    .7032974
       _cons |  -7.024693   .9418348    -7.46   0.000    -8.870655    -5.17873
-------------+----------------------------------------------------------------
inlf         |
         age |    .061538   .0474361     1.30   0.195    -.0314351    .1545111
       agesq |  -.0008165   .0005495    -1.49   0.137    -.0018936    .0002605
      faminc |  -5.36e-06   3.24e-06    -1.65   0.098    -.0000117    9.97e-07
       child |  -.0557783   .0279414    -2.00   0.046    -.1105425   -.0010142
        educ |   .1372242    .021643     6.34   0.000     .0948048    .1796436
       _cons |   -2.42091   1.039939    -2.33   0.020    -4.459153   -.3826664
-------------+----------------------------------------------------------------
lnsigma      |
       _cons |   1.294831   .0352411    36.74   0.000     1.225759    1.363902
-------------+----------------------------------------------------------------
       sigma |   3.650378   .1286434                      3.406752    3.911426
         rho |        .89
------------------------------------------------------------------------------
```

The resulting plot is presented in figure 3. There is a root at $\rho = -0.07$, which is the result reported by Stata. The global maximum is around $\rho = 0.89$. Also note the difference in the log likelihoods reported for these two estimates. At the global maximum of 0.89, the log likelihood is $-1,518.58$, whereas at -0.07 it is $-1,579.50$.

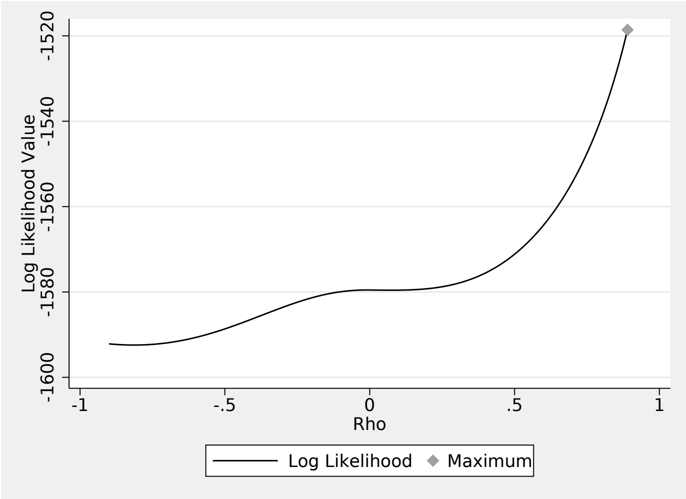


Figure 3. The value of the log-likelihood function for various values of ρ; the dot indicates the point at which the likelihood is maximized

By default, the step size is set to 0.01. It may be advisable to use a smaller step size to obtain a more accurate estimate. This may be especially helpful if the results of `heckman_scanrho` are being passed to `heckman` as initial values.

We can find the correct standard errors by setting initial values for Stata's `heckman` command near the estimates found by `heckman_scanrho`. `heckman_scanrho` returns a matrix that can be passed to `heckman` as the initial values.

We first save this matrix:

```
. matrix startv = e(init_values)
```

We then pass this matrix to `heckman`:

