Labor Supply and the Two-Step Estimator

James J. Heckman
University of Chicago

Econ 350, Winter 2023
This draft, January 30, 2023

Outlines

Objective

Objective for this lecture

In this lecture, we look at a labor supply model and discuss various approaches to identify the key parameters of the model, including the two-step estimator.

A one period model of labor supply

Consider a simple one period model of the labor supply choice (with total time normalized to 1):

$$\max_{\{c,l\}} U(c,l) = \left(\frac{c^{\alpha}-1}{\alpha}\right) + b\left(\frac{l^{\phi}-1}{\phi}\right),$$

such that $c + wl \leq w + A$, where c is consumption, l is leisure, w is the wage rate, and A is non wage income.

The Euler equation is $w = \dfrac{bl^{\phi-1}}{c^{\alpha-1}}$ and the reservation wage is given by:

$$\begin{aligned} w_r &= \left[\frac{bl^{\phi-1}}{c^{\alpha-1}}\right]_{l=1,c=A} = \frac{b}{A^{\alpha-1}} \\ \Longrightarrow \ln w_r &= \ln b + (1-\alpha)\ln A \end{aligned}$$

Assume $\ln b = x\beta + e$, $e \perp\!\!\!\perp (x, A, w)$, $e \sim N(0, \sigma_e^2)$.

Consider three cases:

1. Wages are observed for everyone, i.e. for those participating ($l < 1$) and also for those not participating ($l = 1$);
2. Wages are not known for everyone, but we know the functional form for wages; and
3. Wages are observed for workers only.

We discuss the commonly used methodologies to identify the key parameters of the model in each of these cases below.

Consider three cases:

1. Wages are observed for everyone, i.e. for those participating ($l < 1$) and also for those not participating ($l = 1$);
2. Wages are not known for everyone, but we know the functional form for wages; and
3. Wages are observed for workers only.

We discuss the commonly used methodologies to identify the key parameters of the model in each of these cases below.

Observe wages for everyone

$$
\begin{aligned}
& \Pr(\text{Person works} \,|X, A) \\
= \; & \Pr(\ln w_r \leq \ln w | x, A) \\
= \; & \Pr\left(\frac{e}{\sigma_e} \leq \frac{\ln w - x\beta - (1-\alpha)\ln A}{\sigma_e}\right) \\
= \; & \Phi(c)
\end{aligned}
$$

where

$$
c = \frac{\ln w - x\beta - (1-\alpha)\ln A}{\sigma_e}.
$$

Each cell has common values of w_i, x_i, A_i. For each cell, obtain:

$$\hat{P}_i(D_i = 1|w_i, x_i, A_i) = \text{cell proportion working} = \Phi(\hat{c}_i).$$

Then calculate $\hat{c}_i = \Phi^{-1}(\hat{P}_i)$. Regress $\hat{c}_i$ on

$$\frac{\ln w - x\beta - (1-\alpha)\ln A}{\sigma_e},$$

and obtain estimates of σ_e, β, α.
Note that here, instead of standard normal (Φ), one could use a standard logistic model ($\Lambda(c) = \frac{e^c}{1+e^c}$) or a linear probability model:

$$F(c) = \frac{c}{c_u - c_l}, c_l \leq \frac{e}{\sigma_e} \leq c_u$$

Suppose $D_i = 1$ if agent i works, 0 if not. For each cell, get $\hat{p}(D = 1|w, A)$. By WLLN and Slutsky:

$$
\begin{aligned}
\text{plim}\Phi^{-1}(\hat{p}(D = 1|w, A)) &= \Phi^{-1}(\text{plim}\hat{p}(D = 1|w, A)) \\
&= \Phi^{-1}(p(D = 1|w, A)) \\
&= \frac{\ln w - x\beta - (1-\alpha)\ln A}{\sigma_e}.
\end{aligned}
$$

