London School of Economics
Department of Economics

Vassilis Hajivassiliou
Winter Term 2024

Further Topics in Econometrics (Ec485/Ec518) Answers to Problem Set #2 – Simulation-Based Inference

1. Consider a balanced panel data set with N individuals indexed by $i = 1, \cdots, N$ each observed for $T = 5$ time periods. Consider the Multiperiod Binary Probit model defined by

$$y_{it} = \begin{cases} 1 & iff \quad x'_{it}\beta + \epsilon_{it} > 0 \\ 0 & otherwise \end{cases}$$

Suppose that the errors ϵ are fully independent from all regressor variables. They are also independent across individuals, but have the one-factor error-components structure:

$$\epsilon_{it} = \alpha_i + \nu_{it} \sim N(0, \sigma^2_\alpha + \sigma^2_\nu)$$

where $\alpha_i \sim N(0, \sigma^2_\alpha)$ i.i.d. over i; $\nu_{it} \sim N(0, \sigma^2_\nu)$ i.i.d. over both i and ν; and α_i, ν_{it} fully mutually independent for all i and t.

Consider the 5×1 sequence of binary choices for individual i

$$y_i \equiv \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \\ y_{i5} \end{pmatrix}$$

and denote the stacked vector of the sequences of the binary choices of all individuals by the $5N \times 1$ vector y. The regressors are similarly stacked into the $5N \times k$ matrix X, where k is the number of explanatory variables in the model. Our aim is to estimate the unknown parameters using the observed data set (y, X).

   (a) Can the regressor matrix X contain an intercept (vector of ones)? Why or why not?

   ANSWER:

   It is possible to run a probit model here with a constant since there is no perfect multicollinearity problem. Three points: (1) since the threshold is already normalized to be 0, there is no problem including the constant — it is of course true that the threshold and the intercept cannot be separately. (2) Given the E alpha=0 assumption, there is no confusion between including an intercept and it being confused with the mean of alpha. Finally, (3) had we tried to adopt Fixed-Effects-type of estimation by including N individual intercepts for each a_i, then there would have been multicollinearity with an overall constant. But we cannot estimate all the alphas in any case because of the infinite incidental parameters problem.

(b) Are the variance terms $\sigma_\alpha^2 + \sigma_\nu^2$ identified? Why or why not?
ANSWER:
As with any binary threshold crossing model, the β parameters appear in the likelihood contributions as ratios with $\sqrt{\sigma_\epsilon^2} = \sqrt{\sigma_\alpha^2 + \sigma_\nu^2}$, i.e., as $\beta/\sqrt{\sigma_\alpha^2 + \sigma_\nu^2}$. Consequently, σ_α^2 and σ_ν^2 cannot be separately identified — one option would be to normalize $\sigma_\epsilon^2 = \sigma_\alpha^2 + \sigma_\nu^2 = 1$, which shows that only one of the two variances can be identified even after this normalization.

(c) Suppose an investigator carries out Pooled Binary Probit estimation, whereby the panel data structure is completely ignored and all individual observations are treated as if they came from a single cross-section sample of dimension $5N \times 1$. Discuss the properties of this estimation approach.
ANSWER:
An individual observation of the binary LDV y_{it} is a Bernoulli r.v. conditional on the regressors and the parameters, with (marginal) probability of success:

$$\begin{aligned}\Pr(y_{it} &= 1|X, params) = \Pr(x_{it}'\beta + \epsilon_{it} > 0) \\ &= \Pr(\frac{x_{it}'\beta + \epsilon_{it}}{\sqrt{\sigma_\alpha^2 + \sigma_\nu^2}} > 0) = \Phi\left(\frac{x_{it}'\beta}{\sqrt{\sigma_\alpha^2 + \sigma_\nu^2}}\right)\end{aligned}$$

and of failure:

$$\begin{aligned}\Pr(y_{it} &= 0|X, params) = \Pr(x_{it}'\beta + \epsilon_{it} \leq 0) \\ &= 1 - \Phi\left(\frac{x_{it}'\beta}{\sqrt{\sigma_\alpha^2 + \sigma_\nu^2}}\right)\end{aligned}$$

