Test Statistics and Critical Values

in Selectivity Models

R. Carter Hill
Economics Department
Louisiana State University
Baton Rouge, LA 70803
Phone: 225-578-1490
e-mail: eohill@lsu.edu

Lee C. Adkins
Department of Economics and Legal Studies in Business
Oklahoma State University
Stillwater, OK 74078
Phone: 405-744-8637
e-mail: ladkins@okstate.edu

Keith A. Bender
Economics Department
University of Wisconsin-Milwaukee
Milwaukee, WI 53201
Phone: 414-229-4761
e-mail: kabender@uwm.edu

Abstract: The Heckman two-step estimator (Heckit) for the selectivity model is widely applied in Economics and other social sciences. In this model a nonzero outcome variable is observed only if a latent variable is positive. The asymptotic covariance matrix for two-step estimation procedure must account for the estimation error introduced in the first stage. We examine the finite sample size of tests based on alternative covariance matrix estimators. We do so by using Monte Carlo experiments to evaluate bootstrap generated critical values and critical values based on asymptotic theory.

Keywords: Selection bias, two-stage estimation, robust covariance estimator, bootstrapping, hypothesis tests.

JEL: C12, C15, C24

Version: 7/10/2003 11:47 AM

Acknowledgment: The authors would like to thank Bill Greene for responding to our queries about LIMDEP and making an adjustment in the program. The authors benefited from discussion by participants at the *LSU Advances in Econometrics Conference, November 2002*, Tom Fomby and Viera Chmelarova. Any errors are ours.

1. Introduction

Many researchers use Heckman's (1979) two-step estimation procedure (Heckit) to deal with selectivity in the linear regression model. Selection bias results when the regression dependent variable is observed only when a 'latent' selection variable is positive. While the two-step estimation procedure is easy to implement, i.e., a probit estimation of the selection equation followed by least squares estimation of an augmented regression, the applied literature reveals that researchers (and econometric software vendors) take a variety of approaches when computing standard errors. This is important since the standard errors are the basis for *t*-statistics that are used in significance tests.

The literature on sample selection bias is huge, and we will not attempt a survey. Two articles of note are by Vella (1998) and Puhani (2000). In this paper we focus on two issues.

- First, how do the alternative versions of asymptotic variance-covariance matrices used in selectivity models capture the finite sample variability of the Heckit two-step estimator?
- Second, is it possible to use bootstrapping to improve finite sample inference? Three aspects of this question are:
    - Do bootstrap standard errors match finite sample variability better than nominal standard errors computed from asymptotic covariance matrices?
    - Do critical values of test statistics generated by bootstrapping pivotal *t*-statistics lead to better test size (and power?) than those based on usual asymptotic theory?
    - Does modern software make obtaining bootstrap standard errors and critical values feasible for empirical researchers?

Our plan is to develop the model and Heckman's two-step estimator in Sections 2 and 3, respectively, and then describe the alternative covariance matrix estimators we consider in Section 4. In Section 5 we comment on the practices used in empirical research and some choices available in commercial software. Section 6 contains the design of the Monte Carlo experiment as well as discussion of the output measures we recover. Section 7 presents the results of the Monte Carlo experiment, followed by conclusions in Section 8.

2. *The Selectivity Model*

Following Greene (1997, 974-981) consider a model consisting of two equations. The first equation is the "selection equation," which is defined as

$$z_i^* = w_i'\gamma + u_i, \quad i = 1, \ldots, N \tag{2.1}$$

where z_i^* is a latent variable, γ is a $K\times 1$ vector of parameters, w_i' is a $1\times K$ row vector of observations on K exogenous variables and u_i is a random disturbance. The latent variable is unobservable, but we do observe the dichotomous variable

$$z_i = \begin{cases} 1 & z_i^* > 0 \\ 0 & otherwise \end{cases} \tag{2.2}$$

The second equation is the linear model of interest. It is

$$y_i = x_i'\beta + e_i, \quad i = 1, \ldots, n, \quad N > n \tag{2.3}$$

where y_i is an observable random variable, β is an $M\times 1$ vector of parameters, x_i' is a $1\times M$ vector of exogenous variables and e_i is a random disturbance. We assume that the random disturbances are jointly distributed as

$$\begin{bmatrix} u_i \\ e_i \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & \sigma_e^2 \end{pmatrix}\right] \tag{2.4}$$

A "selectivity problem" arises when y_i is observed only when $z_i = 1$, and if $\rho \neq 0$. In such a situation, the ordinary least squares estimator of β in (2.3) is biased and inconsistent. A consistent estimator is the two-step procedure suggested by Heckman (1979) and clarified by Greene (1981). It is known as "Heckman's two-step estimator," or more simply as "Heckit." The basis for this estimation procedure is the conditional regression function

$$E\left[y_i \mid z_i > 0\right] = E\left[y_i \mid u_i > -w_i'\gamma\right] = x_i'\beta + E\left[e_i \mid u_i > -w_i'\gamma\right] = x_i'\beta + \left(\rho\cdot\sigma_e\right)\lambda_i \tag{2.5}$$

where

$$\lambda_i = \frac{\phi(w_i'\gamma)}{\Phi(w_i'\gamma)} \tag{2.6}$$

is the "inverse Mill's ratio," $\phi(\cdot)$ is the standard normal probability density function evaluated at the argument, and $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable evaluated at the argument. In a regression equation format,

$$y_i = E\left[y_i \mid z_i > 0\right] + v_i = x_i'\beta + \left(\rho \cdot \sigma_e\right)\lambda_i + v_i = x_i'\beta + \beta_\lambda \lambda_i + v_i \tag{2.7}$$

where the random disturbance v_i has conditional mean and variance given by

$$E\left[v_i \mid z_i > 0\right] = 0, \quad \text{var}\left(v_i \mid z_i > 0\right) = \sigma_e^2\left(1 - \rho^2 \delta_i\right) \tag{2.8}$$

with

$$\delta_i = \lambda_i\left(\lambda_i + w_i'\gamma\right) \tag{2.9}$$

Thus the regression error v_i in (2.7) is heteroskedastic.

If λ_i were known and nonstochastic, then the selectivity corrected model (2.7) could be estimated by generalized least squares. Alternatively, the heteroskedastic model (2.7) could be estimated by ordinary least squares, and the heteroskedasticity consistent covariance matrix estimator (HCE) of Huber (1967), Eicker (1967) and White (1980) used for hypothesis testing and construction of confidence intervals. Unfortunately λ_i is not known and must estimated, introducing variability not accounted for by a heteroskedasticity correction.

3. ***Heckit: The Heckman Two-Step Estimator***

The widespread use of Heckit is no doubt in part due to the ease with which it is computed. In the first step of this two-stage estimation process, the method of maximum likelihood is used to estimate the probit model parameters γ, based upon the observable random variable z_i. Denote the MLE as $\hat{\gamma}$ and its asymptotic covariance matrix as V. We will compute V as the negative of the inverse Hessian evaluated at the MLE. Specifically, if f_{i1} is the (Bernoulli) probability function of z_i, then

$$\ln f_{i1} = \ln\left(\left[\Phi\left(w_i'\gamma\right)\right]^{z_i}\left[1 - \Phi\left(w_i'\gamma\right)\right]^{1-z_i}\right) \tag{3.1}$$

and the log-likelihood function for the probit model is $L_1 = \sum_{i=1}^{N} \ln f_{1i}$. For convenience let $\phi_i = \phi\left(w_i'\gamma\right)$ and $\Phi_i = \Phi\left(w_i'\gamma\right)$. Then, for future reference, the gradient vector of the probit model log-likelihood function is

$$\frac{\partial L_1}{\partial \gamma} = \sum_{i=1}^{N} z_i \frac{\phi(w_i'\gamma)}{\Phi(w_i'\gamma)} w_i - \sum_{i=1}^{N} (1-z_i) \frac{\phi(w_i'\gamma)}{1-\Phi(w_i'\gamma)} w_i \tag{3.2}$$

and the Hessian is

$$\frac{\partial^2 L_1}{\partial \gamma \partial \gamma'} = -\sum_{i=1}^{N} \phi_i \left[z_i \frac{\phi_i + (w'\gamma_i)\Phi_i}{\Phi_i^2} + (1-z_i) \frac{\phi_i - (w'\gamma_i)(1-\Phi_i)}{(1-\Phi_i)^2} \right] w_i w_i' \tag{3.3}$$

Then, using the negative of the inverse Hessian evaluated at the MLE, we compute the asymptotic covariance matrix V of the probit estimator as

$$V = -\left[\frac{\partial^2 L_1}{\partial \gamma \partial \gamma'} \right]^{-1} \tag{3.4}$$

Given the MLE $\hat{\gamma}$, we compute $\hat{\lambda}_i = \phi(w_i'\hat{\gamma})/\Phi(w_i'\hat{\gamma})$, $i = 1, \ldots, n$. The variable $\hat{\lambda}_i$ is used as a regressor in the second step equation,

$$y_i = x_i'\beta + \beta_\lambda \hat{\lambda}_i + v_i + \beta_\lambda \left(\lambda_i - \hat{\lambda}_i \right) = x_i'\beta + \beta_\lambda \hat{\lambda}_i + \underset{\sim}{v}_i = \begin{bmatrix} x_i' & \hat{\lambda}_i \end{bmatrix} \begin{bmatrix} \beta \\ \beta_\lambda \end{bmatrix} + \underset{\sim}{v}_i = \underset{\sim}{x}_i{}'\beta^* + \underset{\sim}{v}_i \tag{3.5}$$

Stacking the n complete observations in (3.5) into matrices, the Heckit estimator of the $M + 1$ parameters β^* is the ordinary least squares estimator applied to the augmented regression in (3.5), that is

$$b^* = \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \underset{\sim}{X}' y \tag{3.6}$$

An estimator of σ_e^2 is $\hat{\sigma}_e^2 = \left(\hat{\underset{\sim}{v}}' \hat{\underset{\sim}{v}} / n \right) + \bar{\hat{\delta}} b_\lambda^2$, where $\hat{\underset{\sim}{v}} = y - \underset{\sim}{X} b^*$, b_λ is the least squares estimator of β_λ, $\hat{\delta}_i$ is the plug-in estimator of δ_i in (2.9), and $\bar{\hat{\delta}} = \sum_{i=1}^{n} \hat{\delta}_i / n$. An estimator of ρ is $\hat{\rho}^2 = b_\lambda^2 / \hat{\sigma}_e^2$.

4. Heckit: Alternative Covariance Matrix Estimators Considered in the Monte Carlo

The focus in this paper is the computation of the covariance matrix estimate. There are a number of alternatives used in the literature, some of which are appropriate and some of which are not.