Set up the regression function:

$$
\begin{aligned}
\Phi^{-1}(\hat{p}) &= \Phi^{-1}(p) + \left[\Phi^{-1}(\hat{p}) - \Phi^{-1}(p)\right] \\
&= \frac{\ln w - x\beta - (1-\alpha)\ln A}{\sigma_e} + v.
\end{aligned}
$$

We need to characterize $v = \Phi^{-1}(\hat{p}) - \Phi^{-1}(p)$. By the delta method, assuming g is continuously differentiable, we get:

$$\sqrt{N}(g(\hat{p}) - g(p)) = \left.\frac{\partial g}{\partial p}\right|_{p^*} \sqrt{N}(\hat{p} - p),$$

where $\min(\hat{p}, p) \leq p^* \leq \max(\hat{p}, p)$. Now assume

$$\sqrt{N}(\hat{p} - p) \sim N(0, \sigma_p^2).$$

Applying this to our regression function, we get (using $c^* = \Phi^{-1}(p^*)$)

$$\sqrt{N_i}(\Phi^{-1}(\hat{p}_i) - \Phi^{-1}(p_i)) = \frac{1}{\phi(c_i^*)}\sqrt{N_i}(\hat{p}_i - p_i) \ \forall \text{ cells } i$$

Suppose $\{N_i\} \to \infty$ and errors are uncorrelated asymptotically. Then

$$v_i \xrightarrow{d} N\left(0, \left[\frac{1}{\phi(c_i^*)}\right]^2 \frac{p_i(1-p_i)}{N_i}\right),$$

where $\frac{p_i(1-p_i)}{N_i}$ is the variance of the binary random variable p_i. We obtain the feasible GLS estimator by regressing:

$$\frac{\frac{\Phi^{-1}(\hat{p}_i)}{\phi(c_i^*)}}{\sqrt{\frac{p_i(1-p_i)}{N_i}}} \quad \text{on} \quad \frac{\frac{\frac{\ln w - x\beta - (1-\alpha)\ln A}{\sigma_e}}{\phi(c_i^*)}}{\sqrt{\frac{p_i(1-p_i)}{N_i}}}$$

We can show that the estimates are asymptotically efficient and satisfy the orthogonality condition

$$\sum_{i=1}^{I} \operatorname{plim} \frac{1}{\phi(c_i^*)} \left(\frac{\ln w - x\beta - (1-\alpha)\ln A}{\sigma_e} \right) \left(\Phi^{-1}(\hat{p}_i) - \Phi^{-1}(p_i) \right) = 0.$$

$$L = \prod_{d_i=1} \Phi(c_i) \prod_{d_i=0} \Phi(-c_i) = \prod_i \Phi(c_i[2D_i - 1])$$

MLE gives consistent and asymptotically normal estimates of σ_e, β, α.

Consider three cases:

1. Wages are observed for everyone, i.e. for those participating ($l < 1$) and also for those not participating ($l = 1$);
2. Wages are not known for everyone, but we know the functional form for wages; and
3. Wages are observed for workers only.

We discuss the commonly used methodologies to identify the key parameters of the model in each of these cases below.

Do not observe wages, but wages follow specific functional form

Here we do not observe wages for anyone, but do know that the wages have the following functional form:

$$\ln w = z\gamma + u,$$

where

$$u \sim N(0, \sigma_u^2), (e - u) \perp\!\!\!\perp x, A, z,$$

and

$$(e - u) \sim N(0, \sigma_e^2 + \sigma_u^2 - 2\sigma_{eu}) = N(0, \sigma^{*2}),$$

where $(\sigma^*)^2 = (\sigma_u^2 + \sigma_e^2 - 2\sigma_{ue})$.

Then:

$$\Pr(i \text{ works}) = \Pr\left(\frac{e-u}{\sigma^*} \leq \frac{z\gamma - x\beta - (1-\alpha)\ln A}{\sigma^*}\right).$$

Note: if (e, u) are Extreme Value (Type I), then $(e-u)$ is logistic.