Therefore, the two probabilities can be summarized in the single statement:

$$\begin{aligned}\Pr(y_{it} &= 1|X, params) = \begin{cases} \Phi\left(\frac{x_{it}'\beta}{\sqrt{\sigma_\alpha^2+\sigma_\nu^2}}\right) & if \quad y_{it} = 1 \\ 1 - \Phi\left(\frac{x_{it}'\beta}{\sqrt{\sigma_\alpha^2+\sigma_\nu^2}}\right) & if \quad y_{it} = 0 \end{cases} \\ &= \Phi\left((2y_{it} - 1) \cdot \frac{x_{it}'\beta}{\sqrt{\sigma_\alpha^2 + \sigma_\nu^2}}\right)\end{aligned}$$

This simplification is possible because when $y_{it} = 1$ then $(2y_{it} - 1) = 1$ so the probability equals $\Phi\left(\frac{x'_{it}\beta}{\sqrt{\sigma_\alpha^2+\sigma_\nu^2}}\right)$ and when $y_{it} = 0$ then $(2y_{it} - 1) = -1$ and hence the probability equals $\Phi\left(-\frac{x'_{it}\beta}{\sqrt{\sigma_\alpha^2+\sigma_\nu^2}}\right) = 1 - \Phi\left(\frac{x'_{it}\beta}{\sqrt{\sigma_\alpha^2+\sigma_\nu^2}}\right)$ as required.

Pooling the data amounts to assuming that the errors ϵ_{it} are independent over both i (which they are) and over t (which they are not), in view of the presence of the persistent α_i in all the errors of individual i. So the Pooled Binary Probit uses the likelihood contribution:

$$\ell_{it}(y_{it}|X, parameters) = \prod_{t=1}^{T} \Phi\left((1 - 2y_{it}) \cdot \frac{x'_{it}\beta}{\sqrt{\sigma_\alpha^2 + \sigma_\nu^2}}\right)$$

which is incorrect — the different Phees cannot be multiplied together since the errors of individual i are serially correlated and dependent.

Because of this mis-specification error, the Pooled BP cannot be shown to be CUAN.

(d) Describe the approach of Heckman (1981) for estimating the Random Effects Binary Probit model and explain how his approach improves on the Pooled method of (c). Discuss the order of integration necessary for implementing the Heckman estimation approach.

ANSWER:

Heckman pointed out that *conditionally on the heterogeneity term* α_i,

$$\Pr(y_{it} = 1|\alpha_i, X, params) = \Pr(x'_{it}\beta + \alpha_i + \nu_{it} \le 0) = \Phi\left(\frac{x'_{it}\beta + \alpha_i}{\sqrt{\sigma_\nu^2}}\right)$$

giving the conditional probability and likelihood contribution:

$$\ell_{it}(y_{it}|\alpha_i, X, parameters) = \Phi\left((1 - 2y_{it}) \cdot \frac{x'_{it}\beta + \alpha_i}{\sigma_\nu}\right)$$

But now the only stochastic term is ν_{it}. As long as this component is assumed to be i.i.d. over both i and t, the T terms of individual i can be multiplied together to give the conditional likelihood contribution for all periods of individual i as:

$$\begin{aligned} & \ell_i(y_{i1}, \cdots, y_{it}, \cdots, y_{iT}|\alpha_i, X, parameters) \\ = \quad & \ell_i(y_i|\alpha_i, X, parameters) = \prod_{t=1}^{T} \Phi\left((1 - 2y_{it}) \cdot \frac{x'_{it}\beta + \alpha_i}{\sigma_\nu}\right) \end{aligned}$$

To get the joint density of y_i and a_i, we multiply the conditional by the pdf of a_i to obtain:

$$\begin{aligned}
& \ell_i(y_{i1}, \cdots, y_{it}, \cdots, y_{iT}, \alpha_i | X, parameters) \\
= \quad & \ell_i(y_i, \alpha_i | X, parameters) = \frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha_i}{\sigma_\alpha}\right) \cdot \prod_{t=1}^{T} \Phi\left((1-2y_{it}) \cdot \frac{x'_{it}\beta + \alpha_i}{\sigma_\nu}\right)
\end{aligned}$$

