Econometric Methods: Assignment 1

Bingxue Li

1 MLE and NLLSE

Consider the following model

$$y_i = \begin{cases} 0 & \text{if } y_i^* \le \alpha_{10} \\ 1 & \text{if } \alpha_{10} < y_i^* \le \alpha_{20} \\ 2 & \text{if } \alpha_{20} < y_i^* \end{cases}$$

where y_i^* is a latent variable given by

$$y_i^* = \beta_0 x_i + u_i$$

and where

$$u_i \mid x_i \sim \mathcal{N}(0, 1)$$

Note that here x_i is scalar. The above model is the ordered probit model. In what follows, we will generate samples of n iid observations from the probit model, where $x_i \sim N(0.5, 1)$ and $\theta_0 = (\alpha_{10}, \alpha_{20}, \beta_0)' = (0, 1, 1)'$. Note that the subscript $_0$ denotes the true parameter value.

(a) Write down the objective function for the (conditional) log-likelihood function (multiplied by $-1/n$).

To construct the likelihood, fist compute conditional PMF of y_i [1]:

$$\begin{aligned}
\mathbb{P}(y_i = 0 \mid x_i) &= \mathbb{P}(y_i^* \le \alpha_{10} \mid x_i) = \mathbb{P}(\beta_0 x_i + u_i \le \alpha_{10} \mid x_i) = \Phi(\alpha_{10} - \beta_0 x_i) \\
\mathbb{P}(y_i = 2 \mid x_i) &= \mathbb{P}(\alpha_{20} < y_i^* \mid x_i) = \mathbb{P}(\alpha_{20} < \beta_0 x_i + u_i \mid x_i) = \Phi(\beta_0 x_i - \alpha_{20}) \\
\mathbb{P}(y_i = 1 \mid x_i) &= 1 - \mathbb{P}(y_i = 0 \mid x_i) - \mathbb{P}(y_i = 2 \mid x_i) \\
&= 1 - \Phi(\alpha_{10} - \beta_0 x_i) - \Phi(\beta_0 x_i - \alpha_{20}) \\
&= \Phi(\alpha_{20} - \beta_0 x_i) - \Phi(\alpha_{10} - \beta_0 x_i)
\end{aligned}$$

We write the objective function of ML estimation as log-likelihood of the sample:

$$\begin{aligned}
Q_{n,MLE}(\theta) &= -\frac{1}{n}\Big\{\sum_{i:y_i=0} log\mathbb{P}(y_i = 0 \mid x_i) + \sum_{i:y_i=1} log\mathbb{P}(y_i = 1 \mid x_i) + \sum_{i:y_i=2} log\mathbb{P}(y_i = 2 \mid x_i)\Big\} \\
&= -\frac{1}{n}\Big\{\sum_{i:y_i=0} logF(\alpha_1 - \beta x_i) + \sum_{i:y_i=1} log(F(\alpha_2 - \beta x_i) - F(\alpha_{10} - \beta_0 x_i)) \\
&+ \sum_{i:y_i=2} logF(\beta x_i - \alpha_2)\Big\} \qquad (1)
\end{aligned}$$

[1] $\Phi(\cdot)$ is the cumulative distribution function of $\mathcal{N}(0, 1)$.

$$\hat{\theta}_{MLE} = \arg\min_{\theta \in \Theta} Q_{n,MLE}(\theta) \tag{2}$$

(b) Estimate the model by MLE providing only the objective function to the algorithm. To that end, use the code online, namely the file "Assignment1.R". [Note that you'll have to complete the file as well as the functions that the file is calling.] Report $\hat{\theta} = \left(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}\right)'$.

See Table 1 for estimators.

(c) Write down the nonlinear least squares (NLS) objective function for the ordered probit model.