(1) One alternative is to use the standard estimator for the covariance matrix of the OLS estimator,

$$V_{OLS} = \hat{\sigma}_v^2 \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1}, \quad \hat{\sigma}_v^2 = \frac{\hat{\underset{\sim}{v}}' \hat{\underset{\sim}{v}}}{n - (M+1)} \tag{4.1}$$

This estimator is incorrect for two reasons: it ignores the heteroskedasticity in (2.7), even if we assume λ_i is known, and it ignores that in (3.5) $\hat{\lambda}_i$ is stochastic.

(2) A second possibility is to account for the heteroskedasticity in (2.7), using the estimator

$$V_{HET} = \hat{\sigma}_e^2 \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \left[\underset{\sim}{X}' \left(I - \rho^2 \hat{\Delta} \right) \underset{\sim}{X} \right] \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \quad (4.2)$$

where $\hat{\Delta} = diag\left(\hat{\delta}_i\right)$ is a diagonal matrix with nonzero elements $\hat{\delta}_i$. This estimator is the covariance matrix of the ordinary least squares estimator in the presence of heteroskedasticity of the form in (2.8). This estimator does not account for the fact that $\hat{\lambda}_i$ in (3.5) is stochastic. The error term $\underset{\sim}{v}_i$ in (3.5) is for each observation a function of the probit MLE $\hat{\gamma}$. This means that $\text{cov}(\underset{\sim}{v}_i, \underset{\sim}{v}_j) \neq 0$, which (4.2) does not account for.

(3) In this same spirit, one might use the White (1980) heteroskedasticity consistent estimator (HCE)

$$V_{HCE} = \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \underset{\sim}{X}' D_p \underset{\sim}{X} \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \quad (4.3)$$

where D_p is a diagonal matrix. For the basic heteroskedasticity consistent estimator V_{HC_0} the squared least squares residuals are on the diagonal, $D_0 = diag\left(\hat{\underset{\sim}{v}}_i^2\right)$. This estimator suffers the same flaw as V_{HET}, but is easier to compute since V_{HC_0} is readily available in most regression software.

We also consider a modification of the standard heteroskedasticity consistent estimator V_{HC_0}. This modification is denoted V_{HC_3} by Davidson and MacKinnon (1993, p. 554) and has been studied recently by Long and Ervin (2000). To obtain V_{HC_3} we replace the diagonal matrix D_0 with $D_3 = diag\left(\hat{\underset{\sim}{v}}_i^2 \Big/ \left(1 - \underset{\sim}{x}_i' \left(\underset{\sim}{X}'\underset{\sim}{X} \right)^{-1} \underset{\sim}{x}_i \right)^2 \right)$. Davidson and MacKinnon note (1993, p.70) that in a linear regression model with homoskedastic errors the expected value of these weighted residuals would equal the error variance.

Justification for using V_{HC_0} is found in Lee (1982). Amemiya (1984, p.33) summarizes the argument by noting that the Heckman estimator is consistent under assumptions less restrictive than the usual bivariate normality of the error terms. Under the less restrictive assumptions White's HCE provides a consistent estimator of the asymptotic covariance matrix. The specific form we use is Lee (1982, p. 365) equation (51), which is

$$V_{Lee} = \left(\underset{\sim}{X}'\underset{\sim}{X} \right)^{-1} \left(V_{HC_0} - \underset{\sim}{X}'GVG'\underset{\sim}{X} \right) \left(\underset{\sim}{X}'\underset{\sim}{X} \right)^{-1}$$

where G is the matrix with *i'th* row $\beta_\lambda\, \partial\lambda_i / \partial\gamma'$. See (4.8) below for the derivative.

(4) The asymptotic covariance matrix for b^* was obtained by Heckman (1979, p. 159) and refined by Greene (1981, pp. 795-798). It is, following Greene (1997, p. 981),

$$V_{HECK} = \hat{\sigma}_e^2 \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \left[\underset{\sim}{X}' \left(I - \rho^2 \hat{\Delta} \right) \underset{\sim}{X} + Q \right] \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \tag{4.4}$$

where $Q = \hat{\rho}^2 FVF'$ and $F = \underset{\sim}{X}' \hat{\Delta} W$, with V the asymptotic covariance matrix of the probit estimator, the negative of the inverse Hessian, and with W being the regressor matrix from the probit estimation in the first stage.

(5) An asymptotically equivalent, but seldom if ever used in selectivity models, asymptotic covariance matrix estimator is based on the Murphy-Topel (1985) general result. We are including the MT estimator because of its use in an increasingly wide variety of contexts, and its availability as an automated command in LIMDEP 8.0. In Greene's (1997, p.467) notation

$$y_i = h\left(x_i, \beta^*, w_i, \gamma\right) + v_i = x_i'\beta + \beta_\lambda \lambda_i + v_i = \underset{\sim}{x}_i'\beta^* + v_i \tag{4.5}$$

Let $\operatorname{cov}\left(b^* \mid \hat{\lambda}_i\right) = V_{HET} = \sigma_e^2 V_b$ [see equation (4.2)]. Then the unconditional covariance matrix of the least squares estimator is

$$V_{MT} = \sigma_e^2 V_b + V_b \left[CVC' - RVC' - CVR' \right] V_b \tag{4.6}$$

where V is the asymptotic covariance matrix of the probit estimator. The matrix C is given by

$$C = \sum x_i^0 \cdot \frac{\partial h}{\partial \gamma'}, \quad x_i^0 = \frac{\partial h}{\partial \beta^*} = \underset{\sim}{x}_i, \quad \frac{\partial h}{\partial \gamma'} = \beta_\lambda \frac{\partial \lambda_i}{\partial \gamma'} \tag{4.7}$$

where

$$\frac{\partial \lambda_i}{\partial \gamma'} = -\frac{\phi\left(w_i'\gamma\right)}{\left[\Phi\left(w_i'\gamma\right)\right]^2} \left[\Phi\left(w_i'\gamma\right) \cdot \left(w_i'\gamma\right) + \phi\left(w_i'\gamma\right) \right] w_i' = \lambda_{\gamma i}' \tag{4.8}$$

The matrix R is

$$R = \sum \underset{\sim}{x}_i \hat{\underset{\sim}{v}}_i \frac{\partial \ln f_{i1}}{\partial \gamma'} \tag{4.9}$$

For $i = 1, \ldots, n$ the selectivity indicator $z_i = 1$ in the probit model so that for these observations the i'th term of the log-likelihood function L_1 is $\ln f_{i1} = \Phi\left(w_i'\gamma\right)$ and $\dfrac{\partial \ln f_{i1}}{\partial \gamma'} = \dfrac{\phi\left(w_i'\gamma\right)}{\Phi\left(w_i'\gamma\right)} w_i' = \lambda_i w_i'$.

(6) Hardin (2002, 2003) has introduced a robust variance estimator for two-stage estimators similar in construction to the Murphy-Topel estimator. In it, which we call V_{RMT}, the matrix C in (4.6) is replaced by

$$C^{*} = -\frac{\partial^2 L_2}{\partial\theta_1 \partial\theta_2'} \tag{4.10}$$

where L_2 is the log-likelihood function of the second stage estimator, conditional on the parameters in the first stage. In the context of Heckit the first stage estimator is probit, and $\theta_1 = \gamma$. The second stage is least squares estimation of the augmented model (4.5), so $\theta_2' = \beta^{*\prime} = \begin{pmatrix}\beta' & \beta_\lambda\end{pmatrix}$. To compute the matrix C^* we write the log-likelihood function for the second stage regression as

$$\begin{aligned} L_2 &= \sum_{i=1}^{n}\left[-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\left(y_i - x_i'\beta - \lambda_i\beta_\lambda\right)^2/\sigma^2\right] = \sum_{i=1}^{n}\left[-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}v_i^2/\sigma^2\right] \\ &= \sum_{i=1}^{n}\ln f_{i2} \end{aligned} \tag{4.11}$$

Then,

$$\frac{\delta \ln f_{i2}}{\delta\beta^*} = \frac{1}{\sigma^2}\begin{bmatrix}\left(y_i - x_i'\beta - \lambda_i\beta_\lambda\right)x_i \\ \left(y_i - x_i'\beta - \lambda_i\beta_\lambda\right)\lambda_i\end{bmatrix} \tag{4.12a}$$

and

$$\frac{\partial^2 \ln f_{i2}}{\partial\beta^*\partial\gamma'} = \frac{1}{\sigma^2}\begin{bmatrix}-\beta_\lambda x_i \dfrac{\partial\lambda_i}{\partial\gamma'} \\ v_i\dfrac{\partial\lambda_i}{\partial\gamma'} - \lambda_i\beta_\lambda\dfrac{\partial\lambda_i}{\partial\gamma'}\end{bmatrix} = \begin{bmatrix}-\dfrac{\beta_\lambda}{\sigma^2}x_i \\ \left(\dfrac{v_i}{\sigma^2\lambda_i} - \dfrac{\beta_\lambda}{\sigma^2}\right)\lambda_i\end{bmatrix}\lambda_{\gamma i}' = \underline{x}_i\lambda_{\gamma i}' \tag{4.12b}$$

Finally,

$$C^{*} = -\sum_{i=1}^{n}\underline{x}_i\lambda_{\gamma i}' = -\underline{X}'\Lambda_\gamma \tag{4.13}$$

where $\underline{X}$ is a matrix whose rows are $\underline{x}_i'$ and Λ_γ is a matrix with rows $\lambda_{\gamma i}'$ which is given in (4.8).

The second modification of the Murphy-Topel estimator Hardin introduces is to replace the probit covariance matrix V by the "sandwich" estimator $V_s = V \cdot V_{opg}^{-1} \cdot V$, where

$$V_{opg}^{-1} = \left[\frac{\partial L_1}{\partial\gamma}\frac{\partial L_1}{\partial\gamma'}\right]$$

where L_1 is the probit log-likelihood function, and V_{opg} is the "outer product of the gradient" estimator of the probit covariance matrix.

(7) Our concern is not just the asymptotic reliability of covariance matrix estimators. Selectivity corrections are often applied in samples that are not "large." It is known that standard errors computed from asymptotically valid covariance matrices can seriously understate true estimator variability in finite samples (Horowitz, 1997). In the context of models with limited dependent variables this has been illustrated by Griffiths, Hill and Pope (1987). We use the bootstrap approach (Freedman and Peters, 1984 a,b; Jeong and Maddala, 1993) to compute measures of finite sample variability. As suggested by Jeong and Maddala (1993, p. 577) we resample, with replacement, from the rows of the data matrix to create a large number of "bootstrap samples." LIMDEP and STATA employ this scheme for automated bootstrapping. Using each bootstrap sample we re-estimate the model, storing the parameter estimates as we go. The bootstrap estimate of the estimator standard error is the sample standard deviation of all the estimates.