To obtain the final likelihood contribution without α_i, namely $\ell_i(y_i|X, parameters)$, we must "integrate out" or "marginalize out" α_i from the above formula, i.e.,

$$\begin{aligned}
\ell_i(y_i|X, parameters) &= \int_{-\infty}^{+\infty} \ell_i(y_i, \alpha_i | X, parameters) d\alpha_i \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha_i}{\sigma_\alpha}\right) \cdot \prod_{t=1}^{T} \Phi\left((1-2y_{it}) \cdot \frac{x'_{it}\beta + \alpha_i}{\sigma_\nu}\right) d\alpha_i
\end{aligned}$$

Since Heckman's work, other researchers have shown that the integrand is a non-linear function (the product of the Φs) times the normal density ϕ, the properties of the so-called "Hermite Polynomials" can be exploited to give extremely accurate numerical approximations by so called "Quadrature Rules".

(e) For an individual i, characterize the vector of their conditional expectations of each binary choice given the observed data

$$E(y_i|X) \equiv \begin{pmatrix} Ey_{i1}|X \\ Ey_{i2}|X \\ Ey_{i3}|X \\ Ey_{i4}|X \\ Ey_{i5}|X \end{pmatrix}$$

Hence show how the unknown parameters of the model can be estimated through the Generalized Method of Moments approach of Avery, Hansen, and Hotz (1983). Discuss the order of integration necessary for implementing the AHH approach. Outline the properties of the AHH estimation approach.

ANSWER:

As discussed in the lecture, the marginal expectation functions of Bernoullis are equal to the probability of success, i.e.,

$$E(y_{it}|X) = \Pr(y_{it} = 1|X, params) = 1 - \Phi\left(\frac{x'_{it}\beta}{\sqrt{\sigma_\alpha^2 + \sigma_\nu^2}}\right)$$

This, of course, is a univariate integral — $\Phi(z)$ is the cdf of a N(0,1), i.e., the probability of a value less than or equal to z, meaning the integral from $-\infty$ up to point z.

Also as explained in the lecture, the optimal instruments and weights will involve the Bernoulli variance

$$V(y_{it}|X) = \left[1 - \Phi\left(\frac{x'_{it}\beta}{\sqrt{\sigma^2_\alpha + \sigma^2_\nu}}\right)\right] \cdot \Phi\left(\frac{x'_{it}\beta}{\sqrt{\sigma^2_\alpha + \sigma^2_\nu}}\right)$$

and the derivatives of $E(y_{it}|X)$ and $V(y_{it}|X)$. All these terms will be univariate integrals.

(f) Now suppose that instead of i.i.d. over t and i, the error component ν_{it} follows an ARMA(1,1) process with autoregressive parameter γ and moving average parameter λ. How does the expression for the probability of the observed sequences of binary choices change in this case? I.e., you must characterize the probability $\Pr(y_i|X)$ and explain in which ways it differs from that probability under the one-factor error-components model of Heckman (1981). Discuss the order of integration necessary for the evaluation of $\Pr(y_i|X)$.

ANSWER:

The approach of Heckman (1981) will not work in this case. This is because if we condition on α_i we would again get the expression where the only randomness is in ν_{it}, i.e.:

$$\ell_{it}(y_{it}|\alpha_i, X, parameters) = \Phi\left((1 - 2y_{it}) \cdot \frac{x'_{it}\beta + \alpha_i}{\sigma_\nu}\right)$$

But now the ν_{it} are *dependent* across the different time-periods for individual i because $\nu_{it} \sim ARMA(1,1)$ and hence it is serially correlated/dependent across t. Therefore, to get the likelihood contribution we canot multiply together these terms (whether we do it conditionally on α_i and then marginalize it out or directly).