1. If z, x distinct, then can estimate $\left(\frac{\gamma}{\sigma^*}, \frac{\beta}{\sigma^*}, \frac{1-\alpha}{\sigma^*}\right)$, but can't estimate σ^*.
2. If z, x have elements in common $(x_c = z_c, x_d, z_d)$, then can estimate only:

$$\left(\frac{\beta_d}{\sigma^*}, \frac{\gamma_d}{\sigma^*}, \frac{\gamma_c - \beta_c}{\sigma^*}, \frac{(1-\alpha)}{\sigma^*}\right)$$

and again not σ^*.

Consider three cases:

1. Wages are observed for everyone, i.e. for those participating ($l < 1$) and also for those not participating ($l = 1$);
2. Wages are not known for everyone, but we know the functional form for wages; and
3. Wages are observed for workers only.

We discuss the commonly used methodologies to identify the key parameters of the model in each of these cases below.

Observe wages for workers only

Here we have:

$$
\begin{aligned}
\ln w &= z\gamma + u \\
\ln w_r &= x\beta + (1-\alpha)\ln A + e \\
(e,u) \perp\!\!\!\perp & \quad (x,z,A) \\
y_i = \ln w - \ln w_r &= z\gamma - x\beta - (1-\alpha)\ln A + u - e.
\end{aligned}
$$

Agent i works if $y_i > 0$. This is the Roy model with 2 sectors: market (u) and nonmarket (e).

Following the derivations in the lecture 'Empirical Content of Roy Model,' we get the expression for the expected wages in the market sector, conditional on participation and the x and z variables as:

$$\begin{aligned} & E\left(\ln w \mid \ln w > \ln w_r, x, z, A\right) \\ = \; & z\gamma + \frac{\sigma_{uu} - \sigma_{ue}}{\sigma^*}\lambda\left(\frac{x\beta - z\gamma + (1-\alpha)\ln A}{\sigma^*}\right) \\ = \; & z\gamma + \frac{\sigma_{uu} - \sigma_{ue}}{\sigma^*}\lambda(-c), \end{aligned}$$

where $c = \dfrac{z\gamma - x\beta - (1-\alpha)\ln A}{\sigma^*}$.

- Step 1: Run probit on LFP (labor force participation) decision (as we did in section 4) :

$$(\frac{\hat{\gamma}}{\sigma^*}, \frac{\hat{\beta}}{\sigma^*}, \frac{\hat{\alpha}}{\sigma^*}) = \text{argmax} \sum_i \ln \Phi[c_i(2D_i - 1)]$$

Form:

$$\begin{aligned} \lambda(-\hat{c}_i) &= \frac{\phi(-\hat{c}_i)}{1 - \Phi(-\hat{c}_i)}, \\ \text{where } \hat{c}_i &= \frac{z_i\hat{\gamma} - x_i\hat{\beta} - (1-\hat{\alpha})\ln A_i}{\hat{\sigma}^*} \end{aligned}$$

- Step 2: Estimate $(\hat{\gamma}, \widehat{\frac{\sigma_{uu} - \sigma_{ue}}{\sigma^*}})$ via OLS on

$$\ln w_i = z_i\gamma + \frac{\sigma_{uu} - \sigma_{ue}}{\sigma^*}\lambda(-\hat{c}_i) + v_i$$

using sample of workers only (refer to expression for conditional expectation of market wages derived above).

With one exclusion restriction (1 variable, call it z_1, in z not in x or $\ln A$; let all other z be common with x), we can now identify everything:

$$(\gamma, \beta, \alpha, \sigma^*, \sigma_{uu}, \sigma_{ee}, \sigma_{ue}) \,.$$

We describe below how we recover all the relevant parameters:

(i) Step 1 of 2-step gives $\frac{\gamma_1}{\sigma^*}$ as well as $\left(\frac{(\gamma-\beta)_c}{\sigma^*}, \frac{1-\alpha}{\sigma^*}\right)$.

Step 2 gives γ_1 as well as $\left(\gamma_c, \frac{\sigma_{uu}-\sigma_{ue}}{\sigma^*}\right)$. Solve for σ^*.

