Design of Monte Carlo

Phoebus J. Dhrymes
May 2005 (version May 5)

First we begin by asking: how many variables do we need? Answer: suppose we want 3 x's and 3 more variables to make up the z's; so here we need 6 variables. Then we need 2 variables for u and v. So we need 8 in all.

Second: Generate a sequence of 8 independent N(0,1) from a standard random number generator; the length of the sequence depends on how large we desire the sample to be.

Third: Now we begin to build up the actual variables we are going to use. Thake the last two variables and define

$$u_t = a_1 x_{t7} + a_{12} x_{t8}, \quad v_t = x_{t8},$$

so that

$$Eu_t = Ev_t = 0, \quad \mathrm{Var}(u_t) = a_{11}^2 + a_{12}^2, \quad \mathrm{Cov}(u_t, vt) = a_{12}, \quad \mathrm{Var}(v_t) = 1,$$

as required by the paper.

Fourth: Now we construct the x's and the z's. Let $A_i,\ i = 1, 2,$ be two distinct 3×3 **nonsingular matrices**, and consider

$$x_{t\cdot} = (x_{t1}, x_{t2}, x_{t3})A_1, \quad p_{t\cdot} = (x_{t4}, x_{t5}, x_{t6})A_2, \quad z_{t\cdot} = (x_{t\cdot}, p_{t\cdot}).$$

Fifth: Choose the parameter vectors $\gamma_{\cdot 2}$ and $\gamma_{\cdot 1}$, as well as the scalar parameter β_{21} . Notice, that you already have,

$$\sigma_{11} = a_{11}^2 + a_{12}^2, \quad \sigma_{12} = a_{12}, \quad \sigma_{22} = 1, \quad \sigma_{11\cdot 2} = a_{11}^2.$$

Sixth: For any given draw compute $-z_{t\cdot}\gamma_{\cdot 2}$. If $v_t \geq z_{t\cdot}\gamma_{\cdot 2}$ **assign the value** 1, to the variable $\delta(t)$; if the complement occurs **assign the value** 0 to the variable $\delta(t)$.

Seventh: Construct now

$$y_{t1} = x_{t\cdot}\gamma_{\cdot 1} + \delta(t)\beta_{21} + u_t.$$

This will complete the construction of the sample in accordance to the requirements of the paper. You can then create, say, 1000 replications of samples 50, 100, 500, 1000. If you alter the values of a_{12} you alter the degree of correlation between the two equations, so it is desirable to do one for small, say, .1, one for medium, say .5, and one for large, say, .9 correlations. Noticed that a correlation is:

$$\mathrm{Corr}(u_t, v_t) = \frac{a_{12}}{\sqrt{a_{11}^2 + a_{12}^2}},$$

so it might be necessary to play around with the formula above to determine the proper values of a_{11}, a_{12}.

Now to answer a few questions that occur frequently:

i. In the process of obtaining the data, we do not need to construct y_{t2}; we only need to obtain $\delta(t)$. You then use the $\delta(t)$ you generated **both** to construct y_{t1} **and** to estimate $\gamma_{\cdot 2}$ by probit methods.

ii. No special variables are needed initially; we may want later to make the variables we use for the z's and the x's more realistic, but for the moment hold off.

iii. I'd rather you did not use the material in Eqs. (27) to estimate the parameter $\theta_{\cdot 1}$. I'd rather you use the material in Eqs. (33), (34),(35) and (36) of the paper. The covariance of the limiting distribution is given in Eq. (67) and the matrix M is defined in Eq. (42) of the paper, and its estimation is implicit in the previous discussion.