The correct calculation would be to write:

$$\begin{aligned} &\Pr(y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}|X, parameters) \\ = &\int_{a_5}^{b_5}\int_{a_5}^{b_5}\int_{a_5}^{b_5}\int_{a_5}^{b_5}\int_{a_5}^{b_5} pdf(y^*_{i1}, y^*_{i2}, y^*_{i3}, y^*_{i4}, y^*_{i5}|X, parameters)dy^*_{i1}dy^*_{i2}dy^*_{i3}dy^*_{i4}dy^*_{i5} \end{aligned}$$

where $y^*_{it} = x'_{it}\beta + \epsilon_{it}$ and the limits of integration a_t and b_t are defined as follows:

$$if \quad y_{it} = 1: \quad (a_t = 0, \quad b_t = +\infty)$$

$$if \quad y_{it} = 0: \quad (a_t = -\infty, \quad b_t = 0)$$

(g) Show how probability of the observed choices $\Pr(y_i|X)$ can be characterized through a set of linear inequality constraints on a set of correlated unobserved random variables. Use this framework to explain the Simulation-Based estimation approaches of Boersch-Supan and Hajivassiliou (1993) and Hajivassiliou and McFadden (1997). Discuss the properties of these estimation approaches.

NB: Show how your answer to this part can handle both the original error-component-only specification, as well as the error-components-plus-ARMA(1,1) error structure of part 1(g).

ANSWER:

As explained in the previous part:

$$\begin{aligned} & \Pr(y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}|X, parameters) \\ = & \int_{a_5}^{b_5}\int_{a_5}^{b_5}\int_{a_5}^{b_5}\int_{a_5}^{b_5}\int_{a_5}^{b_5} pdf(y^*_{i1}, y^*_{i2}, y^*_{i3}, y^*_{i4}, y^*_{i5}|X, parameters) dy^*_{i1} dy^*_{i2} dy^*_{i3} dy^*_{i4} dy^*_{i5} \\ = & \Pr(a_1 < y^*_{i1} < b_1, a_2 < y^*_{i2} < b_2, a_3 < y^*_{i3} < b_3, \\ a_4 < & \; y^*_{i4} < b_4, a_5 < y^*_{i5} < b_5 | X, parameters) \end{aligned}$$

where again as explained in the previous part the lower limits a_t and the upper limits b_t are defined by:

$$\begin{aligned} if \quad y_{it} = 1: & \quad (a_t = 0, \quad b_t = +\infty) \\ if \quad y_{it} = 0: & \quad (a_t = -\infty, \quad b_t = 0) \end{aligned}$$

Boersch-Supan and Hajivassiliou defined the "Smooth Recursive Conditioning" simulator based on the Cholesky factorization of the VCov matrix of the total ϵ errors — subsequently known as the GHK simulator. Their simulation-based estimation method is thus "Maximum Smoothly Simulated Likelihood" (or MSL with GHK).

Hajivassiliou and McFadden proposed the "Gibbs Resampling Simulator" for approximating the Likelihood Scores:

$$\frac{\partial \ell_i / \partial \theta}{\ell_i}$$

Their estimation method was thus the "Method of Simulated Scores" (or MSS with Gibbs Resampling).

Both simulation strategies can be formulated as generating draws from (latent) distributions that are restricted over regions defined by a set of linear inequalties — as illustrated above.

2. Consider the Multiperiod Autoregressive Binary Probit model defined by:

$$y_{it} = \begin{cases} 1 & iff \quad \delta y_{i,t-1} + x'_{it}\beta + \epsilon_{it} > 0 \\ 0 & otherwise \end{cases}$$

Suppose that the errors ϵ are fully independent from all regressor variables. They are also independent across individuals, but have the one-factor error-components structure:

$$\epsilon_{it} = \alpha_i + \nu_{it} \sim N(0, \sigma_\alpha^2 + \sigma_\nu^2)$$

where $\alpha_i \sim N(0, \sigma_\alpha^2)$ i.i.d. over i; $\nu_{it} \sim N(0, \sigma_\nu^2)$ i.i.d. over both i and ν; and α_i, ν_{it} fully mutually independent for all i and t.