Use σ^* to solve for $(\sigma_{uu}-\sigma_{ue}, \beta, \alpha)$.

(ii) Look at residuals from step 2:

$$\ln w_i = z_i\gamma + \frac{\sigma_{uu}-\sigma_{ue}}{\sigma^*}\lambda(-\hat{c}_i) + v_i$$

Then following results in an earlier lecture, we have:

$$\begin{aligned} E(v_i^2) &= \sigma_{uu}\left(\rho^2\left[1-\lambda^2(-c_i)-\lambda(-c_i)c_i\right]+\left[1-\rho^2\right]\right) \\ &= \sigma_{uu}-\sigma_{uu}\rho^2\left[\lambda_i^2+\lambda_i c_i\right]. \end{aligned}$$

Regression of $\hat{v}_i^2$ on $\hat{\lambda}_i^2 + \hat{\lambda}_i\hat{c}_i$ gives consistent estimates of (σ_{uu}, ρ). Solve for σ_{ue}. Use $\sigma^* = \sigma_{uu} + \sigma_{ee} - 2\sigma_{ue}$ to solve for σ_{ee}.

Without any exclusion restrictions on z, we can only identify (γ, σ_{uu}): We cannot uniquely identify σ_{ee} or σ_{ue}.

(i) Step 2 of 2-step gives $\left(\gamma, \dfrac{\sigma_{uu} - \sigma_{ue}}{\sigma^*}\right)$.

(ii) Obtain (σ_{uu}, ρ) from residual regression as above.

(iii) To solve for $(\sigma_{ee}, \sigma_{ue})$ either normalize $\sigma_{ee} = 1$ or $\sigma_{ue} = 0$. If we assume $\sigma_{ee} = 1$, then from step 2 we obtain $\dfrac{\sigma_{uu} - \sigma_{ue}}{1 + \sigma_{uu} - 2\sigma_{ue}}$, from which we can obtain σ_{ue}. If we assume $\sigma_{ue} = 0$, then from step 2 we get $\dfrac{\sigma_{uu}}{\sigma_{uu} + \sigma_{ee}}$ and can solve for σ_{ee}.

Durbin's problem (1970)

(See also Newey (1984) and Newey and McFadden (1994), and also refer handout on the Durbin problem.)
In step 2 of the two-step estimation, setting for simplicity $-\hat{c} = x\hat{\beta}$, we have the regression:

$$\ln w = z\gamma + \sigma\lambda(x\hat{\beta}) + \sigma\left[\lambda(x\beta) - \lambda(x\hat{\beta})\right] + v.$$

OLS gives consistent estimates of γ, σ but the variance of the OLS estimates γ, σ equal the usual OLS variance matrix plus an additional term due to the $\sigma(\lambda - \hat{\lambda})$ term, so we have heteroskedasticity and extra variability.

$$
\begin{aligned}
\lambda(x\hat{\beta}) &= \lambda(x\beta) + \frac{\partial \lambda}{\partial c} x(\hat{\beta} - \beta) + o(\cdot) \\
\sqrt{N}(\hat{\lambda} - \lambda) &= \frac{\partial \lambda}{\partial c} \sqrt{N} x(\hat{\beta} - \beta) + o(\cdot) \\
&\xrightarrow{d} N\left(0, \frac{\partial \lambda}{\partial c} x \Sigma_{\beta} x' \frac{\partial \lambda}{\partial c}'\right) \\
\Sigma_{\beta} &= asy.var\left(\sqrt{N}(\hat{\beta} - \beta)\right).
\end{aligned}
$$

Sampling distribution of the OLS coefficient is:

$$
\begin{pmatrix} \hat{\gamma} \\ \hat{\sigma} \end{pmatrix} = \begin{pmatrix} \gamma \\ \sigma \end{pmatrix} + \begin{pmatrix} z'z & z'\hat{\lambda} \\ \hat{\lambda}'z & \hat{\lambda}'\hat{\lambda} \end{pmatrix}^{-1} \begin{pmatrix} z \\ \hat{\lambda} \end{pmatrix} \left(\sigma(\hat{\lambda} - \lambda) + v\right),
$$

where $Z'Z = \sum_{i=1}^{n} Z_i Z_i'$, $Z'\hat{\lambda} = \sum_{i=1}^{n} Z_i \lambda_i$, etc.