```
. heckman wage educ exper expersq city,
> select(inlf = age agesq faminc child educ) from(startv, copy)
Iteration 0:   log likelihood = -1518.5761
Iteration 1:   log likelihood = -1507.4948
Iteration 2:   log likelihood = -1486.0866
Iteration 3:   log likelihood = -1480.4704
Iteration 4:   log likelihood = -1480.0801
Iteration 5:   log likelihood = -1480.0792
Iteration 6:   log likelihood = -1480.0792
Heckman selection model                         Number of obs     =        753
(regression model with sample selection)              Selected    =        428
                                                      Nonselected =        325
                                                Wald chi2(4)      =      86.48
Log likelihood = -1480.079                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
wage         |
        educ |   .6677479   .0767392     8.70   0.000     .5173419    .8181539
       exper |   .0641402   .0326946     1.96   0.050       .00006    .1282203
     expersq |  -.0008423   .0010354    -0.81   0.416    -.0028717    .0011871
        city |   .0253211   .1930123     0.13   0.896    -.3529761    .4036183
       _cons |  -7.561446   1.001272    -7.55   0.000    -9.523904   -5.598989
-------------+----------------------------------------------------------------
inlf         |
         age |  -.0138954   .0311176    -0.45   0.655    -.0748848    .0470939
       agesq |   .0001649   .0003628     0.45   0.649    -.0005462     .000876
      faminc |  -6.29e-06   2.38e-06    -2.64   0.008     -.000011   -1.62e-06
       child |  -.0058901    .015534    -0.38   0.705    -.0363362     .024556
        educ |   .1572075   .0198417     7.92   0.000     .1183184    .1960966
       _cons |   -1.40227   .6973208    -2.01   0.044    -2.768993    -.035546
-------------+----------------------------------------------------------------
     /athrho |   2.872177   .2425491    11.84   0.000      2.39679    3.347565
    /lnsigma |   1.441164   .0398443    36.17   0.000     1.363071    1.519257
-------------+----------------------------------------------------------------
         rho |   .9936188   .0030856                      .9835706    .9975292
       sigma |   4.225612   .1683666                      3.908175    4.568831
      lambda |   4.198647   .1729926                      3.859588    4.537707
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0): chi2(1) = 198.92       Prob > chi2 = 0.0000
```

We are confident that these are the true maximum likelihood estimates from the plot in figure 3. Note that `heckman_scanrho` found that the value of ρ was 0.89 rather than 0.99 because, by default, `heckman_scanrho` will step through values of ρ equal to -0.90, -0.89, …, 0.89, and 0.90.

◁

▷ Comparing heckman estimates with those of heckman_fixedrho

Finally, we want to mention the differences between `heckman` and `heckman_fixedrho` when it is provided with the same value of ρ that was found by `heckman`. We use the specification from the previous example but return to using log wage instead of raw wage. First, we call `heckman`:

```
. heckman lwage educ exper expersq city,
> select(inlf = age agesq faminc child educ)
Iteration 0:   log likelihood = -917.31049  (not concave)
Iteration 1:   log likelihood = -916.06289
Iteration 2:   log likelihood = -912.80399
Iteration 3:   log likelihood = -911.72529
Iteration 4:   log likelihood = -911.72356
Iteration 5:   log likelihood = -911.72356
Heckman selection model                         Number of obs     =        753
(regression model with sample selection)              Selected    =        428
                                                      Nonselected =        325
                                                Wald chi2(4)      =      27.06
Log likelihood = -911.7236                      Prob > chi2       =     0.0000

             |  Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
lwage        |
        educ |    .065685   .0166021     3.96   0.000     .0331455    .0982245
       exper |   .0225033   .0130791     1.72   0.085    -.0031313    .0481379
     expersq |  -.0002975   .0003815    -0.78   0.436    -.0010452    .0004502
        city |   .0551856   .0655468     0.84   0.400    -.0732837    .1836549
       _cons |   .5283402   .2483942     2.13   0.033     .0414966    1.015184
-------------+----------------------------------------------------------------
inlf         |
         age |   .1125716   .0566393     1.99   0.047     .0015607    .2235826
       agesq |  -.0014949   .0006643    -2.25   0.024    -.0027968    -.000193
      faminc |   .0000114   3.77e-06     3.03   0.002     4.03e-06    .0000188
       child |  -.0883215   .0342274    -2.58   0.010    -.1554061    -.021237
        educ |   .0730682   .0217579     3.36   0.001     .0304235    .1157128
       _cons |  -2.846018   1.209941    -2.35   0.019    -5.217458   -.4745777
-------------+----------------------------------------------------------------
     /athrho |  -1.105663   .1342836    -8.23   0.000    -1.368854   -.8424716
    /lnsigma |  -.1969906   .0532265    -3.70   0.000    -.3013126   -.0926686
-------------+----------------------------------------------------------------
         rho |  -.8025238    .047799                     -.8784306    -.687116
       sigma |   .8211984   .0437095                      .7398465    .9114955
      lambda |  -.6590312   .0696919                     -.7956249   -.5224376
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0): chi2(1) = 16.96        Prob > chi2 = 0.0000
```

We now call `heckman_fixedrho` with the same value of ρ that was found by `heckman`:

```
. heckman_fixedrho lwage educ exper expersq city,
> select(inlf = age agesq faminc child educ) rho(-0.8025238)
initial:       log likelihood = -1698.4165
alternative:   log likelihood = -1332.7154
rescale:       log likelihood = -1332.7154
rescale eq:    log likelihood = -1037.9961
Iteration 0:   log likelihood = -1037.9961
Iteration 1:   log likelihood =  -975.5364
Iteration 2:   log likelihood = -923.50073
Iteration 3:   log likelihood = -921.99429
Iteration 4:   log likelihood = -921.99373
Iteration 5:   log likelihood = -921.99373
Heckman with a specified value of rho                Number of obs =    753
                                                     Wald chi2(4)  =  54.17
Log likelihood = -921.99373                          Prob > chi2   = 0.0000
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
lwage						
educ	.0852383	.0146806	5.81	0.000	.0564647	.1140118
exper	.0338607	.0129758	2.61	0.009	.0084287	.0592927
expersq	-.0005855	.0003863	-1.52	0.130	-.0013426	.0001717
city	.0508706	.0668876	0.76	0.447	-.0802267	.181968
_cons	.0800034	.2031224	0.39	0.694	-.3181091	.4781159
inlf						
age	.1951647	.0761932	2.56	0.010	.0458287	.3445007
agesq	-.0025797	.0008888	-2.90	0.004	-.0043217	-.0008377
faminc	9.96e-06	4.72e-06	2.11	0.035	7.08e-07	.0000192
child	-.1733396	.0452645	-3.83	0.000	-.2620565	-.0846227
educ	.1007933	.0280072	3.60	0.000	.0459002	.1556863
_cons	-4.737372	1.646762	-2.88	0.004	-7.964967	-1.509778
lnsigma						
_cons	-.325461	.0343588	-9.47	0.000	-.3928029	-.258119
sigma	.7221944	.0248137			.6751618	.7725033
rho	-.8025238					

The coefficients between these two are noticeably different. The source of difference is the maximization procedure being employed. The procedure used by `heckman` is preferable to the one used by `heckman_fixedrho` because it uses information about the first and second derivatives of the log-likelihood function. Notice that the log-likelihood value obtained by `heckman` is greater than the one obtained by `heckman_fixedrho` (-911.72 compared with -921.99). This is important: because `heckman_scanrho` calls `heckman_fixedrho`, the user needs to verify that the results found by `heckman_scanrho` do in fact improve the log likelihood relative to the results found by `heckman`.

◁

5 Discussion and conclusion

Several extensions of this work are possible. We maintained the bivariate normality assumption in all of these models. This could be relaxed in several ways. Altonji, Elder, and Taber (2005) use Heckman and Singer's (1984) approach for allowing for deviations for normality, which involves treating the stochastic terms as having a discrete component. Modeling the stochastic terms as a mixture of normals would also allow for some deviations from normality.

Another extension would be to change the maximization procedure used. Stata has several options for maximizing likelihood functions, which differ in whether derivatives of the likelihood functions are provided. The maximization methods that we used do not use any information about derivatives of the likelihood function. As a result, there may be specifications for which `heckman` converges but `heckman_fixedrho` does not converge (in addition to potential differences between the estimators as mentioned in the last example). This is because `heckman` uses information about the first and second derivative of the likelihood function.

The commands that we presented can be used to examine the sensitivity of regression results to sample selection or endogenous treatment and to verify that the results of a Heckman model are the global maximum of the likelihood function.