Horowitz (1997) argues that while the bootstrap can be used to approximate standard errors of estimation, it is preferable to use the bootstrap to obtain critical values for *t*-statistics that are used as a basis for hypothesis testing. The crux of the argument is that bootstrap standard errors converge to the true standard errors as the sample size gets larger, but that bootstrapped critical values of test statistics converge to the true critical values at an even faster rate. Thus, instead of computing bootstrap standard errors and using these as a basis for a new *t*-statistic, it is perhaps better to bypass the computation of the standard error and just compute the critical value of the commonly used asymptotic *t*-statistic.

Using standard econometric practice, to test the null hypothesis $H_0 : \beta_k = \beta_k^0$ against the alternative $H_0 : \beta_k \neq \beta_k^0$ we use the *t*-statistic $t = \left(b_k - \beta_k^0\right) / se\left(b_k\right)$, where $se\left(b_k\right)$ is a valid asymptotic standard error for the consistent estimator b_k. For example, $se\left(b_k\right)$ might be the Heckit standard error from V_{HECK}, or the standard error computed from the Murphy-Topel asymptotic covariance matrix V_{MT}. The null hypothesis is rejected if $| t |\geq t_c$, where t_c is the critical value from the standard normal or *t*-distribution.

The problem is that using the standard critical values with *t*-statistics derived from asymptotic theory leads to tests with incorrect size, or probability of type I error. Horowitz suggests that we can obtain improved critical values using the bootstrap. In each bootstrap sample we obtain the *t*-statistic value for the (true in the sample) hypothesis $H_0 : \beta_k = b_k$. That is, we compute for each bootstrap sample $t_b = \left(\hat{\beta}_b - b_k\right) / se_b$, where $\hat{\beta}_b$ is an estimate from a consistent estimator in the b'th bootstrap sample, and se_b is the value of the asymptotic standard error of the estimator $\hat{\beta}_b$. In the bootstrap world it is b_k, the estimate based on the original and full sample, that plays the role of the true parameter value. Hence the statistic t_b is centered at the "true in the sample" parameter value, and it is asymptotically pivotal. The absolute values of these t-statistics are sorted by magnitude and the positive t-critical value t_c is chosen to be the upper α-percentile value. We will report, from the Monte Carlo experiment that we outline in the next section, both the standard errors computed by the bootstrap, and the bootstrap t-critical values.

5. Approaches in the Empirical Literature: Some Comments

At least four different strategies are being used to obtain standard errors in selectivity models appearing in the applied literature.

Heckman's Asymptotic Covariance Matrix: Many authors cite the analytic results of Heckman (1979) or Greene (1981), who derive expressions for asymptotic covariance matrices in selectivity models, as the source of their standard errors. We may include in this group the authors who indicate they use software packages LIMDEP or STATA which will compute the Heckman asymptotic covariance matrix. It is our view that authors of empirical papers should clearly state which software is being used, which version, and the commands for options actually employed. Not doing so hinders the work of other researchers in the area and obscures calculations that may be incorrect. For example, early versions of STATA 6.0 contained an error when computing the Heckman asymptotic covariance matrix, a reminder that we should all keep our software updated.

Maximum Likelihood Estimator: Less seldom used in applications is the method of maximum likelihood. The most recent versions of LIMDEP and STATA include both two-step estimator and maximum likelihood options. In addition, both packages offer a Huber/White/sandwich estimator for the asymptotic covariance matrix when maximum likelihood estimation is chosen. Examples in the applied literature are Hunter (2000) and Wu and Kwok (2002). Nawata (1994) notes that standard software routines for finding the MLE may not converge, or may converge to a local rather than a global maximum. This may explain why the two-step estimator is widely used instead of maximum likelihood.

Nawata (1994) offers an alternative to standard optimization routines for maximizing the log-likelihood function. Nawata and Nagase (1996) compare the finite sample properties of the MLE and Heckit. They conclude, confirming the result in Nawata (1993), that a key indicator to the likely performance of Heckit relative to the MLE is the collinearity between the systematic portion of the selection equation and the regressors in the equation of interest. If the selection (probit) equation and the equation of interest have a substantial number of variables in common then the Heckit estimator is not a good choice relative to the MLE. Nawata and McAleer (2001) compare the *t*-test, likelihood ratio test and Lagrange multiplier test of the hypothesis that the errors in the selection equation and equation of interest are uncorrelated using maximum likelihood. They find that even if there is no collinearity between the regressors in the two equations the *t*-test based on maximum estimates performs poorly due to poor variance estimates.

<u>White's Heteroskedasticity Consistent Estimator</u>: We discovered a large group of authors who rely upon White's HCE in selectivity models. Puhani (2000, p.55) states: "In order to obtain a simple and consistent estimator of the asymptotic variance-covariance matrix, Lee (1982, p.364f.) suggests to use White's (1980) method." We are extremely curious about this point. We note first that blind use of the HCE does not account for the fact that the parameters of the first stage probit model are estimated. Second, Lee (1982, p.365) indicates that he suggests HCE only for part of his two-step estimator's (which is a generalization of the standard Heckit estimator) variance-covariance matrix. We are unsure whether applied researchers have relied on Lee's (1982, p.364) sentence "A relatively simpler approach which avoids the above complication is to adopt the method in White (1980)," assuming it means that a simple heteroskedasticity correction is adequate asymptotically in the two-step estimation process. Amemiya (1984, p.33) summarizes the argument by noting that the Heckman estimator is consistent under assumptions less restrictive than the usual bivariate normality of the error terms. Under the less restrictive assumptions White's HCE provides a consistent estimator of the asymptotic covariance matrix.

Simpson (1986, p. 801), in the context of a model somewhat more involved than the usual selectivity model, acknowledges that

> "Ordinary least squares estimates … will be consistent, but the resulting standard errors will be incorrect and the estimation of correct standard errors will be complex. Therefore, the estimates of the standard errors reported … use White's (1980) procedure to adjust for heteroskedasticity induced by sample selection. *This procedure does not correct for the fact that γ is unknown and must be estimated*." (Italics added).

Similarly, Ermisch and Wright (1993, p. 123) obtain *t*-statistics using OLS standard errors, but then say,

> "When the *t*-statistics are computed from standard errors corrected for heteroskedasticity using White's (1980) method, conclusions about statistical significance are the same. These *t*-statistics are not, however, the appropriate ones…. While they correct for heteroskedasticity*, they do not allow for the fact that the λ regressor is estimated*. (Italics added.) The correct asymptotic standard errors for the dichotomous selection case are derived in Heckman (1979, pp. 158-59)."

<u>Ordinary Least Squares:</u> Another strategy is to use least squares regression. The ill-conditioning in the Heckit regression model, which includes the inverse Mills ratio, becomes severe when the variables in the probit selection equation are highly correlated with the variables in the regression equation. The ill-conditioning in the augmented regression occurs in this case because the inverse Mills ratio, despite the fact that it is a nonlinear function the selection equation variables, is well approximated by a linear function over broad ranges. In such cases Monte Carlo evidence exists [summarized in Puhani (2000)] supporting the small sample estimation efficiency of OLS relative to Heckit and Heckit alternatives.

<u>Recent Developments in Covariance Matrix Estimation:</u> In recent years the standard approaches to computing standard errors have been expanded in two directions. First, the use of bootstrapping has become more refined. Originally used as an empirical method for computing standard errors, bootstrapping is now seen as a basis for computing empirically relevant critical values for hypothesis tests [Horowitz (1997), Deis and Hill (1998), Horowitz and Savin (2000)]. The advantage of this approach is that it leads to the use of a test procedure having Type I error specified by the researcher in finite samples. As Horowitz and Savin (2000) point out, this feature facilitates power comparisons among alternative tests.

Newer versions of the software STATA and LIMDEP have automated bootstrap commands. These can be used to obtain bootstrap standard errors for any estimation procedure via resampling the data. With the appropriate modifications the pivotal statistics can be resampled, making bootstrap critical values easy to obtain. Some sample programs are provided in the appendix to this paper showing the commands required.

"Sandwich" covariance matrix estimators are being used in a wide variety of contexts. The heteroskedasticity-consistent covariance matrix estimator (HCE) introduced by Huber (1967), Eicker (1967) and White (1980) is ubiquitous. The software STATA offers robust covariance estimation as an option for virtually all models that can be estimated by maximum likelihood [STATA 7.0 *User's Guide*, p.254]. These robust covariance estimators are used to obtain "consistent estimates of the covariance matrix under misspecified working covariances as well as under heteroscedastic errors" [Kauermann and Carroll (2001, p. 1387)]. Recently Hardin (2002) has proposed a sandwich estimator for two-step models, a class of models that includes Heckit. He develops a sandwich covariance matrix estimator similar in construction to the Murphy-Topel (1985) estimator for two-step estimators, and he provides instructions for using STATA to implement the estimator. LIMDEP 8.0 offers a general procedure for two-step estimators with the Murphy-Topel covariance matrices automated for some special cases.

It is our conjecture that bootstrapping a pivotal *t*-statistic, using a standard error from any consistent asymptotic covariance matrix estimator for the Heckit model (or any other), should yield critical values such that hypothesis tests are the proper size. Then the question is "Do any of the alternative covariance matrix estimators lead to power advantages?" We have not yet investigated this question.

6. *The Monte Carlo Design*

The Monte Carlo Design we employ is that of Zuehlke and Zeman (1991) with a modification used by Nawata and Nagase (1996). The performance of each covariance estimator is examined under various circumstances likely to affect their performances. The sample size, severity of censoring, degree of selection bias and the correlation between the independent variables in the selection and regression equations are all varied within the Monte Carlo experiment.

The specific model we employ consists of the selection equation

$$z_i^* = \gamma_1 + \gamma_2 w_i + u_i = \gamma_1 + 1 w_i + u_i, \quad i = 1, \dots, N \tag{6.1}$$

The sample size N = 100 and 400. The value of the parameter γ_1 controls the degree of censoring. Following Zuehlke and Zeman (1991) we specifiy $\gamma_1 = [-.96, 0 \text{ or } .96]$, which correspond to expected sub-samples of size n equaling 25%, 50% and 75% of N, given that $u_i \sim nid(0,1)$

The regression equation of interest is

$$y_i = \beta_1 + \beta_2 x_i + e_i = 100 + 1x_i + e_i, \quad i = 1, \dots, n, \quad N > n \tag{6.2}$$

We assume the regression error $e_i \sim nid(0, \sigma_e = 1)$. The selection bias is controlled through the error correlation $\rho = [0, .5 \text{ or } 1]$ in (2.4).