Rearranging we get:

$$\sqrt{N}\left[\begin{pmatrix}\hat{\gamma}\\ \hat{\sigma}\end{pmatrix}-\begin{pmatrix}\gamma\\ \sigma\end{pmatrix}\right]$$
$$=\begin{pmatrix}\frac{z'z}{N} & \frac{z'\hat{\lambda}}{N}\\ \frac{\hat{\lambda}'z}{N} & \frac{\hat{\lambda}'\hat{\lambda}}{N}\end{pmatrix}^{-1}\begin{pmatrix}\frac{z}{\sqrt{N}}\\ \frac{\hat{\lambda}}{\sqrt{N}}\end{pmatrix}\left(\sigma(\hat{\lambda}-\lambda)+v\right)$$

Taylor expanding around the true β and taking probability limits of each element on the rhs gives:

$$\frac{z'\hat{\lambda}}{N} = \frac{z'\left[\lambda + \frac{\partial \lambda}{\partial c} x(\hat{\beta} - \beta)\right]}{N} = \frac{z'\lambda}{N} + \frac{z'\frac{\partial \lambda}{\partial c} x(\hat{\beta} - \beta)}{N} \xrightarrow{p} \frac{z'\lambda}{N}$$

$$\begin{aligned}
\frac{\hat{\lambda}'\hat{\lambda}}{N} &= \frac{\left[\lambda + \frac{\partial \lambda}{\partial c} x(\hat{\beta} - \beta)\right]'\left[\lambda + \frac{\partial \lambda}{\partial c} x(\hat{\beta} - \beta)\right]}{N} \\
&= \frac{\lambda'\lambda}{N} + 2\frac{\lambda'\frac{\partial \lambda}{\partial c} x}{N\sqrt{N}}\sqrt{N}(\hat{\beta} - \beta) \\
&\quad + \left[\frac{\frac{\partial \lambda}{\partial c} x}{N}\sqrt{N}(\hat{\beta} - \beta)\right]'\left[\frac{\frac{\partial \lambda}{\partial c} x}{N}\sqrt{N}(\hat{\beta} - \beta)\right] \\
&\xrightarrow{p} \frac{\lambda'\lambda}{N}
\end{aligned}$$

$$\begin{aligned}
\frac{z'\sigma(\hat{\lambda}-\lambda)}{\sqrt{N}} &= \frac{-z'\sigma(\frac{\partial\lambda}{\partial c}x(\hat{\beta}-\beta))}{\sqrt{N}} \\
&= -\sigma\frac{z'\left(\frac{\partial\lambda}{\partial c}\right)x}{N}\sqrt{N}(\hat{\beta}-\beta) \\
\xrightarrow{p} -\sigma\Sigma_1 N(0,\Sigma_\beta) &= N(0,\sigma^2\Sigma_1\Sigma_\beta\Sigma_1'),
\end{aligned}$$

$$\frac{z'v}{\sqrt{N}} \xrightarrow{d} N(0,\sigma_z^2\sigma_v^2)$$

$$\frac{\hat{\lambda}'[\sigma(\lambda-\hat{\lambda})+v]}{\sqrt{N}} = -\sigma\frac{\left[\lambda+\frac{\partial\lambda}{\partial c}x(\hat{\beta}-\beta)\right]'\left[\frac{\partial\lambda}{\partial c}x(\hat{\beta}-\beta)\right]}{\sqrt{N}} + \frac{\left[\lambda+\frac{\partial\lambda}{\partial c}x(\hat{\beta}-\beta)\right]' v}{\sqrt{N}}$$