As for the outputs you need to prepare:

i. The mean of the parameter estimates, for $\gamma_{\cdot 1}$, β_{21} and $\sigma_{11\cdot 2}$, separately. Their bias, which is the difference between the means you have obtained and the true parameterswhich you know. Their (squre root of) mean square error.

ii. The mean of the covariance matrices you have estimated, from Eqs. (67) and (41), (42)and a comparison with the true covariance matrix, which you can easily construct from the way you obtained the data for the study, i.e. a comparison of $\hat{\sigma}_{11.2}\tilde{Q}^{-1}$ with $\sigma_{11.2}M^{-1}$.

iii. Do the same for the parameters and their covariance matrix as estimated by the "linear probability" method.

iv. To compare the two methods for efficiency: compute from the way you constructed the data the matrices $\frac{1}{\sigma_{11}}C_2$ from Eq. (72) and $\frac{1}{\sigma_{11.2}}C_1$ from Eq. (78), and their difference ω from Eq. (81). If the **characteristic roots** of the difference are **all positive** the estimator of the paper is efficient relative to the "linear probability" estimator.(The linear probability estimator is encouraged by Heckman (1978), p. 947.

Connection to Heckman's Model

The connection between the model in my paper and that in Heckman (1978) as given in his equations (1a), (1b) on p. 932 of that paper is as follows: (1) $\gamma_1 = 0$; (2) upon substituting from (1a) into (1b) we can rewrite the latter as

$$y^*_{2i} = X_{1i}\gamma_2\alpha_1 + \alpha_2 X_{2i} + (\beta_2 + \gamma_2\beta_1)d_i + (\gamma_2 U_{1i} + U_{2i});$$

if we apply his **principal assumption** and make the identification with the variables in Dhrymes:

$$D\, u_t = H\, U_{1i}, D\, v_t = H\, (\gamma_2 U_{1i}+U_{2i}), D\, z_{t\cdot} = H\, (X_{1i}, X_{2i}), D\, \gamma_{\cdot 2} = H\, (\gamma_2\alpha_1, \alpha_2)',$$

etc., the Heckman model is identical with that in Dhrymes with the further identification: D $x_{t\cdot} = X_{1i}$ H, D $\gamma_{\cdot 1} = \alpha_1$ H. Hence, we shall deal exclusively with the model in Dhrymes. In that model the basic error terms are the $w_{t\cdot} = (u_t, v_t)$ which are assumed i.i.d. mean zero and nonsingular covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & 1 \end{bmatrix}$$

Noting that

$$\Sigma^{-1} = \frac{1}{\sigma_{11}(1-\rho^2)} \begin{bmatrix} 1 & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix},$$

we can write the joint density of (u, v) in terms of the **conditional density of** v **given** u times the marginal density of u , i.e.

$$f(u_t, v_t) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}(v_t - \frac{\sigma_{12}}{\sigma_{11}}u_t)^2} \times \frac{1}{\sqrt{2\pi\sigma_{11}}} e^{-\frac{1}{2\sigma_{11}}u_t^2}.$$

Because the Jacobean of the transformation of u_t to y_t is one, we need only substitute to obtain the joint density of y_t and v_t as

$$f(y_t, v_t) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}[v_t - \frac{\sigma_{12}}{\sigma_{11}}(y_t - x_{t\cdot}\gamma_{\cdot 1} - \beta_{21}\delta(t))]^2} \times \frac{1}{\sqrt{2\pi\sigma_{11}}} e^{-\frac{1}{2\sigma_{11}}(y_t - x_{t\cdot}\gamma_{\cdot 1} - \beta_{21}\delta(t))^2}.$$

The probability that $\delta(t) = 1$ is found by

$$\int_{-z_{t\cdot}\gamma_{\cdot 2}}^{\infty} f(y_t, v_t) dv_t = \frac{1}{\sqrt{2\pi\sigma_{11}}} e^{-\frac{1}{2\sigma_{11}}(y_t - x_{t\cdot}\gamma_{\cdot 1} - \beta_{21}\delta(t))^2} \times I_1,$$

where

$$I_1 = \int_{-z_{t\cdot}\gamma_{\cdot 2}}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}(v_t - \frac{\sigma_{12}}{\sigma_{11}}(y_t - x_{t\cdot}\gamma_{\cdot 1} - \beta_{21}\delta(t))^2} dv_t.$$

To evaluate the last integral, make the change in variable

$$\xi = \frac{1}{(1-\rho^2)^{(1/2)}}(v_t - g(t)),$$

where

$$g(t) = \frac{\sigma_{12}}{\sigma_{11}}(y_t - x_{t\cdot}\gamma_{\cdot 1} - \beta_{21}\delta(t)).$$