The remaining element of the experiment is the correlation between the regressor *w* in the selection equation and the regressor *x* in the regression equation. Following Nawata and Nagase (1996) we specify this correlation to be $\rho_{xw} = [.90, .95 \text{ or } 1]$. Varying γ_1, ρ and ρ_{xw} as describes gives us 27 Monte Carlo design points for each sample size *N*. For each design point and sample size we generate 500 Monte Carlo samples. We use 200 bootstrap samples.

The Monte Carlo results we report are designed to measure

(i) The finite sample accuracy of nominal standard errors of the Heckit estimator based on the alternative covariance matrix estimators described in section 4. The accuracy is evaluated by comparing the average nominal standard errors to the Monte Carlo estimates' standard deviations.

(ii) The size of *t*-tests of significance based on each standard error method for the slope parameter in the regression equation, based on the usual asymptotic critical values ± 1.645 for tests of nominal 10% size and ± 1.96 for nominal 5% size tests. That is, we compute the *t*-statistic

$$t_m = \frac{\hat{\beta}_m - \beta}{se_m} \tag{6.3}$$

where $\hat{\beta}_m$ and β are Heckit estimates for parameters in the augmented regression equation from the Monte Carlo sample and the true parameter value, respectively, and se_m is a standard error computed from the Monte Carlo sample. The true null hypothesis is rejected if $|t_m| \geq t_c$, with t_c = 1.645 or 1.96. The test size is the percentage of times in 500 Monte Carlo samples we reject the null hypothesis. Clearly we prefer tests of an assumed nominal size of .05 or .10 to have actual size close to those values.

(iii) The size of *t*-tests of significance based on bootstrapping critical values of the distribution of the "pivotal" statistic associated with each standard error method. Specifically, in each Monte Carlo sample *m* we resample, with replacement, from the rows of the data matrix $[y\ z\ x\ w]$ to obtain a bootstrap sample *b* of size *N*. A pivotal statistic is obtained for each of 200 bootstrap samples by computing

$$t_b = \frac{\hat{\beta}_b - \hat{\beta}_m}{se_b} \tag{6.4}$$

where $\hat{\beta}_b$ and $\hat{\beta}_m$ are Heckit estimates for parameters in the augmented regression equation from the bootstrap and Monte Carlo samples respectively, and se_b is a standard error computed from the bootstrap sample. Recall that for the m'th Monte Carlo sample, the estimate $\hat{\beta}_m$ based on the full sample is the "true in the sample" value and thus using it to center the t-statistic makes it asymptotically pivotal. The bootstrap critical values are obtained by taking the .90 and .95 percentiles of empirical distribution of $|t_b|$, for the nominal 10% and 5% size tests, respectively. The size of a two-tailed test using the bootstrap critical values is computed by calculating the percentage of the time the t-statistic in (6.3) falls above or below the computed percentiles. That is, a nominal 5% test is obtained by rejecting the (true) null hypothesis if $|t_m| \ge t_c$, with t_c being the bootstrap critical value, which is the 95th percentile of the sorted $|t_b|$ values. The actual test size using the bootstrap critical values is the percentage of times in 500 Monte Carlo samples we reject the null hypothesis.

7. The Monte Carlo Results

When viewing the tabled results recall that the interpretation of the parameters controlling the experimental design:

- γ_1 controls the degree of censoring in the sample. When $\gamma_1 = -0.96, 0.0, 0.96$ the sizes of the selected sub-samples of size n are approximately 25%, 50% and 75 %, respectively, of the original sample of N observations.
- ρ is the correlation between the errors in the regression and selection equation. It takes the values 0.0 (under which OLS on the sub-sample is BLUE), 0.5 and 1.0. As the value of ρ increases the magnitude of the selection problem, as measured by $\beta_\lambda = \rho \cdot \sigma_e$, increases.
- ρ_{xw} is the correlation between the explanatory variables in the selection equation and the regression model of interest. The parameter ρ_{xw} takes the values 0.90, 0.95 and 1.0. The latter case represents the undesirable situation in which the explanatory variables in the selection equation are identical to those in the regression equation.
- The columns labeled OLS are results that use the standard errors from V_{OLS} in (4.1). Columns labeled HCE0 contain results based on the covariance matrix estimator V_{HC_0}, and so on.

Table 1.a reports for $N = 100$ the average of the nominal (asymptotic) standard errors relative to the true estimator standard error as measured by the sampling variation of the Heckit estimator of the parameter β_2. The

column labeled MCSE is the standard error of the estimates in the Monte Carlo simulation. This is the true finite sample variability. Several cases are of interest:

- When $\gamma_1 = -0.96$ and ρ_{xw}=1 the Monte Carlo standard errors become significantly larger than in the other cases. This is an indication of the general difficulties encountered when the independent variables used in each of the two-steps are the same, especially if the sample for the regression is heavily censored.
- When $\rho = 0$, the OLS standard error should accurately reflect the sampling variation of the Heckit estimator since it is BLUE. For the extreme design point $\gamma_1 = -0.96$ or 0.0, and $\rho_{xw} = 1$, the OLS estimator understates the Heckit estimator variability. The other estimators HCE3, HECK (and its near image, MT) and BOOT all overestimate the variability in this case. Bootstrap standard errors measure finite sample variation quite well when $\rho = 0$ except for the case when $\rho_{xw} = 1$. Given $\rho = 0$ and $\rho_{xw} = 1$, the performance of HCE3, HECK and BOOT improve as the degree of censoring diminishes.
- Of the two White HCE estimators, the usual HCE0 tends to understate the finite sample variation of Heckit and HCE3 tends to overstate it, but by a relatively small margin. Both estimators improve as censoring is reduced.
- When the degree of censoring is large the Murphy-Topel and Robust Murphy-Topel estimators are not reliable for this sample size. For less censored cases, in this example, the usual MT estimator performs better than the robust version, RMT. Indeed, the MT estimator mirrors the results of HECK even in this small sample.
- The LEE estimator is appropriate when errors are non-normal, but in this case it severely understates the true sample variation of Heckit when censoring is severe or moderate. In our experiment we did not consider non-normal errors or the robustness of alternative standard error estimators.
- Comparing HECK and BOOT, we see that HECK does not fare well when censoring is large, but it becomes progressively better as the degree of censoring declines. When censoring is severe ($\gamma_1 = -0.96$) the bootstrap estimator of sampling variation is strongly preferred, except for $\rho_{xw} = 1$. If censoring is moderate ($\gamma_1 = 0.0$) BOOT is preferred when $\rho = 0$. For other cases of moderate censoring HECK and BOOT are similar, except in the unusual case $\rho = 0.5$ and $\rho_{xw} = .95$. In the case with mild censoring ($\gamma_1 =$ 0.96) HECK and BOOT perform similarly.

Table 1.b reports for $N = 400$ the same information as Table 1.a. Having more data is better, and all estimators more accurately capture finite sampling variation of the Heckit estimator. Furthermore, the same relative relationships exist between them. When the degree of censoring is large, HECK does not do as well as BOOT or HCE3. The bootstrap standard error BOOT also provides a close measure of finite sample variation in other designs. The exceptions are the designs $\gamma_1 = 0.0$, $\rho = 0.5$ and $\rho_{xw} = 0.90$ and $\gamma_1 = 0.96$, $\rho = 0$ and $\rho_{xw} = 1$. In these anomalous cases all standard errors understate the true variation.

Table 2.a and 2.b report for $N = 100$ and $N = 400$, respectively, the $\alpha = .05$ critical values for asymptotic tests. These critical values were computed by sorting the absolute values of the difference between the Monte Carlo Heckit estimates and the true parameter value divided by the relevant asymptotic standard error. Given the evidence in Table 1 it is not surprising that these critical values are not ±1.96, as asymptotic theory would predict.

- When $N = 100$ and $\gamma_1 = -0.96$ (censoring is severe), the empirically determined critical values are quite different from ±1.96 for all estimators in almost all cases. This is especially true for $\rho_{xw} = 1$.
- It is worth noting, however, that even when $N = 100$, for $\gamma_1 = 0.0$ or 0.96 (less severe censoring) and $\rho = 0.5$ or 1.0 (moderate or severe selection bias) with $\rho_{xw} = 0.90$ or $\rho_{xw} = 0.95$, then the usual Heckit standard error, HCE3 and BOOT are in a range consistent with the usual 1.96.
- For $N = 100$, using the bootstrap standard error, BOOT, produces nominal critical values closer to ±1.96 than HECK when censoring is severe, but the results illustrate that simply using bootstrap standard errors does not solve the problem of testing in small samples.
- When $N = 400$ using the bootstrap standard error yields a *t*-statistic whose critical values are close to ±1.96 even in the severe censoring case. Compare, for example, the critical values based on BOOT with those from HECK (and the others) when censoring is severe and $\rho_{xw} = 1$. Blindly comparing a Heckit *t*-statistic to ±1.96 is a less troubling when using the bootstrap standard error than the alternatives. In this context HCE3 also provides a definite improvement over HECK.
- When $\gamma_1 = 0.0$ or 0.96 (less severe censoring) and $\rho_{xw} = 1$, BOOT, HCE3 and other estimators exhibit critical values that are larger than 1.96.

In Tables 3.a and 3.b we report for $N = 100$ and $N = 400$, respectively, the sizes of the nominal $\alpha = .05$ asymptotic tests. These sizes are computed by calculating the percentage of true null hypotheses rejected in the Monte Carlo experiment, when the *t*-statistics based on alternative standard error estimators are compared to ±1.96.

- When $N = 100$, HECK is does not produce tests of predicted size when censoring is severe ($\gamma_1 = -0.96$) or if $\rho_{xw} = 1$. Using the bootstrap standard error in these cases is certainly better, except when $\rho_{xw} = 1$. In other cases using the bootstrap standard error is a good choice; comparing the resulting *t*-statistic to ±1.96 leads to tests of approximately the correct size. The exception to this rule is when censoring is moderate ($\gamma_1 = 0.0$), $\rho = 0$ (so that OLS is BLUE) and $\rho_{xw} = 1$.
- When $N = 100$, HCE3 results in less size distortion than HCE0, which rejects too frequently, a finding consistent with its "too small" standard errors noted in Table 1.a. If the usual critical values ±1.96 are employed, then using standard error HCE3 by far the best alternative in our simulation when censoring is severe and $\rho_{xw} = 1$. In fact using HCE3 seems a good choice overall, with its only real problem occurring when censoring is moderate and the error correlation is $\rho = 0.5$ and $\rho_{xw} = 1$.
- When $N = 400$ using *t*-statistics based on the bootstrap standard error yield sizes closer to the nominal $\alpha = .05$ value than using HECK, especially for $\rho_{xw} = 1$.
- MT and HECK perform similarly. RMT under-rejects in all cases when censoring is not severe. HCE3 is again preferable to HCE0 over all degrees of censoring and in virtually all cases.