To compute the nonlinear least squares residuals, first construct the parametric regression function $m(x_i, \theta)$ which is nonlinear in θ such that:

$$y_i = m(x_i, \theta) + v_i \quad \text{with} \quad \mathbb{E}[v_i \mid x_i] = 0$$

Notice that under this nonlinear least squares specification, we have $\mathbb{E}[y_i \mid x_i] = m(x_i, \theta)$:

$$\begin{aligned} m(x_i, \theta_0) := \mathbb{E}[y_i \mid x_i] &= 0 \cdot \mathbb{P}(y_i = 0 \mid x_i) + 1 \cdot \mathbb{P}(y_i = 1 \mid x_i) + 2 \cdot \mathbb{P}(y_i = 2 \mid x_i) \\ &= \Phi(\beta_0 x_i - a_{10}) + \Phi(\beta_0 x_i - a_{20}) \end{aligned} \tag{3}$$

We define the objective function minimizing residual squared:

$$Q_{n,NLLSE}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\{y_i - m(x_i, \theta)\}^2 = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \{\Phi(\beta x_i - a_1) + \Phi(\beta x_i - a_2)\}\}^2 \tag{4}$$

$$\hat{\theta}_{NLLSE} = \arg\min_{\theta \in \Theta} Q_{n,NLLSE}(\theta) \tag{5}$$

(d) Estimate the model by NLS providing only the objective function to the algorithm and report $\hat{\theta}$.

See Table 1 for estimators.

2 GMM: Just-ID

Let $m(x_i, \theta)$ be such that $E(y_i \mid x_i) = m(x_i, \theta_0)$. Then, we have

$$E\left[y_i - m(x_i, \theta_0) \mid x_i\right] = 0$$

This, in turn, implies that

$$E\left[\left(y_i - m(x_i, \theta_0)\right) g(x_i)\right] = 0 \tag{6}$$

for any function $g(x_i)$.
(e) Write down the GMM objective function based on (1) with $g(x_i) = (1, x_i, x_i^2)'$. (Hint: Note that given $g(x_i) = (1, x_i, x_i^2)'$, the model is "just-identified", i.e., there are as many equations as parameters, such that the choice of $\hat{W}$ does not matter. Hence, you may choose $\hat{W} = I_3$.)

To define the moment conditions, let $h(w_i, \theta_0) := (y_i - m(x_i, \theta_0))g(x_i)$, where $w_i = (x_i, y_i)'$.

$$\begin{aligned} h(w_i, \theta) &= \{y_i - (\Phi(\beta x_i - a_1) + \Phi(\beta x_i - a_2))\}g(x_i) \\ &= \begin{bmatrix} \{y_i - (\Phi(\beta x_i - a_1) + \Phi(\beta x_i - a_2))\} \\ \{y_i - (\Phi(\beta x_i - a_1) + \Phi(\beta x_i - a_2))\}x_i \\ \{y_i - (\Phi(\beta x_i - a_1) + \Phi(\beta x_i - a_2))\}x_i^2 \end{bmatrix} \end{aligned} \tag{7}$$

$$\mathbb{E}[h(w_i, \theta_0)] = 0 \tag{8}$$

Note that $dim(h(\cdot)) = dim(g(\cdot)) = dim(\Theta) = 3$, this is a case of just-identified GMM, we can simply set $\hat{W} = I_3$, a symmetric weight matrix, and set the objective function as weighted Euclidean norm:

$$Q_{n,Just-ID-GMM}(\theta) = [\frac{1}{n}\sum_{i=1}^{n} h(w_i, \theta)]' \cdot I_3 \cdot [\frac{1}{n}\sum_{i=1}^{n} h(w_i, \theta)] \tag{9}$$

$$\hat{\theta}_{Just-ID-GMM} = \arg\min_{\theta \in \Theta} Q_{n,Just-ID-GMM}(\theta) \tag{10}$$

(f) Estimate the model by using the GMM objective function from (e) providing only the objective function to the algorithm and report $\hat{\theta}$.
See Table 1 for estimators.