6 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 21-4
. net install st0658      (to install program files, if available)
. net get st0658          (to install ancillary files, if available)
```

7 Acknowledgments

Jonathan Cook thanks Victor Jarosiewicz, Arshad Rahman, Sarah Wolfolds, Thomas Zuehlke, and an anonymous referee for helpful comments.

The views expressed in this article are the views of the authors and do not necessarily reflect the views of the authors' employers or any other entities with which the authors may be associated.

8 References

Altonji, J. G., T. E. Elder, and C. R. Taber. 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy* 113: 151–184. https://doi.org/10.1086/426036.

Aobdia, D. 2019. Do practitioner assessments agree with academic proxies for audit quality? Evidence from PCAOB and internal inspections. *Journal of Accounting and Economics* 67: 144–174. https://doi.org/10.1016/j.jacceco.2018.09.001.

Arabmazar, A., and P. Schmidt. 1982. An investigation of the robustness of the Tobit estimator to non-normality. *Econometrica* 50: 1055–1063. https://doi.org/10.2307/1912776.

Blundell, R. W., and J. L. Powell. 2004. Endogeneity in semiparametric binary response models. *Review of Economic Studies* 71: 655–679. https://doi.org/10.1111/j.1467-937X.2004.00299.x.

Certo, S. T., J. R. Busenbark, H.-s. Woo, and M. Semadeni. 2016. Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal* 37: 2639–2657. https://doi.org/10.1002/smj.2475.

Chan, J. Y., and J. A. Cook. 2020. Inferring Zambia's HIV prevalence from a selected sample. *Applied Economics* 52: 4236–4249. https://doi.org/10.1080/00036846.2020.1733477.

Choudhary, P., K. Merkley, and K. Schipper. 2019. Auditors' quantitative materiality judgments: Properties and implications for financial reporting reliability. *Journal of Accounting Research* 57: 1303–1351. https://doi.org/10.1111/1475-679X.12286.

Cook, J., and S. Siddiqui. 2020. Random forests and selected samples. *Bulletin of Economic Research* 72: 272–287. https://doi.org/10.1111/boer.12222.

Downey, D. H., J. C. Bedard, and C. M. Boland. 2020. Monitoring quality of group audits: Internal and regulatory inspections of component auditors of U.S. issuers. Working paper.

Heckman, J., and B. Singer. 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52: 271–320. https://doi.org/10.2307/1911491.

Heckman, J. J. 1976. The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement*, ed. S. V. Berg. Vol. 5, 475–492. Cambridge, MA: National Bureau of Economic Research.

———. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46: 931–959. https://doi.org/10.2307/1909757.

———. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–161. https://doi.org/10.2307/1912352.

———. 1990. Varieties of selection bias. *American Economic Review* 80: 313–318.

Maddala, G. S., and L.-F. Lee. 1976. Recursive models with qualitative endogenous variables. In *Annals of Economic and Social Measurement*, ed. S. V. Berg. Vol. 5, 525–545. Cambridge, MA: National Bureau of Economic Research.

Mroz, T. A. 1987. The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* 55: 765–799. https://doi.org/10.2307/1911029.

Olsen, R. J. 1982. Distributional tests for selectivity bias and a more robust likelihood estimator. *International Economic Review* 23: 223–240. https://doi.org/10.2307/2526473.

Robinson, P. M. 1982. On the asymptotic properties of estimators of models containing limited dependent variables. *Econometrica* 50: 27–41. https://doi.org/10.2307/1912527.

Rodemeier, M. 2020. Buy baits and consumer sophistication: Theory and field evidence from large-scale rebate promotions. Working paper, University of Muenster.

Tran, T. Q., and V. T. T. Dinh. Forthcoming. Provincial governance and financial inclusion: Micro evidence from a Rural Vietnam. *International Public Management Journal*. https://doi.org/10.1080/10967494.2021.1964009.

Van de Ven, W. P. M. M., and B. M. S. Van Praag. 1981. The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics* 17: 229–252. https://doi.org/10.1016/0304-4076(81)90028-2.

Wolfolds, S. E., and J. Siegel. 2019. Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument. *Strategic Management Journal* 40: 432–462. https://doi.org/10.1002/smj.2995.

Zuehlke, T. W. 2017. Use of quadratic terms in Type 2 Tobit models. *Applied Economics* 49: 1706–1714. https://doi.org/10.1080/00036846.2016.1223831.

About the authors

Jonathan Cook is an applied economist.

Joon-Suk Lee has worked as a financial economist at a regulatory agency since 2015.

Noah Newberger is a senior data scientist with empirical research interests.