Tables 4.a and 4.b contain the sizes of the nominal $\alpha = .05$ tests based on bootstrap critical values using the pivotal statistic in equation (6.4). The usefulness of bootstrapping critical values is immediately obvious.

- When $N = 100$, using bootstrap critical values for the *t*-statistic based on HECK provides tests of close to proper size in all cases other than $\rho_{xw} = 1$. The Murphy-Topel (MT) results are very similar to HECK and the robust version RMT is close, though with a few more extreme values.
- HCE3 seems to perform relatively well with the bootstrap critical values, even when $\rho_{xw} = 1$ and censoring is severe ($\gamma_1 = -.96$) or not ($\gamma_1 = 0.96$).

- If $N = 400$, using the bootstrap critical values with HCE3 again seems a good alternative across all designs. For smaller amounts of censoring ($\gamma_1 = 0.96$) there is very little difference across the empirical sizes no matter which standard error is used.
- If $N = 400$, using HECK with bootstrap critical values is a good choice when censoring is not severe.

8. *Summary and Conclusions*

Our research goals for this work were:

- First, how do the alternative versions of asymptotic variance-covariance matrices used in selectivity models capture the finite sample variability of the Heckit two-step estimator? The answer depends on the degree of censoring and on whether the explanatory variables in the selection and regression equation differ or not. With severe censoring and if the explanatory variables in the two equations are identical, then none of the asymptotic standard error formulations is reliable in small samples. In larger samples the bootstrap does a good job in reflecting estimator variability, as does the White HCE in which the diagonal matrix of squared residuals is weighted, element-wise, by the squared reciprocals of the diagonal of an identity less the "hat" matrix, which we have called V_{HC_3}, following Davidson and MacKinnon (1993). These are both better choices than the usual Heckit asymptotic covariance matrix in the sense of coming closer to the finite sample variation or being slightly conservative. If selection and regression equations have different variables, and if the sample is small, the bootstrap does relatively well if censoring is strong or moderate.
- Second, is it possible to use bootstrapping to improve finite sample inference? Three aspects of this question are:
    - Do bootstrap standard errors match finite sample variability better than nominal standard errors computed from asymptotic covariance matrices? As noted in the previous summary point, the answer to this is yes, unless censoring is severe, and the explanatory variables in the selection and regression equations are the same.
    - Do critical values of test statistics generated by bootstrapping pivotal *t*-statistics lead to better test size than those based on usual asymptotic theory? The answer is yes, and this is the most interesting and useful finding in the current research.

    - If censoring is severe, or when the sample is small, we recommend using HCE3 with bootstrapped critical values when carrying out tests about the slope parameters, or when constructing interval estimates.
    - If the sample is large and censoring is moderate or less severe, then using the standard Heckit standard errors with bootstrapped critical values is satisfactory.
- Does modern software make obtaining bootstrap standard errors and critical values feasible for empirical researchers? Yes. Sample programs in the appendix show that it is simple to obtain bootstrap standard errors, and almost as simple to obtain bootstrap critical values with Heckit or HCE3 standard errors.

9. References

Amemiya, T., 1984. Tobit Models: A Survey. Journal of Econometric 24, 3--61.

Amemiya, T., 1985. Advanced Econometrics. Cambridge, MA: Harvard University Press.

Davidson, R. and MacKinnon, J., 1993. Estimation and Inference in Econometrics, New York: Oxford University Press.

Deis, D. and Hill, R.C., 1998. An Application of the Bootstrap Method to the Simultaneous Equations Model of the Demand and Supply of Audit Services. Contemporary Accounting Research 15, 83--99.

Eicker, F., 1963. Asymptotic Normality and Consistency of the Least Squares Estimators For Families of Linear Regressions. Annals of Mathematical Statistics 34, 447--456.

Ermisch, J.F. and Wright, R.E., 1993. Wage Offers and Full-time and Part-time Employment by British Women. The Journal of Human Resources 28, 111--133.

Freedman, D. and Peters, S., 1984a. Bootstrapping and Econometric Model: Some Empirical Results. Journal of Business & Economic Statistics 2, 150--158.

Freedman, D. and Peters, S., 1984b. Bootstrapping a Regression Equation: Some Empirical Results. Journal of the American Statistical Association 79, 97--106.

Greene, W.H., 1981. Sample Selection Bias as a Specification Error: Comment. Econometrica 49, 795--798.

Greene, W.H., 1997. Econometric Analysis, 3rd Edition, Upper Saddle River, NJ: Prentice-Hall.

Griffiths, W.E., Hill, R.C., and Pope, P., 1987. Small Sample Properties of Probit Estimators. Journal of the American Statistical Association 73, 191--193.

Hardin, J.W., 2002. The Robust Variance Estimator for Two-Stage Models. The Stata Journal 2, 253--266.

Hardin, J.W., 2003. The Sandwich Estimate of Variance, in: Hill, R.C. and Fomby, T.B. (Eds.), Advances in Econometrics, Volume 17: Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later. North-Holland: Elsevier, pp. XXX--XXX.

Heckman, J.J., 1979. Sample Selection Bias as a Specification Error. Econometrica 47, 153--161.

Horowitz, J.L., 1997. Bootstrap Methods in Econometrics: Theory and Numerical Performance, in: Kreps, D.M. and Wallis, K.F. (Eds.), Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress, Vol. III. Cambridge: Cambridge University Press, pp. 188--222.

Horowitz, J.L. and Savin, N.E., 2000. Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach. Journal of Econometrics 95, 375--389.

Huber, P.J., 1967. The Behavior of Maximum Likelihood Estimation Under Nonstandard Conditions, in: LeCam, L.M. and Neyman, J. (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1. Berkeley: University of California Press, pp. 221--233.

Hunter, L.W., 2000. What Determines Job Quality in Nursing Homes? Industrial and Labor Relations Review 53, 463--481.

Jeong, J. and Maddala, G.S., 1993. A Perspective on Application of Bootstrap Methods in Econometrics, in Maddala, G.S., Rao, C.R., and Vinod, H.D. (Eds.), Handbook of Statistics, Volume 11. New York: North-Holland, pp. 573--610.

Kauermann, G. and Carroll, R.J., 2001. A Note on the Efficiency of Sandwich Covariance Matrix Estimation. Journal of the American Statistical Association 96, 1387--1396.

Lee, L.-F., 1982. Some Approaches to the Correction of Selectivity Bias. Review of Economic Studies 49, 355--372.

Long, J.S. and Ervin, L.H., 2000. Correcting for Heteroscedasticity with Heteroscedasticity Consistent Standard Errors in the Linear Regression Model: Small Sample Considerations. American Statistician 54, 217--224.

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.

Murphy, K.M. and Topel, R.H., 1985. Estimation and Inference in Two-Step Econometric Models, Journal of Business & Economic Statistics 34, 370--379.

Nawata, K., 1993. A Note on the Estimation of Models with Sample Selection Biases. Economics Letters 42, 15-24.

Nawata, K., 1994. Estimation of Sample Selection Bias Models by the Maximum Likelihood Estimator and Heckman's Two-Step Estimator. Economics Letters 45, 33--40.

Nawata, K. and McAleer, M., 2001. Size Characteristics of Tests for Sample Selection Bias: A Monte Carlo Comparison and Empirical Example. Econometric Reviews 20, 105--112.

Nawata, K. and Nagase, N., 1996, Estimation of Sample Selection Bias Models. Econometric Reviews 15, 387--400.

Pindyck, R.S. and Rubinfeld, D.L., 1998. Econometric Models and Economic Forecasts, 4th Edition. Boston, MA: Irwin-McGraw-Hill.

Puri, M., 1996. Commerical Banks in Investment Banking. Conflict of Interest or Certification Role? Journal of Financial Economics 40, 373--401.

Puhani, P.A., 2000. The Heckman Correction for Sample Selection and its Critique. Journal of Economic Surveys 14, 53--68.

Marshall, R.C. and Zarkin, G.A., 1987. The Effect of Job Tenure on Wage Offers. Journal of Labor Economics 53, 301--324.

Simpson, W., 1986. Analysis of Part-time Pay in Canada. Canadian Journal of Economics 19, 798--807.

Vella, F., 1998. Estimating Models with Sample Selection Bias: A Survey. Journal of Human Resources 33, 127--172.

White, H.J., 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica 48, 817--838.

Wu, C. and Kwok, C.C.Y., 2002. Why Do US Firms Choose Global Equity Offerings? Finanacial Management, Summer, 47--65.

Zuehlke, T.W. and Zeman, A.R., 1991. A Comparison of Two-Stage Estimators of Censored Regression Models. Review of Economics and Statistics 73, 185--188.

Table 1.a Nominal standard errors relative to the Monte Carlo standard errors, N=100
Coefficient: β_2