3 GMM: Under/Over-ID

Let $p_y(x_i, \theta)$ be such that $P(y_i = y \mid x_i) = p_y(x_i, \theta_0)$. Note that we also have

$$E[\mathbb{1}(y_i = y) - p_y(x_i, \theta_0) \mid x_i] = 0$$

for $y \in \{0, 1, 2\}$. This, in turn, implies that

$$E[(\mathbb{1}(y_i = y) - p_y(x_i, \theta_0))\, g(x_i)] = 0 \tag{11}$$

for $y \in \{0, 1, 2\}$ and for any function $g(x_i)$.

(g) Why can we not identify (or estimate) θ_0 from the three moment conditions

$$E[\mathbb{1}(y_i = y) - p_y(x_i, \theta_0)] = 0$$

for $y \in \{0,1,2\}$?

This corresponds to the case of under-identified GMM, for $y = 0, 1$:

$$\mathbb{E}[\mathbb{1}(y_i = 0) - p_0(x_i, \theta_0)] = 0 \tag{12}$$
$$\mathbb{E}[\mathbb{1}(y_i = 1) - p_1(x_i, \theta_0)] = 0 \tag{13}$$

Combining 12 and 13, we have

$$\mathbb{E}[\mathbb{1}(y_i = 0) + \mathbb{1}(y_i = 1) - p_0(x_i, \theta_0) - p_1(x_i, \theta_0)] = 0$$
$$\Rightarrow \mathbb{E}[\{1 - \mathbb{1}(y_i = 2)\} - \{1 - p_2(x_i, \theta_0)\}] = 0 \Rightarrow \mathbb{E}[\mathbb{1}(y_i = 2) - p_2(x_i, \theta_0)] = 0$$

Note that one of the three moment conditions is a linear combination of the other two, the moment condition can be reduced to $\mathbb{E}h(w_i, \theta_0) = \mathbb{E}\left[\left(\mathbb{1}\left(y_i = y\right) - p_y\left(x_i, \theta_0\right)\right) g\left(x_i\right)\right] = 0$ for $y \in \{0,1\}$ and for any function $g\left(x_i\right)$. $dim(h(\cdot)) = 2 < dim(\Theta) = 3$. This lead to a system with three unknows and two equations. We say that this GMM model is under-identified, meaning that there is insufficient conditions to identify the parameters such that we do not have unique minimum with limiting objective function for a consistent estimator.

(h) Write down the GMM objective function based on (2) with $g\left(x_i\right) = \left(1, x_i\right)'$ for $y \in \{0,1\}$. Please use $\hat{W} = I_4$. Note that the model is "over-identified" such that the choice of $\hat{W}$ does matter.

We construct the following moment conditions:

$$\mathbb{E}\begin{bmatrix} \mathbb{1}(y_i = 0) - p_0(x_i) \\ \{\mathbb{1}(y_i = 0) - p_0(x_i)\}x_i \\ \mathbb{1}(y_i = 1) - p_1(x_i) \\ \{\mathbb{1}(y_i = 1) - p_1(x_i)\}x_i \end{bmatrix} = \mathbb{E}[h(w_i, \theta)] = 0 \tag{14}$$

$$Q_{n,Over-ID-GMM}(\theta) = [\frac{1}{n}\sum_{i=1}^{n} h(w_i, \theta)]'\hat{W}[\frac{1}{n}\sum_{i=1}^{n} h(w_i, \theta)] \tag{15}$$

$$\hat{\theta}_{Over-ID-GMM} = \arg\min_{\theta \in \Theta} Q_{n,Over-ID-GMM}(\theta) \tag{16}$$

Note that this is an overidentification case as $dim(h(\cdot)) = 4 > dim(\Theta) = 3$. We compute the GMM-estimator with $\hat{W}$ set to be I_4, although this does not give us an efficient GMM-estimator, it still satisfies consistency and asymptotic normality. Alternatively, we can consider the two-step GMM.

(i) Estimate the model by using the GMM objective function from (h) providing only the objective function to the algorithm and report $\hat{\theta}$.