Note further that the lower limit of the integral is thus transformed to

$$-h(t) = -\frac{z_{t\cdot}\gamma_{\cdot 2} + g(t)}{(1-\rho^2)^{(1/2)}}$$

and thus

$$\Pr(\delta(t) = 1) = \int_{-h(t)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1/2)\xi^2} d\xi = 1 - \Phi(-h(t)).$$

By complementation

$$\Pr(\delta(t) = 0) = \int_{-\infty)}^{-h(t)} \frac{1}{\sqrt{2\pi}} e^{-(1/2)\xi^2} d\xi = \Phi(-h(t)).$$

where Φ is the cdf of a unit normal. Bearing in mind that for a unit normal cdf $\Phi(-x) = 1 - \Phi(x)$, we can rewrite the joint density above more suggestively as

$$f(y_t, \delta(t)) = \left[\frac{1}{\sqrt{2\pi\sigma_{11}}} e^{-\frac{1}{2\sigma_{11}}(y_t - x_{t\cdot}\gamma_{\cdot 1} - \beta_{21}\delta(t))^2} \Phi(h(t))\right]^{\delta(t)}$$

$$\times \left[\frac{1}{\sqrt{2\pi\sigma_{11}}} e^{-\frac{1}{2\sigma_{11}}(y_t - x_{t\cdot}\gamma_{\cdot 1} - \beta_{21}\delta(t))^2} (1 - \Phi(h(t))\right]^{1-\delta(t)}.$$

The log-likelihood function may thus be written as

$$L = -\frac{T}{2}[\ln(2\pi) + \ln\sigma_{11}] + \sum_{t=1}^{T} \left[-\frac{1}{2\sigma_{11}}(y_t - x_{t\cdot}\gamma_{\cdot 1} - \beta_{21}\delta(t))^2 + \ln(\Phi(h(t))\right] \delta(t)$$

$$+ \sum_{t=1}^{T} \left[-\frac{1}{2\sigma_{11}}(y_t - x_{t\cdot}\gamma_{\cdot 1} - \beta_{21}\delta(t))^2 + \ln(1 - \Phi(h(t))\right] [1 - \delta(t)]$$

$$= -\frac{T}{2}[\ln(2\pi) + \ln\sigma_{11}] - \frac{1}{2\sigma_{11}}(y - X\eta_{\cdot 1})' D(y - X\eta_{\cdot 1})$$

$$+e' D\Phi_1 - \frac{1}{2\sigma_{11}}(y - X^*\eta_{\cdot 1})'(I - D)(y - X^*\eta_{\cdot 1}) + e' D\Phi_2, \quad \text{where}$$

$$\Phi_1 = (\ln\Phi(h(1)), \ldots, \ln\Phi(h(T)))', \quad \Phi_2 = (\ln(1 - \Phi(h(1))), \ldots, \ln(1 - \Phi(h(T))))'.$$

The first order conditions with respect to $\eta_{\cdot 1}$ are given by

$$\frac{\partial L}{\partial \eta_{\cdot 1}} = -\frac{1}{\sigma_{11}}[(X' D X + X^{*\prime}(I - D)X^*)\eta_{\cdot 1} - (X' D + X^{*\prime})y]$$

$$+\frac{\sigma_{12}}{\sigma_{11}^2(1 - \rho^2)^{(1/2)}}[X' D_1 D e - X^{*\prime} D_2(I - D)e] = 0$$

$$e = (1, 1, \ldots, 1)', \quad D_1 = \mathrm{diag}\left(\frac{\phi(h(1))}{\Phi(h(1))}, \ldots, \frac{\phi(h(T))}{\Phi(h(T))}\right),$$

$$D_2 = \mathrm{diag}\left(\frac{\phi(h(1))}{1 - \Phi(h(1))}, \ldots, \frac{\phi(h(T))}{1 - \Phi(h(T))}\right),$$

with the additional requirement that $\delta = 1$ in D_1, while $\delta = 0$ in D_2, whenever they appear in $h(t)$.

This has a certain similarities with the first order conditions for obtaining $\theta_{\cdot 1}$ in the procedure adopted in Dhrymes, but it is far more complicated.