γ_1	ρ	ρ_{xw}	MCSE	OLS	HCE0	HCE3	HECK	MT	RMT	LEE	Boot
-0.96	0.0	0.90	0.3879	0.9830	0.8865	1.0694	0.9277	0.9270	0.9971	0.8700	0.9923
-0.96	0.0	0.95	0.5654	0.9868	0.8852	1.0988	0.9571	0.9564	1.0808	0.8837	1.0284
-0.96	0.0	1.00	1.4435	0.8550	0.6240	1.2476	1.2703	1.2555	1.2485	0.9763	1.4686
-0.96	0.5	0.90	0.3796	1.0066	0.9107	1.0793	0.9557	0.9556	0.9779	0.9054	1.0346
-0.96	0.5	0.95	0.7730	1.0203	0.8775	1.1497	0.9063	0.9050	1.0484	0.8593	1.0210
-0.96	0.5	1.00	1.3504	0.8554	0.6831	1.0929	1.1962	1.1827	1.1726	0.8989	1.2761
-0.96	1.0	0.90	0.2679	0.9476	0.8611	1.0579	0.9103	0.9101	0.9738	0.8758	1.0019
-0.96	1.0	0.95	0.4568	0.9137	0.8295	1.0870	0.9117	0.9111	0.9903	0.8414	1.0435
-0.96	1.0	1.00	1.9669	0.8628	0.7620	1.0860	1.3601	1.3675	1.3373	0.9200	1.1929
0.00	0.0	0.90	0.3131	1.0177	0.9655	1.0705	0.9886	0.9877	1.1285	0.9635	1.0152
0.00	0.0	0.95	0.3901	1.0063	0.9362	1.0654	1.0259	1.0222	1.2529	0.9381	1.0186
0.00	0.0	1.00	0.7672	0.9378	0.8461	1.0414	1.1127	1.0966	1.0906	0.9610	1.0871
0.00	0.5	0.90	0.2966	0.9811	0.9370	1.0348	0.9503	0.9494	1.0354	0.9276	0.9869
0.00	0.5	0.95	0.2904	0.9389	0.7811	1.1728	1.0037	0.9982	1.3274	0.8900	1.1165
0.00	0.5	1.00	0.6089	0.8743	0.7671	1.0205	1.0587	1.0387	1.0435	0.8958	1.0490
0.00	1.0	0.90	0.2121	0.9640	0.9029	1.0333	0.9611	0.9558	1.0926	0.9356	0.9735
0.00	1.0	0.95	0.2728	0.9088	0.9199	1.0548	0.9937	0.9853	1.2130	0.9139	0.9897
0.00	1.0	1.00	0.5708	0.7867	0.7927	0.9392	0.9993	0.9769	0.9738	0.8800	0.9681
0.96	0.0	0.90	0.1996	1.0309	0.9808	1.0662	1.0213	1.0173	1.7107	0.9769	1.0060
0.96	0.0	0.95	0.2625	0.9917	0.9514	1.0307	0.9841	0.9792	1.2908	0.9456	0.9646
0.96	0.0	1.00	0.3075	1.0136	0.9637	1.0443	1.0560	1.0372	1.0508	1.0208	1.0291
0.96	0.5	0.90	0.1996	0.9957	0.9610	1.0288	0.9909	0.9881	1.4615	0.9650	0.9834
0.96	0.5	0.95	0.2585	1.0170	0.9638	1.0644	1.0111	1.0062	1.4529	0.9726	0.9961
0.96	0.5	1.00	0.2967	0.9830	0.9018	1.0746	1.0675	1.0553	1.1216	0.9842	1.0610
0.96	1.0	0.90	0.1888	1.0125	0.9813	1.0606	1.0228	1.0165	1.3934	0.9865	1.0029
0.96	1.0	0.95	0.1975	0.9735	0.9575	1.0591	1.0353	1.0165	1.6213	0.9888	0.9845
0.96	1.0	1.00	0.2725	0.9297	0.8682	1.0424	1.0678	1.0451	1.1670	0.9735	1.0310

Table 1.b Nominal standard errors relative to the Monte Carlo standard errors, N=400
Coefficient: β_2

γ_1	ρ	ρ_{xw}	MCSE	OLS	HCE0	HCE3	HECK	MT	RMT	LEE	Boot
-0.96	0.0	0.90	0.2388	0.9658	0.9419	0.9905	0.9517	0.9516	0.9662	0.9230	0.9695
-0.96	0.0	0.95	0.3108	1.0394	0.9988	1.0544	1.0250	1.0248	1.0603	0.9913	1.0274
-0.96	0.0	1.00	0.8865	1.0087	0.9258	1.0422	1.1326	1.1263	1.1234	0.9715	1.0391
-0.96	0.5	0.90	0.2047	0.9403	0.9186	0.9631	0.9302	0.9300	0.9466	0.9184	0.9404
-0.96	0.5	0.95	0.3032	0.9442	0.9164	0.9666	0.9307	0.9302	0.9552	0.9197	0.9478
-0.96	0.5	1.00	0.8982	0.9134	0.8637	0.9848	1.0635	1.0575	1.0532	0.9010	0.9758
-0.96	1.0	0.90	0.1440	0.8999	0.9024	0.9508	0.9319	0.9295	0.9361	0.8939	0.9311
-0.96	1.0	0.95	0.1933	1.0029	0.9542	1.0033	0.9718	0.9693	0.9833	0.9513	0.9786
-0.96	1.0	1.00	0.5204	0.8198	0.8906	1.0814	1.1485	1.1340	1.1334	0.9338	1.0567
0.00	0.0	0.90	0.1398	1.0055	0.9805	1.0144	1.0027	1.0026	1.0727	0.9813	0.9980
0.00	0.0	0.95	0.1853	1.0034	0.9890	1.0190	1.0005	1.0000	1.0517	0.9860	1.0030
0.00	0.0	1.00	0.2964	1.0171	0.9411	1.0613	1.0565	1.0538	1.0552	0.9676	1.0519
0.00	0.5	0.90	0.1405	0.9445	0.9344	0.9601	0.9472	0.9461	0.9871	0.9378	0.9497
0.00	0.5	0.95	0.1795	1.0264	1.0090	1.0423	1.0298	1.0284	1.1029	1.0018	1.0260
0.00	0.5	1.00	0.3006	1.0294	1.0116	1.0753	1.0810	1.0742	1.0740	1.0142	1.0578
0.00	1.0	0.90	0.1063	1.0008	1.0015	1.0265	1.0159	1.0118	1.0206	1.0004	1.0200
0.00	1.0	0.95	0.1412	0.8934	0.9371	0.9723	0.9739	0.9634	1.0001	0.9400	0.9625
0.00	1.0	1.00	0.2503	0.7969	0.8545	0.9801	1.0071	0.9893	0.9970	0.8980	0.9633
0.96	0.0	0.90	0.1111	0.9841	0.9701	0.9894	0.9818	0.9817	1.1602	0.9661	0.9759
0.96	0.0	0.95	0.1138	1.0576	1.0454	1.0714	1.0580	1.0574	1.2216	1.0505	1.0575
0.96	0.0	1.00	0.1769	0.9365	0.9182	0.9487	0.9533	0.9502	0.9554	0.9369	0.9431
0.96	0.5	0.90	0.0997	0.9909	0.9797	1.0010	0.9955	0.9941	1.1672	0.9802	0.9893
0.96	0.5	0.95	0.1376	0.9956	0.9874	1.0095	1.0003	0.9975	1.1762	0.9831	1.0013
0.96	0.5	1.00	0.1643	1.0011	0.9894	1.0332	1.0421	1.0354	1.0474	1.0014	1.0290
0.96	1.0	0.90	0.0966	0.9825	0.9757	0.9941	0.9963	0.9887	1.0744	0.9811	0.9890
0.96	1.0	0.95	0.1079	0.8835	0.9273	0.9545	0.9596	0.9468	1.0662	0.9410	0.9567
0.96	1.0	1.00	0.1409	0.8950	0.9288	0.9600	0.9861	0.9672	0.9778	0.9635	0.9733

Table 2.a Nominal .05 critical values for asymptotic tests, N=100
Coefficient: β_2

γ_1	ρ	ρ_{xw}	OLS	HCE0	HCE3	HECK	MT	RMT	LEE	BOOT
-0.96	0.0	0.90	2.1564	2.4997	2.0921	2.2959	2.2960	2.2913	2.4987	2.2475
-0.96	0.0	0.95	1.9765	2.4784	2.0300	1.9034	1.9148	1.8836	2.4783	2.1307
-0.96	0.0	1.00	2.1854	3.3944	2.2879	1.0708	1.0896	1.0890	2.4424	1.4372
-0.495	0.5	0.90	2.0495	2.4514	2.0473	2.1605	2.1606	2.1153	2.4514	2.0689
-0.96	0.5	0.95	1.9881	2.5483	2.0967	2.4759	2.4722	2.4870	2.5119	2.1011
-0.96	0.5	1.00	2.1838	2.9830	2.1977	1.2258	1.2676	1.2799	2.2716	1.5520
-0.96	1.0	0.90	2.1766	2.5664	2.1281	2.2015	2.2045	1.9579	2.5658	2.1364
-0.96	1.0	0.95	2.1372	2.5000	1.9635	2.0984	2.0981	1.9684	2.4867	2.0471
-0.96	1.0	1.00	2.3388	2.8207	2.1199	1.2288	1.2198	1.2635	2.2081	1.6392
0.00	0.0	0.90	1.9518	2.0702	1.8953	1.9681	1.9701	1.8695	2.0635	2.0093
0.00	0.0	0.95	2.1152	2.2941	2.0492	1.8647	1.8697	1.7915	2.2503	2.0836
0.00	0.0	1.00	2.0411	2.3298	1.8919	1.2812	1.2866	1.3344	1.9916	1.6607
0.00	0.5	0.90	2.0201	2.1913	2.0182	2.0889	2.0950	1.9332	2.1881	2.1184
0.00	0.5	0.95	2.0782	2.8374	2.1033	1.9094	1.9186	1.6178	2.7738	1.9887
0.00	0.5	1.00	2.3039	2.7945	2.3425	1.5838	1.6462	1.6463	2.4435	2.0065
0.00	1.0	0.90	2.0709	2.3460	2.0931	2.0664	2.0850	2.0038	2.3231	2.1118
0.00	1.0	0.95	2.1359	2.2299	1.9406	1.9634	1.9823	1.6567	2.2175	2.0637
0.00	1.0	1.00	2.3178	2.7172	2.3249	2.0401	2.0613	1.9249	2.3276	2.1305
0.96	0.0	0.90	1.9096	2.0825	1.9320	1.8630	1.8879	1.6744	2.0453	2.0021
0.96	0.0	0.95	1.9942	2.1298	1.9958	1.9569	1.9955	1.8357	2.0626	2.1470
0.96	0.0	1.00	1.9253	2.1346	1.9263	1.6599	1.7813	1.7885	1.9637	1.9374
0.96	0.5	0.90	2.1139	2.2656	2.1211	2.1102	2.1126	1.7838	2.2339	2.2155
0.96	0.5	0.95	1.9368	1.9385	1.7726	1.9358	1.9411	1.6993	1.9153	1.9506
0.96	0.5	1.00	2.0130	2.2224	1.9966	1.7260	1.7470	1.6652	2.0930	1.8365
0.96	1.0	0.90	2.0607	2.1647	2.0214	2.0333	2.0370	1.5831	2.1646	2.0948
0.96	1.0	0.95	1.9905	2.0239	1.8937	1.8877	1.9386	1.4887	1.9670	1.9886
0.96	1.0	1.00	2.1119	2.3250	2.0976	1.8567	1.8631	1.7598	2.1743	2.0130

Table 2.b Nominal .05 critical values for asymptotic tests, N=400
Coefficient: β_2