Table 1: Class of Estimators for Ordinal Probit Model

Estimation Method (True)	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}$
True	$\alpha_{10} = 0$	$\alpha_{20} = 1$	$\beta_0 = 1$
MLE	0.04470834	1.01192508	1.03455275
NLLSE	0.1412001	0.8505605	1.0323160
Just-ID GMM	0.06148362	0.92212035	0.98221870
Over-ID GMM	0.1257366	1.2311598	1.2365714

4 Monte-Carlo Simulation

Now, the Monte Carlo part starts. From now on, use the file "Assignment1_MC.R" and feel free to copy+paste any lines you may need from the file "Assignment1.R". Throughout the Monte Carlo part the number of iterations equals 1000.
(j) Report the frequencies (or estimated probabilities) over 1000 Monte Carlo simulations with which the four estimators lie within a neighborhood of 0.1 for $n = 300$ and $n = 500$. Comment on what you find.

To obtain an estimator for $\mathbb{P}\{\|\hat{\theta} - \theta\| < 0.1\}$, which corresponds to a measure of consistency performance, we conduct the following experiment: 1. Randomly draw n (=300 or 500) independent pairs of (x_i, y_i) from the true population distribution F. 2. Estimate $\hat{\theta}$ with our 4 algorithm. Calculate $\mathbb{1}\{\|\hat{\theta} - \theta\| < 0.1\}$. We repeat such experiment for $B(= 1000)$ times and obtain $\left(\mathbb{1}\{\|\hat{\theta}_b - \theta\| < 0.1\}\right)_{b=1}^{B}$, this is a random sample of size B from the distribution of $\mathbb{P}\{\|\hat{\theta} - \theta\| < 0.1\}$. And we simply apply a method of moment estimator such that $\hat{\mathbb{P}}\{\|\hat{\theta} - \theta\| < 0.1\} = \frac{\sum_{b=1}^{B} \mathbb{1}\{\|\hat{\theta}_b - \theta\| < 0.1\}}{B}$.

	MLE	NLLSE	Just-ID GMM	Over-ID GMM
n=300	0.272	0.027	0.276	0.209
n=500	0.438	0.052	0.433	0.34

1. For all of the 4 estimators, the estimated probability that estimators fall into the a small neighborhood of true parameters are improved substantially once increasing replications' sample size n from 300 to 500. This observation corresponds to the definition of consistent estimators such that we have $\hat{\theta}_b$ close to truth under larger sample size under each iteration b.
2. Monte-Carlo simulation with nonlinear least square estimator has lower probability of being close to the true parameter compared with the other three estimators. Consistency of NLLSE are satisfied with finite moments of $y_i, m(\cdot)$, and conditional mean identification. (See Appendix) Then, one possible reason is that NLLSE has larger variance compared with the other three and converge to the true value slower.
3. Transforming our nonlinear least square estimator into a GMM-style estimator makes the consistency property more pronounced. We incorporate more variability according to the value of y_i in addition to $m(\cdot)$ specification with GMM moment

conditions, whereas NLLSE specify a uniform parametric functional form evaluated regardless of y_i.

(k) Plot a histogram for $\hat{\beta}$ for all four estimators using n = 300. What do you find? Are your findings conformable with the Central Limit Theorem (CLT)?