Consider the 5×1 sequence of binary choices for individual i

$$y_i \equiv \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \\ y_{i5} \end{pmatrix}$$

and denote the stacked vector of the sequenes of the binary choices of all individuals by the $5N \times 1$ vector y. The regressors are similarly stacked into the $5N \times k$ matrix X, where k is the number of explanatory variables in the model. Our aim is to estimate the unknown parameters using the observed data set (y, X).

(a) What additional problems are caused by the presence of the lagged dependent variable as regressor, i.e., by $\delta \neq 0$? How did Heckman (1981) propose to handle these extra problems?

ANSWER:

To incorporate the presence of the lagged dependent discrete response term $y_{i,t-1}$ as an extra regressor, Heckman modified his approach as follows: First *conditionally on the heterogeneity term* α_i,

$$\begin{aligned} \Pr(y_{it} &= 1|\alpha_i, X, params) \\ &= \Pr(\delta y_{i,t-1} + x'_{it}\beta + \alpha_i + \nu_{it} \leq 0) \\ &= \Phi\left(\frac{\delta y_{i,t-1} + x'_{it}\beta + \alpha_i}{\sqrt{\sigma_\nu^2}}\right) \end{aligned}$$

giving the conditional probability and likelihood contribution:

$$\ell_{it}(y_{it}|\alpha_i, X, parameters) = \Phi\left((1 - 2y_{it}) \cdot \frac{\delta y_{i,t-1} + x'_{it}\beta + \alpha_i}{\sigma_\nu}\right)$$

This involves the additional assumption that the discrete response y_{it} follows a "Markov scheme" whereby conditional on its single past value $y_{i,t-1}$, the randomness left is only through the ϵ_{it} term.

Again conditional on α_i the only stochastic term is ν_{it}. As long as this component is assumed to be i.i.d. over both i and t, the T terms of individual i can be multiplied together to give the conditional likelihood contribution for all periods of individual i as:

$$\begin{aligned} & \ell_i(y_{i1}, \cdots, y_{it}, \cdots, y_{iT} | y_{i0}, \alpha_i, X, parameters) \\ = & \ \ell_i(y_i | y_{i0}, \alpha_i, X, parameters) \\ = & \prod_{t=1}^{T} \Phi\left((1-2y_{it}) \cdot \frac{\delta y_{i,t-1} + x'_{it}\beta + \alpha_i}{\sigma_\nu}\right) \end{aligned}$$

*NB: the formula now *conditions* also on the initial condition y_{i0} since for the first period $t = 1$, the RHS will have the lagged y in period $t = 0$. Since there are N individuals each with their own initial condition, and since $N \to \infty$, the number of initial conditions grows to infinity as well.*

To get the joint density of y_i and a_i, we multiply the conditional by the pdf of a_i to obtain:

$$\begin{aligned} & \ell_i(y_{i1}, \cdots, y_{it}, \cdots, y_{iT}, \alpha_i | y_{i0}, X, parameters) \\ = & \ \ell_i(y_i, \alpha_i | y_{i0}, X, parameters) \\ = & \ \frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha_i}{\sigma_\alpha}\right) \cdot \prod_{t=1}^{T} \Phi\left((1-2y_{it}) \cdot \frac{\delta y_{i,t-1} + x'_{it}\beta + \alpha_i}{\sigma_\nu}\right) \end{aligned}$$

To obtain the final likelihood contribution without α_i, namely $\ell_i(y_i|X, parameters)$, we must "integrate out" or "marginalize out" α_i from the above formula, i.e.,

$$\begin{aligned} & \ell_i(y_i | y_{i0}, X, parameters) \\ = & \int_{-\infty}^{+\infty} \ell_i(y_i, \alpha_i | y_{i0}, X, parameters) d\alpha_i \\ = & \int_{-\infty}^{+\infty} \frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha_i}{\sigma_\alpha}\right) \cdot \prod_{t=1}^{T} \Phi\left((1-2y_{it}) \cdot \frac{\delta y_{i,t-1} + x'_{it}\beta + \alpha_i}{\sigma_\nu}\right) d\alpha_i \end{aligned}$$