γ_1	ρ	ρ_{xw}	OLS	HCE0	HCE3	HECK	MT	RMT	LEE	BOOT
-0.96	0.0	0.90	2.0500	2.1451	2.0523	2.0785	2.0795	2.0447	2.1435	2.1276
-0.96	0.0	0.95	1.8944	2.0885	1.9700	1.9001	1.8986	1.8787	2.0874	2.0033
-0.96	0.0	1.00	1.8752	2.2606	2.0678	1.3247	1.3388	1.3454	2.1320	1.9484
-0.96	0.5	0.90	2.1030	2.1336	2.0331	2.1268	2.1280	2.0869	2.1313	2.0827
-0.96	0.5	0.95	2.1649	2.1965	2.0492	2.1724	2.1748	2.1238	2.1959	2.1395
-0.592	0.5	1.00	2.0592	2.2756	2.0304	1.4440	1.4656	1.4657	2.0921	1.9184
-0.96	1.0	0.90	2.2114	2.2487	2.1376	2.1462	2.1491	2.1304	2.2427	2.2424
-0.96	1.0	0.95	2.0727	2.1301	2.0216	2.1494	2.1520	2.1324	2.1224	2.1424
-0.96	1.0	1.00	2.3443	2.4222	2.0593	1.8965	1.8980	1.8969	2.2716	1.9422
0.00	0.0	0.90	1.9659	2.0334	1.9766	1.9674	1.9665	1.9243	2.0330	1.9672
0.00	0.0	0.95	1.8788	1.9846	1.9231	1.8731	1.8791	1.8339	1.9735	1.9741
0.00	0.0	1.00	1.9069	2.1461	2.0072	1.6711	1.6868	1.6884	2.1198	1.9156
0.00	0.5	0.90	2.0941	2.1415	2.0938	2.0622	2.0679	1.9699	2.1296	2.1316
0.00	0.5	0.95	1.9059	1.9633	1.9035	1.9073	1.9237	1.7781	1.9625	1.9302
0.00	0.5	1.00	1.9118	1.9348	1.8304	1.7409	1.7682	1.7681	1.9340	1.8751
0.00	1.0	0.90	1.9238	1.8755	1.8320	1.9160	1.9227	1.8550	1.8742	1.8836
0.00	1.0	0.95	2.1775	2.2601	2.1856	2.0703	2.0854	1.9107	2.2142	2.1287
0.00	1.0	1.00	2.3878	2.5730	2.2710	1.9941	2.0103	1.9815	2.4375	2.1631
0.96	0.0	0.90	1.9685	2.0711	2.0390	1.9733	1.9735	1.8247	2.0570	2.0502
0.96	0.0	0.95	1.9170	1.8881	1.8575	1.8954	1.9034	1.6659	1.8753	1.8722
0.96	0.0	1.00	2.1018	2.2210	2.1754	1.9464	1.9670	1.9777	2.1772	2.0706
0.96	0.5	0.90	1.9689	2.0432	2.0042	1.9731	1.9788	1.7784	2.0251	2.0590
0.96	0.5	0.95	1.9287	2.0100	1.9722	1.9223	1.9223	1.8189	2.0015	2.0391
0.96	0.5	1.00	1.9138	2.0252	1.9444	1.7525	1.7698	1.7754	1.9883	1.9547
0.96	1.0	0.90	1.9865	1.8974	1.8620	1.9589	1.9754	1.8189	1.8956	1.9626
0.96	1.0	0.95	2.1878	2.0817	2.0182	2.0688	2.0788	1.8654	2.0299	2.0396
0.96	1.0	1.00	2.1758	2.2229	2.1715	2.0610	2.0626	2.0304	2.1527	2.1981

Table 3.a Sizes of the nominal .05 asymptotic tests, N=100
Coefficient: β_2

γ_1	ρ	ρ_{xw}	OLS	HCE0	HCE3	HECK	MT	RMT	LEE	BOOT
-0.96	0.0	0.90	0.0780	0.1180	0.0580	0.0940	0.0940	0.0900	0.1180	0.0920
-0.96	0.0	0.95	0.0540	0.0940	0.0600	0.0380	0.0380	0.0380	0.0940	0.0680
-0.96	0.0	1.00	0.0720	0.1940	0.0740	0.0000	0.0000	0.0000	0.1360	0.0040
-0.96	0.5	0.90	0.0680	0.1020	0.0580	0.0760	0.0760	0.0700	0.1020	0.0640
-0.96	0.5	0.95	0.0560	0.1000	0.0500	0.1040	0.1040	0.1040	0.0980	0.0640
-0.96	0.5	1.00	0.0740	0.1500	0.0700	0.0020	0.0020	0.0020	0.0900	0.0180
-0.96	1.0	0.90	0.0660	0.1000	0.0680	0.0680	0.0680	0.0480	0.1000	0.0660
-0.96	1.0	0.95	0.0760	0.1300	0.0500	0.0680	0.0680	0.0500	0.1300	0.0540
-0.96	1.0	1.00	0.0940	0.1640	0.0620	0.0000	0.0020	0.0020	0.0940	0.0200
0.00	0.0	0.90	0.0480	0.0660	0.0460	0.0500	0.0500	0.0420	0.0660	0.0600
0.00	0.0	0.95	0.0660	0.0880	0.0580	0.0400	0.0420	0.0320	0.0820	0.0620
0.00	0.0	1.00	0.0540	0.0900	0.0440	0.0000	0.0000	0.0000	0.0520	0.0100
0.00	0.5	0.90	0.0580	0.0740	0.0500	0.0600	0.0600	0.0480	0.0740	0.0660
0.00	0.5	0.95	0.0680	0.1620	0.0700	0.0400	0.0460	0.0200	0.1540	0.0500
0.00	0.5	1.00	0.0780	0.1580	0.1000	0.0100	0.0120	0.0120	0.1160	0.0520
0.00	1.0	0.90	0.0700	0.1060	0.0620	0.0720	0.0720	0.0580	0.1000	0.0760
0.00	1.0	0.95	0.0720	0.0820	0.0420	0.0500	0.0540	0.0280	0.0780	0.0540
0.00	1.0	1.00	0.1140	0.1360	0.0880	0.0600	0.0620	0.0440	0.1020	0.0620
0.96	0.0	0.90	0.0420	0.0640	0.0440	0.0440	0.0440	0.0240	0.0580	0.0640
0.96	0.0	0.95	0.0520	0.0760	0.0500	0.0460	0.0540	0.0380	0.0620	0.0700
0.96	0.0	1.00	0.0420	0.0660	0.0460	0.0140	0.0180	0.0260	0.0500	0.0460
0.96	0.5	0.90	0.0680	0.0860	0.0680	0.0640	0.0640	0.0340	0.0820	0.0800
0.96	0.5	0.95	0.0480	0.0460	0.0380	0.0440	0.0460	0.0300	0.0420	0.0480
0.96	0.5	1.00	0.0540	0.0920	0.0520	0.0140	0.0160	0.0080	0.0700	0.0360
0.96	1.0	0.90	0.0620	0.0680	0.0540	0.0640	0.0640	0.0180	0.0660	0.0600
0.96	1.0	0.95	0.0560	0.0640	0.0420	0.0460	0.0460	0.0060	0.0520	0.0520
0.96	1.0	1.00	0.0680	0.1140	0.0620	0.0360	0.0360	0.0200	0.0860	0.0500

Table 3.b Sizes of the nominal .05 asymptotic tests, N=400
Coefficient: β_2

γ_1	ρ	ρ_{xw}	OLS	HCE0	HCE3	HECK	MT	RMT	LEE	BOOT
-0.96	0.0	0.90	0.0580	0.0740	0.0560	0.0620	0.0620	0.0580	0.0740	0.0700
-0.96	0.0	0.95	0.0460	0.0680	0.0500	0.0460	0.0460	0.0420	0.0680	0.0520
-0.96	0.0	1.00	0.0460	0.0880	0.0580	0.0000	0.0000	0.0000	0.0720	0.0480
-0.96	0.5	0.90	0.0740	0.0760	0.0540	0.0740	0.0740	0.0680	0.0760	0.0680
-0.96	0.5	0.95	0.0620	0.0780	0.0600	0.0680	0.0680	0.0680	0.0780	0.0720
-0.96	0.5	1.00	0.0560	0.0860	0.0520	0.0000	0.0020	0.0020	0.0680	0.0460
-0.96	1.0	0.90	0.0780	0.0880	0.0780	0.0660	0.0660	0.0600	0.0880	0.0800
-0.96	1.0	0.95	0.0600	0.0700	0.0540	0.0620	0.0640	0.0640	0.0700	0.0620
-0.96	1.0	1.00	0.1140	0.1020	0.0600	0.0400	0.0400	0.0380	0.0800	0.0460
0.00	0.0	0.90	0.0500	0.0620	0.0540	0.0520	0.0520	0.0420	0.0620	0.0500
0.00	0.0	0.95	0.0420	0.0540	0.0420	0.0380	0.0380	0.0360	0.0520	0.0500
0.00	0.0	1.00	0.0420	0.0660	0.0500	0.0120	0.0160	0.0180	0.0640	0.0460
0.00	0.5	0.90	0.0640	0.0720	0.0680	0.0600	0.0620	0.0520	0.0720	0.0760
0.00	0.5	0.95	0.0480	0.0500	0.0460	0.0460	0.0460	0.0320	0.0500	0.0480
0.00	0.5	1.00	0.0400	0.0480	0.0400	0.0160	0.0200	0.0200	0.0460	0.0340
0.00	1.0	0.90	0.0420	0.0400	0.0360	0.0360	0.0380	0.0300	0.0400	0.0420
0.00	1.0	0.95	0.0800	0.0780	0.0700	0.0540	0.0560	0.0440	0.0720	0.0640
0.00	1.0	1.00	0.1140	0.1260	0.0980	0.0540	0.0560	0.0520	0.1080	0.0800
0.96	0.0	0.90	0.0520	0.0640	0.0580	0.0520	0.0520	0.0320	0.0640	0.0660
0.96	0.0	0.95	0.0440	0.0460	0.0420	0.0440	0.0440	0.0260	0.0460	0.0380
0.96	0.0	1.00	0.0600	0.0720	0.0640	0.0480	0.0500	0.0500	0.0620	0.0580
0.96	0.5	0.90	0.0500	0.0580	0.0540	0.0500	0.0500	0.0240	0.0560	0.0600
0.96	0.5	0.95	0.0460	0.0540	0.0520	0.0420	0.0440	0.0380	0.0540	0.0540
0.96	0.5	1.00	0.0400	0.0620	0.0480	0.0260	0.0260	0.0240	0.0520	0.0460
0.96	1.0	0.90	0.0540	0.0480	0.0480	0.0480	0.0500	0.0400	0.0480	0.0500
0.96	1.0	0.95	0.0920	0.0800	0.0640	0.0700	0.0740	0.0300	0.0620	0.0640
0.96	1.0	1.00	0.0820	0.0800	0.0720	0.0560	0.0600	0.0560	0.0680	0.0700