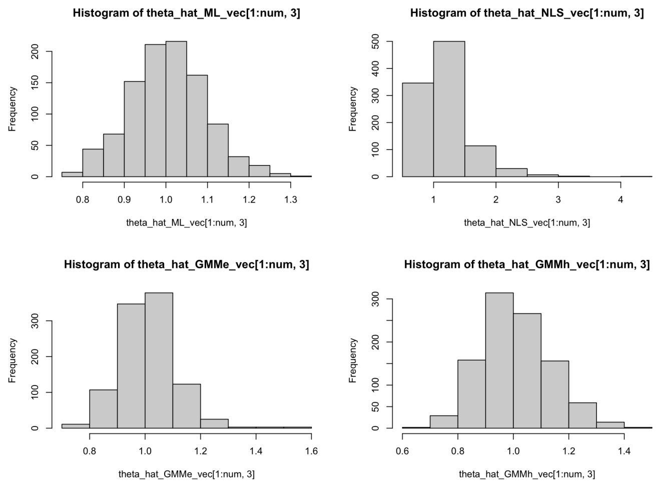

Figure 1: 1000 Simulation outcome

The nonlinear least squared estimator seem to be heavily right tailed, and does not exhibit CLT under current specification (sample size = 300, iteration = 1000). I plot the histogram for larger sample size, 1000, and the problem persists. 1000 iterations of Monte-Carlo should perform good by Law of Large Number. Due to such phenomenon, I verify if NLLSE here satisfies conditions for asymptotic theorem of extremum estimators in the Appendix.

(l) Report the 3×3 (sample) variance matrices of the four estimators over the 1000 Monte Carlo simulations for $n = 300$. Briefly comment on these matrices.

		$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\beta}$
MLE	$\hat{\alpha_1}$	0.0087	0.0051	0.0031
	$\hat{\alpha_2}$	0.0051	0.0109	0.0057
	$\hat{\beta}$	0.0031	0.0057	0.0089
NLS	$\hat{\alpha_1}$	0.1103	-0.1752	-0.0876
	$\hat{\alpha_2}$	-0.1752	0.3744	0.1967
	$\hat{\beta}$	-0.0876	0.1967	0.1179
Just-ID GMM	$\hat{\alpha_1}$	0.0226	-0.0114	-0.0030
	$\hat{\alpha_2}$	-0.0114	0.0309	0.0138
	$\hat{\beta}$	-0.0030	0.0138	0.0138
Over-ID GMM	$\hat{\alpha_1}$	0.0098	0.0081	0.0056
	$\hat{\alpha_2}$	0.0081	0.0204	0.0133
	$\hat{\beta}$	0.0056	0.0133	0.0152

Table 2: Variance-Covariance Matrix of $\hat{\theta}$ for Different Estimators

1. The $\hat{\theta}$ variance-covariance matrix estimation through Monte-Carlo simulation shows that $\hat{\theta}_{MLE}$ has the lowest variance-covariance, which corresponds to the property of Cramér–Rao Bounds for MLE. Note that when we prove the MLE reaches lowest variance-covariance, we prove against arbitary functional alternative from $\{w_i\}_i$ information set to θ.
2. By contrast, NLLSE has much larger covariance-variance matrix estimated compared with the other three estimators. This is improved through Just-ID GMM formulation with conditions transformed from NLLSE conditional expectation function. As argued before, the Just-ID GMM moment condition incorporate more variability from y_i itself rather than evaluating over a uniform parametric function $m(\cdot)$. Both estimation methods give some negative covariance between estimators in the same direction. Again, this might be caused by the parametric function $m(\cdot)$ that we are using in both case and needs further investigation. It might fade to 0 as we increase sample size and number of monte-carlo experiments.

(m) Report the average estimates over the 1000 Monte Carlo simulations (using $n = 300$) of $\left(\hat{J}_2(\hat{\theta})\right)^{-1}/n$. You may use the hessian spit out by R's optim function as your $\hat{J}_2(\hat{\theta})$. What do you find? How does this (average) estimate compare to your finding in (l)?

$$y_i = \begin{cases} 0 & \text{with probability } \Phi(\alpha_{10} - \beta_0 \cdot x_i) \\ 1 & \text{with probability } \Phi(\alpha_{20} - \beta_0 \cdot x_i) - \Phi(\alpha_{10} - \beta_0 \cdot x_i) \\ 2 & \text{with probability } \Phi(\beta_0 \cdot x_i - \alpha_{20}) \end{cases} \tag{17}$$

$$L_n(\theta) = \sum_{i=1}^{n} \sum_{j=0}^{2} \mathbb{1}\{Y_i = j\} \cdot log\mathbb{P}_j(Y_i = j|x_i, \theta) \tag{18}$$

The variance-covariance matrix of Maximum Likelihood Estimator defined with Hessian Matrix $\nabla\nabla_{\theta\theta}L_n(\theta)$ is computed as:

	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\beta}$
$\hat{\alpha_1}$	0.008349096	0.004913903	0.003020629
$\hat{\alpha_2}$	0.004913903	0.010950772	0.005628395
$\hat{\beta}$	0.003020629	0.005628395	0.008635254

Table 3: Variance-Covariance Matrix of $\hat{\theta_{MLE}}$ for Computed Through Hessian

The Hessian-computed variance-covariance matrix is very close to, almost coincide with, direct MoM estimator of the variance-covariance matrix by Monte-Carlo simulation and systematically smaller, which is reasonable as Hessian-evaluated MoM variance-covariance estimator should be more precise than simple MoM variance-covariance estimator.

(n) Write down Σ as defined in Theorem 4 (see slides) for the NLS estimator.

To achieve NLLSE satisfying asymptotic normality, the following conditions from general asymptotic normality theorem for extremum estimators should be satisfied:
i) $\theta_0 \in int(\Theta)$
ii) $Q_{n,NLLSE}$ is twice continuously differentiable in a small neighborhood of θ_0
iii) $\sqrt{n}\nabla_\theta Q_{n,NLLSE}(\theta_0) \to \mathcal{N}(0, \Sigma)$
iv) there exists $H(\theta)$ that is continuous at θ_0 and $\sup_{\theta\in\mathcal{N}} \|\nabla_{\theta\theta}Q_n(\theta) - H(\theta)\| \xrightarrow{p} 0$
v) $H \equiv H(\theta_0)$ is nonsingular.
Then $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1})$.
To derive conditions needed for NLLSE specified above for general condition iii),

$$\sqrt{n}\nabla_\theta Q_{n,NLLSE}(\theta_0) = \sqrt{n}\nabla_\theta \frac{1}{n}\sum_{i=1}^{n}\{y_i - m(x_i, \theta)\}^2 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial m(x_i, \theta_0)}{\partial\theta}e_i \xrightarrow{d} \mathcal{N}(0, \Sigma) \tag{19}$$

The last part is achieved through CLT with the convergence object $\Sigma = \mathbb{E}\left[e^2 \frac{\partial m(x,\theta_0)}{\partial\theta}\frac{\partial m(x,\theta_0)}{\partial\theta'}\right]$ (in addtion, we need regularity conditions like $E[y^4] < \infty$, $E\left|\frac{\partial m(x,\theta_0)}{\partial\theta}\right|^4 < \infty$ for CLT to apply). Substitute the functional form of $m(x_i, \theta_0)$, and we obtain

$$\Sigma = \mathbb{E}\left[e^2 \frac{\partial m(x,\theta_0)}{\partial \theta}\frac{\partial m(x,\theta_0)}{\partial \theta'}\right] = \mathbb{E}\left[(y - m(x,\theta_0))^2 \frac{\partial m(x,\theta_0)}{\partial \theta}\frac{\partial m(x,\theta_0)}{\partial \theta'}\right] \tag{20}$$

$$\frac{\partial m(x,\theta_0)}{\partial \theta} = \begin{bmatrix} \frac{\partial \Phi(\beta_0 x_i - a_{10}) + \Phi(\beta_0 x_i - a_{20})}{\partial \alpha_1} \\ \frac{\partial \Phi(\beta_0 x_i - a_{10}) + \Phi(\beta_0 x_i - a_{20})}{\partial \alpha_2} \\ \frac{\partial \Phi(\beta_0 x_i - a_{10}) + \Phi(\beta_0 x_i - a_{20})}{\partial \beta} \end{bmatrix} = \begin{bmatrix} -\phi(\beta_0 x_i - a_{10}) \\ -\phi(\beta_0 x_i - a_{20}) \\ -\phi(\beta_0 x_i - a_{10})x_i + \phi(\beta_0 x_i - a_{20})x_i \end{bmatrix} \tag{21}$$

$$\frac{\partial m(x,\theta_0)}{\partial \theta'} = \{\frac{\partial m(x,\theta_0)}{\partial \theta}\}^T \tag{22}$$