*But the problem would still not be solved, however, because of the N initial conditions that each of these contributions will depend on. Heckman investigated this issue and found that it is *insurmountable* to overcome it we would need to characterize the distribution of y_{i0} and marginalize it out — just like we did to take care of α_i. The problem, however, is that the distribution of y_{i0} is not known — and in fact Heckman showed that (a) it *cannot* *be* a well-defined, stable distribution like the distribution of y_{it} because it will depend* *on the actual paths of all the y_{it}s in the past; and (b) whatever the distribution of y_{i0} is, it *cannot* be the same as that of y_{it} for $t = 1, \cdots, T$. This problem remains unresolved for the last 30 years.*

(b) Explain the terms "State Dependence" and "Persistent Unobserved Heterogeneity" and discuss how these two properties can be distinguished and tested formally.

ANSWER:

"State Dependence" refers to how significantly the probability of y_{it} today is affected by the observed discrete response in the previous period, i.e., by $y_{i,t-1}$. So this can be quantified by assessing the significance of the δ parameter in front of the lagged discrete response term on the RHS.

"Persistent Unobserved Heterogeneity" (or more generallay just "Unobserved Persistence") refers to how important the term α_i is relative to the overall ϵ_{it} error. This can be quantified by assessing how large σ^2_α relative to $\sigma^2_\epsilon = \sigma^2_\alpha + \sigma^2_\nu$. More generally, "Unobserved Persistence" overall can be quantified by the persistence in α_i and possible serial correlation in ν_{it}, if for example the latter folows ARMA-type schemes.

Heckman's fundamental point was that we need to model explicitly for the *both* sources of temporal dependence to get meaningful, reliable estimates:

If for example we left out the lagged dependent indicator and only allowed for the persistent α_i (and/or allow for ν_{it} to be serially correlated — which we can only do through Simulation-Based Estimators), the error autocorrelation would appear misleadingly strong because it would try to capture the importance of the state-dependence term we left out.

Conversely, if we included the $y_{i,t-1}$ term but didn't allow for the error term to be serially correlated (either through α_i and/or through serial correlation in the ν_{it}), then the δ coefficient would be seriously upwardly biased/inconsistent because it would try to capture the error serial dependence/persistence we left out.

(c) Show how the probability of the observed choices $\Pr(y_i|X)$ can be characterized through a set of linear inequality constraints on a set of correlated unobserved random variables. Use this framework to explain the Simulation-Based estimation approaches of Boersch-Supan and Hajivassiliou (1993) and Hajivassiliou and McFadden (1997). Discuss the properties of these estimation approaches for the more complicated error specification that, in addition to the heterogeneity term α_i, the ν_{it} error component follows an ARMA(1,1) process with autoregressive parameter γ and moving average parameter λ.

ANSWER:

This is entirely analogous to the discussion in 1.g above with

one very major modification:

$$
\begin{aligned}
& \Pr(y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5} | y_{i0}, X, parameters) \\
= & \int_{a_5}^{b_5}\int_{a_5}^{b_5}\int_{a_5}^{b_5}\int_{a_5}^{b_5}\int_{a_5}^{b_5} pdf(y^*_{i1}, y^*_{i2}, y^*_{i3}, y^*_{i4}, y^*_{i5} | X, parameters) dy^*_{i1} dy^*_{i2} dy^*_{i3} dy^*_{i4} dy^*_{i5} \\
= & \Pr(a_1 < y^*_{i1} < b_1, a_2 < y^*_{i2} < b_2, a_3 < y^*_{i3} < b_3, \\
a_4 < & y^*_{i4} < b_4, a_5 < y^*_{i5} < b_5 | X, parameters) \\
(2) = & \Pr(a_1 < \delta y_{i0} + x'_{it}\beta + \epsilon_{i1} < b_1, \\
a_2 < & \delta y_{i1} + x'_{it}\beta + \epsilon_{i2} < b_2, \\
a_3 < & \delta y_{i2} + x'_{it}\beta + \epsilon_{i3} < b_3, \\
a_4 < & \delta y_{i3} + x'_{it}\beta + \epsilon_{i4} < b_4, \\
a_5 < & \delta y_{i4} + x'_{it}\beta + \epsilon_{i5} < b_5 | X, parameters) \\
(3) = & \Pr(a_1 - \delta y_{i0} < x'_{i1}\beta + \epsilon_{i1} < b_1 - \delta y_{i0}, \\
a_2 - \delta y_{i1} < & x'_{i2}\beta + \epsilon_{i2} < b_2 - \delta y_{i1}, \\
a_3 - \delta y_{i2} < & x'_{i3}\beta + \epsilon_{i3} < b_3 - \delta y_{i2}, \\
a_4 - \delta y_{i3} < & x'_{i4}\beta + \epsilon_{i4} < b_4 - \delta y_{i3}, \\
a_5 - \delta y_{i4} < & x'_{i5}\beta + \epsilon_{i5} < b_5 - \delta y_{i4} | X, parameters) \\
(4) = & \Pr(a_1 - \delta y_{i0} < u_{i1} < b_1 - \delta y_{i0}, \\
a_2 - \delta y_{i1} < & u_{i2} < b_2 - \delta y_{i1}, \\
a_3 - \delta y_{i2} < & u_{i3} < b_3 - \delta y_{i2}, \\
a_4 - \delta y_{i3} < & \delta y_{i3} + u_{i4} < b_4 - \delta y_{i3}, \\
a_5 - \delta y_{i4} < & u_{i5} < b_5 - \delta y_{i4} | X, parameters) \\
(5) = & \Pr(a_1 - \delta y_{i0} < u_{i1} < b_1 - \delta y_{i0}, \\
a_2 - \delta y_{i1} < & u_{i2} < b_2 - \delta y_{i1}, \\
a_3 - \delta y_{i2} < & u_{i3} < b_3 - \delta y_{i2}, \\
a_4 - \delta y_{i3} < & \delta y_{i3} + u_{i4} < b_4 - \delta y_{i3}, \\
a_5 - \delta y_{i4} < & u_{i5} < b_5 - \delta y_{i4} | X, parameters)
\end{aligned}
$$

where the latent random variables that are being limited are redefined as:

$$u_{it} \equiv x'_{it}\beta + \epsilon_{it}$$

and the lower limits c_t and the upper limits d_t now have *four* cases, depending on what happened to the discrete indicator in the previous period:

$$
\begin{array}{llll}
if & y_{it} = 1, y_{i,t-1} = 1: & (c_t = -\delta, & d_t = +\infty) \\
if & y_{it} = 1, y_{i,t-1} = 0: & (c_t = 0, & d_t = +\infty) \\
if & y_{it} = 0, y_{i,t-1} = 1: & (c_t = -\infty, & d_t = -\delta) \\
if & y_{it} = 0, y_{i,t-1} = 0: & (c_t = -\infty, & d_t = 0)
\end{array}
$$

This is because the limits are defined by:

$$\begin{aligned} c_t &= a_t - \delta y_{i,t-1} \\ d_t &= b_t - \delta y_{i,t-1} \end{aligned}$$

In this fashion, we are able to charactarize in the same fashion to joint probability $\Pr(y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}|y_{i0}, X, parameters)$ as a set of linear inequality constraints on (possibly heavily serially correlated) latent random variables.

3. Consider the Simultaneous Binary Liquidity and Ordered Response Employment model of Hajivassiliou and Ioannides (2005). The model is estimated using a Panel Data set on the S_{it} and E_{it} dependent discrete variables and the matrix of explanatory factors for the two sides, X_S and X_E respectively.