Table 4.a Sizes of the nominal .05 bootstrap tests, N=100
Coefficient: β_2

γ_1	ρ	ρ_{xw}	OLS	HCE0	HCE3	HECK	MT	RMT	LEE
-0.96	0.0	0.90	0.0740	0.0620	0.0620	0.0720	0.0720	0.0860	0.0660
-0.96	0.0	0.95	0.0520	0.0400	0.0460	0.0480	0.0480	0.0780	0.0480
-0.96	0.0	1.00	0.0400	0.0600	0.0420	0.0100	0.0120	0.0120	0.0680
-0.96	0.5	0.90	0.0540	0.0560	0.0580	0.0520	0.0520	0.0560	0.0560
-0.96	0.5	0.95	0.0440	0.0560	0.0580	0.0640	0.0640	0.1000	0.0560
-0.96	0.5	1.00	0.0380	0.0540	0.0440	0.0220	0.0260	0.0380	0.0460
-0.96	1.0	0.90	0.0540	0.0620	0.0640	0.0620	0.0620	0.0580	0.0620
-0.96	1.0	0.95	0.0380	0.0380	0.0520	0.0640	0.0620	0.0680	0.0400
-0.96	1.0	1.00	0.0320	0.0400	0.0400	0.0200	0.0220	0.0260	0.0460
0.00	0.0	0.90	0.0540	0.0420	0.0420	0.0500	0.0480	0.0520	0.0440
0.00	0.0	0.95	0.0600	0.0460	0.0440	0.0640	0.0640	0.0900	0.0500
0.00	0.0	1.00	0.0300	0.0220	0.0200	0.0020	0.0040	0.0060	0.0180
0.00	0.5	0.90	0.0600	0.0480	0.0500	0.0540	0.0520	0.0560	0.0460
0.00	0.5	0.95	0.0860	0.1060	0.0800	0.0860	0.0860	0.0800	0.1180
0.00	0.5	1.00	0.0700	0.0780	0.0720	0.0500	0.0540	0.0560	0.0820
0.00	1.0	0.90	0.0680	0.0680	0.0720	0.0680	0.0660	0.0800	0.0700
0.00	1.0	0.95	0.0560	0.0440	0.0440	0.0540	0.0540	0.0560	0.0480
0.00	1.0	1.00	0.0620	0.0560	0.0560	0.1060	0.0980	0.1020	0.0700
0.96	0.0	0.90	0.0560	0.0460	0.0480	0.0560	0.0520	0.0740	0.0500
0.96	0.0	0.95	0.0660	0.0600	0.0620	0.0640	0.0680	0.0820	0.0580
0.96	0.0	1.00	0.0340	0.0340	0.0320	0.0220	0.0280	0.0520	0.0240
0.96	0.5	0.90	0.0700	0.0620	0.0620	0.0680	0.0680	0.0820	0.0640
0.96	0.5	0.95	0.0400	0.0360	0.0340	0.0480	0.0480	0.0560	0.0400
0.96	0.5	1.00	0.0520	0.0440	0.0460	0.0580	0.0620	0.0860	0.0600
0.96	1.0	0.90	0.0520	0.0500	0.0480	0.0560	0.0560	0.0480	0.0500
0.96	1.0	0.95	0.0380	0.0360	0.0380	0.0420	0.0420	0.0580	0.0380
0.96	1.0	1.00	0.0660	0.0560	0.0560	0.0820	0.0800	0.0760	0.0760

Table 4.b Sizes of the nominal .05 bootstrap tests, N=400
Coefficient: β_2

γ_1	ρ	ρ_{xw}	OLS	HCE0	HCE3	HECK	MT	RMT	LEE
-0.96	0.0	0.90	0.0660	0.0660	0.0640	0.0640	0.0640	0.0680	0.0660
-0.96	0.0	0.95	0.0480	0.0500	0.0500	0.0460	0.0460	0.0500	0.0500
-0.96	0.0	1.00	0.0560	0.0480	0.0420	0.0280	0.0280	0.0280	0.0620
-0.96	0.5	0.90	0.0700	0.0560	0.0560	0.0700	0.0700	0.0640	0.0560
-0.96	0.5	0.95	0.0760	0.0720	0.0680	0.0780	0.0760	0.0760	0.0720
-0.96	0.5	1.00	0.0560	0.0500	0.0460	0.0340	0.0340	0.0340	0.0580
-0.96	1.0	0.90	0.0820	0.0640	0.0640	0.0760	0.0760	0.0840	0.0620
-0.96	1.0	0.95	0.0660	0.0600	0.0600	0.0640	0.0660	0.0640	0.0600
-0.96	1.0	1.00	0.0640	0.0560	0.0520	0.1260	0.1200	0.1140	0.0760
0.00	0.0	0.90	0.0480	0.0500	0.0520	0.0540	0.0540	0.0620	0.0500
0.00	0.0	0.95	0.0400	0.0320	0.0320	0.0480	0.0460	0.0520	0.0320
0.00	0.0	1.00	0.0560	0.0520	0.0520	0.0540	0.0540	0.0560	0.0560
0.00	0.5	0.90	0.0740	0.0760	0.0760	0.0740	0.0740	0.0680	0.0760
0.00	0.5	0.95	0.0480	0.0420	0.0420	0.0500	0.0500	0.0540	0.0420
0.00	0.5	1.00	0.0380	0.0300	0.0260	0.0620	0.0640	0.0620	0.0340
0.00	1.0	0.90	0.0340	0.0260	0.0260	0.0380	0.0380	0.0440	0.0260
0.00	1.0	0.95	0.0620	0.0580	0.0580	0.0680	0.0660	0.0740	0.0580
0.00	1.0	1.00	0.0960	0.0840	0.0840	0.0900	0.0880	0.0920	0.0920
0.96	0.0	0.90	0.0620	0.0600	0.0600	0.0620	0.0620	0.0740	0.0580
0.96	0.0	0.95	0.0380	0.0360	0.0360	0.0420	0.0400	0.0420	0.0360
0.96	0.0	1.00	0.0600	0.0520	0.0540	0.0700	0.0660	0.0680	0.0580
0.96	0.5	0.90	0.0600	0.0640	0.0640	0.0640	0.0640	0.0600	0.0640
0.96	0.5	0.95	0.0460	0.0420	0.0420	0.0520	0.0480	0.0620	0.0460
0.96	0.5	1.00	0.0520	0.0460	0.0460	0.0580	0.0520	0.0480	0.0520
0.96	1.0	0.90	0.0440	0.0400	0.0400	0.0440	0.0440	0.0540	0.0400
0.96	1.0	0.95	0.0580	0.0500	0.0500	0.0680	0.0620	0.0840	0.0480
0.96	1.0	1.00	0.0700	0.0640	0.0640	0.0760	0.0740	0.0740	0.0640

Computational Appendix

In this appendix we provide bare bones code in Stata 8.0 and LIMDEP 8.0 to carry out bootstrapping in the Heckit model. The data, code files and output are available at

`http://www.bus.lsu.edu/economics/faculty/chill/personal/Heckit.htm`

The Example

The data are from Thomas Mroz's widely cited example (1987, "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765-799) and consist of 753 observations from the 1987 Panel Study of Income Dynamics.

The linear model of interest has the logarithm of wage dependent on education, experience and age:

$$\ln(wage) = \beta_1 + \beta_2 ed + \beta_3 ex + \beta_4 ex^2 + \beta_5 age + e \tag{A.1}$$

However, many in the sample are observed with zero wages, suggesting that there may be sample selection bias; omitted factors in the wage equation may be correlated with a decision to work. The selection equation considered is:

$$z^* = \gamma_1 + \gamma_2 ed + \gamma_3 ex + \gamma_4 ex^2 + \gamma_5 age + \gamma_6 klt6 + v \tag{A.2}$$

where *klt6* is the number of children less than 6 years old present. The latent variable, z^*, is only observed in one of two states; either one works for wages (z=1) or does not (z=0). The correlation between v and e is ρ.

Stata 8.0: The .do file

```
clear
#delimit ;
use c:\stata\mroz;
heckman lwage educ exper expersq age,
   select(educ exper expersq age kidslt6) twostep;

scalar b1=_b[exper];
scalar se1=_se[exper];

di "   Coefficient b1 " b1   "  Standard Error  "  se1;
set seed 1431241 ;
bootstrap "heckman lwage educ exper expersq age,
    select (educ exper expersq age kidslt6) twostep
    "  "_b[exper] _se[exper] e(lambda) e(rho)",
    reps(400) level(95) saving(bsstboot) replace;
drop _all;

use bsstboot;

gen ttest1=(_bs_1 - b1)/ _bs_2;
gen abt=abs(ttest1);

_pctile abt, p(95);
di " The 97.5% critical value is " r(r1) ;
summarize ttest1 abt, detail ;
histogram ttest1, bin(20) normal ;
```

program notes: If the estimated ρ is outside [−1,1], Stata truncates the estimate and makes the error variance computation consistent. To obtain results comparable to LIMDEP, use the **`rhotrunc`** *option. We are not advocating one approach or the other.*

LIMDEP 8.0 command file

```
reset
read;file=c:\mroz.raw;nobs=753;nvar=22;names=inlf,hours,kidslt6,kidsge6,
          age,educ,wage,repwage,hushrs,husage,huseduc,huswage,faminc,mtr,motheduc,fatheduc,unem,
          city,exper,nwifeinc,lwage,expersq$

? M defines coefficient of interest.  NSAM is number of bootstraps
? In this example same variables are in selection and regression equations


namelist; w=one,educ,exper,expersq,age,kidslt6 $
namelist; x=one,educ,exper,expersq,age $
calc; m=3; nsam=400; k = col(x)+1$
matrix; boott = init(nsam,k,0); bootest = init(nsam,k,0) $

probit  ; lhs=inlf; rhs=w; hold $
select  ; lhs=lwage; rhs=x$
matrix  ; truebeta = b $

calc; iter=0 $

proc
calc; iter=iter+1 $
draw;n=753;replace$
probit  ; lhs=inlf; rhs=w; hold $
select  ; lhs=lwage; rhs=x$

? tvec = absolute values of t-stats centered at sample estimates

matrix; bhat = b; se = vecd(varb); sevec = esqr(se); rse = diri(sevec);
        dvec = bhat-truebeta; tvec = dirp(dvec,rse); tvec = dirp(tvec,tvec);
        tvec = esqr(tvec); boott(iter,*) = tvec'; bootest(iter,*) = bhat' $
endproc

exec; n=nsam; silent $
matrix; tm=part(boott,1,nsam,m,m); bm=part(bootest,1,nsam,m,m) $
sample; 1-400 $
create; boot_bm=bm$
dstat; rhs=boot_bm$

? To sort t-statistics must create variable

create; boot_tm=tm$
sort; lhs = boot_tm $
calc; nval = .95*nsam+1 $
matrix; list; tc = part(boot_tm,nval,nval,1,1) $

stop $
```

program notes: If the estimated ρ is outside [−1,1], LIMDEP truncates the estimate. Reminder--in Project Settings specify enough memory for sample size of 753 observations.