$$\Sigma = \mathbb{E}[(y - m(x,\theta_0))^2 \cdot \tag{23}$$

$$\begin{bmatrix} \phi^2(\beta_0 x - a_{10}) & \phi(\beta_0 x - a_{10}) \cdot \phi(\beta_0 x - a_{20}) & \phi(\beta_0 x - a_{10}) \cdot \{-\phi(\beta_0 x - a_{10}) + \phi(\beta_0 x - a_{20})\} \cdot x \\ ... & \phi^2(\beta_0 x - a_{20}) & \phi(\beta_0 x_i - a_{20}) \cdot \{-\phi(\beta_0 x - a_{10}) + \phi(\beta_0 x - a_{20})\} \cdot x \\ ... & ... & \{-\phi(\beta_0 x - a_{10}) + \phi(\beta_0 x - a_{20})\}^2 x^2 \end{bmatrix}] \tag{24}$$

(o) Derive G and Ω as defined in Theorem 7 (see slides) for the GMM estimator in (h).

The Over-ID GMM G is an mk(4x3) Jacobian matrix with respect to θ evaluated at θ_0, and Ω 4 by 4 weight matrix:

$$\begin{aligned} G &= \mathbb{E}[\frac{\partial h(w,\theta_0)}{\partial \theta}] \\ &= \mathbb{E}\begin{bmatrix} \frac{\partial \mathbb{1}(y_i=0) - p_0(x_i)}{\partial \alpha_1} & \frac{\partial \mathbb{1}(y_i=0) - p_0(x_i)}{\partial \alpha_2} & \frac{\partial \mathbb{1}(y_i=0) - p_0(x_i)}{\partial \beta} \\ \frac{\partial \{\mathbb{1}(y_i=0) - p_0(x_i)\} \cdot x_i}{\partial \alpha_1} & \frac{\partial \{\mathbb{1}(y_i=0) - p_0(x_i)\} \cdot x_i}{\partial \alpha_2} & \frac{\partial \{\mathbb{1}(y_i=0) - p_0(x_i)\} \cdot x_i}{\partial \beta} \\ \frac{\partial \mathbb{1}(y_i=1) - p_1(x_i)}{\partial \alpha_1} & \frac{\partial \mathbb{1}(y_i=1) - p_1(x_i)}{\partial \alpha_2} & \frac{\partial \mathbb{1}(y_i=1) - p_1(x_i)}{\partial \beta} \\ \frac{\partial \{\mathbb{1}(y_i=1) - p_1(x_i)\} \cdot x_i}{\partial \alpha_1} & \frac{\partial \{\mathbb{1}(y_i=1) - p_1(x_i)\} \cdot x_i}{\partial \alpha_2} & \frac{\partial \{\mathbb{1}(y_i=1) - p_1(x_i)\} \cdot x_i}{\partial \beta} \end{bmatrix} \\ &= \mathbb{E}\begin{bmatrix} -\phi(\alpha_1 - \beta \cdot x_i) & 0 & \phi(\alpha_1 - \beta \cdot x_i) \cdot x_i \\ -\phi(\alpha_1 - \beta \cdot x_i) \cdot x_i & 0 & \phi(\alpha_1 - \beta \cdot x_i) \cdot x_i^2 \\ \phi(\alpha_1 - \beta \cdot x_i) & -\phi(\alpha_2 - \beta \cdot x_i) & \{\phi(\alpha_2 - \beta \cdot x_i) - \phi(\alpha_1 - \beta \cdot x_i)\} \cdot x_i \\ \phi(\alpha_1 - \beta \cdot x_i) \cdot x_i & -\phi(\alpha_2 - \beta \cdot x_i) \cdot x_i & \{\phi(\alpha_2 - \beta \cdot x_i) - \phi(\alpha_1 - \beta \cdot x_i)\} \cdot x_i^2 \end{bmatrix} \end{aligned} \tag{25}$$

$$\Omega = \mathbb{E}[h(w,\theta_0)h(w,\theta_0)'] \tag{26}$$

$h(w, \theta_0)$ is defined in question (h).