   Dropping the i index for simplicity, define two latent dependent variables $y^*_{1t} \equiv S^*_t$ and $y^*_{2t} \equiv E^*_t$ that are the underpinnings of S_t and E_t according to:

   $$S_t = \begin{cases} 1 & iff \quad S^*_t > 0 \\ 0 & otherwise \end{cases} \qquad E_t = \begin{cases} -1 & iff \quad E^*_t < \theta^- \\ 0 & iff \quad \theta^- < E^*_t < \theta^+ \\ +1 & iff \quad E^*_t > \theta^+ \end{cases}$$

   Also dropping the t subscript for ease of notation, we consider the model with spillover effects on both sides:

   $$y^*_1 \equiv S^* = \mathbf{1}(y^*_2 < \theta^-)\delta_{01} + \mathbf{1}(y^*_2 > \theta^+)\delta_{02} + x_1\beta_1 + \epsilon_1$$
   $$y^*_2 \equiv E^* = \mathbf{1}(y^*_1 > 0)\kappa_0 + x_2\beta_2 + \epsilon_2$$

   The contemporaneous spillover effect $\delta_0 E$ on the RHS of S^* into $\delta_{01}\mathbf{1}(\mathbf{E} = -\mathbf{1}) + \boldsymbol{\delta_{02}}\mathbf{1}(\mathbf{E} = \mathbf{1})$, i.e., into separate terms for the overemployment and the under/unemployment indicators.

   Consider the observed sequence of discrete responses S_{it} and E_{it} for individual i in period t. Stack these into their vector of choices for all periods:

   $$(S_i, E_i|X_S, X_E) \equiv \begin{pmatrix} S_{i1} \\ E_{i1} \\ S_{i2} \\ E_{i2} \\ \vdots \\ S_{i,T-1} \\ E_{i,T-1} \\ S_{iT} \\ E_{iT} \end{pmatrix}$$

   Define the joint probability of the observed discrete responses for all individuals:

   $$\Pr(S_1, E_1, S_2, E_2, \cdots, S_{N-1}, E_{N-1}, S_N, E_N|X_S, X_E)$$

Show how this probability can be characterized through a set of linear inequality constraints on a set of correlated unobserved random variables ϵ_{1it} and ϵ_{2it} for all i and t. Use this framework to explain the Simulation-Based estimation approaches of Boersch-Supan and Hajivassiliou (1993) and Hajivassiliou and McFadden (1997). Discuss the properties of these estimation approaches for the more complicated error specifications for ϵ_{1it} and ϵ_{2it} when they include both persistent heterogeneity terms like α_i, as well as error components which follow ARMA processes.

ANSWER:

In terms of the GHK simulator described above, the probability of a pair (S, E) is equivalent to the probability:

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} < \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} < \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

where $(\epsilon_1, \epsilon_2)' \sim N((\mu_1, \mu_2)', \Sigma_\epsilon)$, and a and b are:

*** TABLE 1 — see over ***

Define the 2×1 vectors a_{it}, b_{it}, and ϵ_{it}. Stacking all the T_i periods of observation for individual i gives the $2 \cdot T_i \times 1$ vectors a_i, b_i, and ϵ_i, where ϵ_i has the $2 \cdot T_i \times 2 \cdot T_i$ variance-covariance matrix with structure characterized by the precise serial correlation assumptions made on the ϵ_{it}s. In particular, one-factor random effect assumptions will imply an equicorrelated block structure on Σ_ϵ, while our most general assumption of one-factor random effects *combined with* an AR(1) process for each error implies that Σ_ϵ combines equicorrelated and Toeplitz-matrix features.

Through this representation, the probability of a complete sequence of the observable (S, E) behaviour for individual household i, conditionally on the initial conditions S_{i0} and E_{i0}, is given by: $P(S_1, \cdots, S_{T_i}, E_1, \cdots, E_{T_i}) = Prob(a_i < \epsilon_i < b_i)$. Consequently, our approach incorporates fully: (a) the contemporaneous correlations in ϵ_{it}; (b) the one-factor plus AR(1) serial correlations in ϵ_i; and (c) the dependency of S_{it} on E_{it}, and vice versa. The possible endogeneity of S_{i0} and E_{i0} is handled by the approach described next.