$$\frac{\partial p_0(x_i)}{\partial \alpha_1} = \phi(\alpha_1 - \beta \cdot x_i) \tag{27}$$

$$\frac{\partial p_0(x_i)}{\partial \alpha_2} = 0 \tag{28}$$

$$\frac{\partial p_0(x_i)}{\partial \beta} = -x_i \cdot \phi(\alpha_1 - \beta \cdot x_i) \tag{29}$$

$$\frac{\partial p_1(x_i)}{\partial \alpha_1} = -\phi(\alpha_1 - \beta \cdot x_i) \tag{30}$$

$$\frac{\partial p_1(x_i)}{\partial \alpha_2} = \phi(\alpha_2 - \beta \cdot x_i) \tag{31}$$

$$\frac{\partial p_1(x_i)}{\partial \beta} = -x_i \cdot \phi(\alpha_2 - \beta \cdot x_i) + x_i \cdot \phi(\alpha_1 - \beta \cdot x_i) \tag{32}$$

$$\tag{33}$$

(p) Report the average estimate over the 1000 Monte Carlo simulations (using $n =$ 300) of $\left(\hat{G}'\hat{G}\right)^{-1} \hat{G}'\hat{\Omega}\hat{G} \left(\hat{G}'\hat{G}\right)^{-1} / n$ (i.e., the estimate of the finite-sample variance of GMM). The estimators $\hat{G}$ and $\hat{\Sigma}$ should be based on the G and Ω you obtained in (o); replace expected values by sample averages and θ_0 by $\hat{\theta}$. What do you find? How does this (average) estimate compare to your finding in (l)?

	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\beta}$
$\hat{\alpha_1}$	0.009435183	0.008105278	0.005630466
$\hat{\alpha_2}$	0.008105278	0.021123584	0.013681486
$\hat{\beta}$	0.005630466	0.013681486	0.015255727

Table 4: Estimated Variance-Covariance Matrix of $\hat{\theta}_{Over-ID-GMM}$ Computed Through Monte-Carlo of Asymptotic Variance-Covariance Matrix

The average estimates of asymptotic variance-covariance almost coincide with the Monte-Carlo average empirical variance for the over-identified GMM estimator.

5 Appendix

Verify the consistency of NLLSE estimator:

$$y_i = m(x_i, \theta) + v_i \quad \text{with} \quad \mathbb{E}[v_i \mid x_i] = 0 \tag{34}$$

$$m(x_i, \theta_0) = \Phi(\beta_0 x_i - a_{10}) + \Phi(\beta_0 x_i - a_{20}) \tag{35}$$

Rewrite the regression model with extremum estimator framework:

$$Q_{n,NLLSE}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\{y_i - m(x_i, \theta)\}^2 = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \{\Phi(\beta x_i - a_1) + \Phi(\beta x_i - a_2)\}\}^2 \tag{36}$$

Consistency Theorem of NLLSE
Suppose
1. $\{y_i, x_i\}_i$ i.i.d. and $\mathbb{E}[y|m] = m(x, \theta)$ is almost surely only if $\theta = \theta_0$ (conditional mean identification). 2. Θ is compact. 3. $m(x, \theta)$ almost surely continuous at each $\theta \in \Theta$. 4. $\mathbb{E}[y^2] < \infty$ and $\mathbb{E}[sup_{\theta\in\Theta}|m(x, \theta)|^2] < \infty$. Then, $\hat{\theta}_{NLLSE} \xrightarrow{p} \theta_0$.
Note that condition 2 is directly assumed. Condition 3 is satisfied because of continuity inherited from CDF of normal random variable. Condition 4 is satisfied with ordinal probit setting and boundedness of CDFs. The conditional mean identification, i.e. unqiue minimum of $m(x, \theta)$ over Θ, is achieved by the monotonicity of CDF and intake functions. Therefore, it is likely that the simulation setting generates this tailed-distribution of NLLSE estimators under a large asymptotic variance.