$$
\begin{aligned}
= \quad & -\sigma \frac{\lambda' \left[\frac{\partial \lambda}{\partial c} x(\hat{\beta} - \beta) \right]}{\sqrt{N}} \\
& -\sigma \frac{\left[\frac{\partial \lambda}{\partial c} x(\hat{\beta} - \beta) \right]' \left[\frac{\partial \lambda}{\partial c} x(\hat{\beta} - \beta) \right]}{\sqrt{N}} \\
& + \frac{\lambda' v}{\sqrt{N}} + \frac{\left[\frac{\partial \lambda}{\partial c} x(\hat{\beta} - \beta) \right]' v}{\sqrt{N}}
\end{aligned}
$$

$$
\begin{aligned}
= \quad & -\sigma \frac{\lambda' \left[\frac{\partial \lambda}{\partial c} x \right]}{N} \sqrt{N}(\hat{\beta} - \beta) \\
& -\sigma \frac{\left[\frac{\partial \lambda}{\partial c} x \right]' \left[\frac{\partial \lambda}{\partial c} x \right]}{N\sqrt{N}} \left[\sqrt{N}(\hat{\beta} - \beta) \right]' \left[\sqrt{N}(\hat{\beta} - \beta) \right] \\
& + \frac{\lambda' v}{\sqrt{N}} + \frac{\left[\frac{\partial \lambda}{\partial c} x \right]' v}{N} \sqrt{N}(\hat{\beta} - \beta)
\end{aligned}
$$

$$\xrightarrow{P} -\sigma \frac{\lambda^{/} \left[\frac{\partial \lambda}{\partial c} x\right] \sqrt{N}(\hat{\beta} - \beta)}{N}$$

$$-\sigma \frac{\left[\frac{\partial \lambda}{\partial c} x\right]' \left[\frac{\partial \lambda}{\partial c} x\right]}{N\sqrt{N}} \left[\sqrt{N}(\hat{\beta} - \beta)\right]' \left[\sqrt{N}(\hat{\beta} - \beta]\right]$$

$$+\frac{\lambda' v}{\sqrt{N}}$$

$$\xrightarrow{P} -\sigma\Sigma_2 N(0,\Sigma_\beta) = N(0,\sigma^2\Sigma_2\Sigma_\beta\Sigma_2') + N(0,\sigma_\lambda^2\sigma_v^2),$$

where

$$\Sigma_2 = \text{plim}\frac{\left[\lambda'\left(\frac{\partial\lambda}{\partial c}x\right)\right]}{N},$$

assuming

$$v \perp\!\!\!\perp \frac{\partial\lambda}{\partial c}x \Rightarrow \text{plim}\frac{\left[\frac{\partial\lambda}{\partial c}x\right]' v}{N} = 0$$

and $\frac{\lambda' v}{\sqrt{N}} \xrightarrow{d} N(0,\sigma_\lambda^2\sigma_v^2)$.

Putting this all together and assuming that the random components of the first and second step are independent (i.e. sequence of estimates $\hat{\beta}$ is independent of v), we get:

$$\sqrt{N}\left[\left(\begin{array}{c}\hat{\gamma}\\ \hat{\sigma}\end{array}\right)-\left(\begin{array}{c}\gamma\\ \sigma\end{array}\right)\right]\xrightarrow{d} N\left[0,V\right],$$

where

$$V=\sigma_v^2 Q_0^{-1}+\sigma^2 Q_0^{-1}Q_1\Sigma_\beta Q_1' Q_0^{-1}$$

$$Q_0\equiv E\left(\begin{array}{cc}z'z & z'\lambda\\ \lambda' z & \lambda'\lambda\end{array}\right), Q_1\equiv\left(\begin{array}{c}\Sigma_1\\ \Sigma_2\end{array}\right).$$

Note that since $\hat{\beta}$ is estimated using MLE, we have:

$$\Sigma_\beta=-E\left[\frac{\partial^2 f(x;\beta)}{\partial\beta\partial\beta'}\right]^{-1}$$

In the non-independence case we have shown that:

$$\sqrt{N}\left[\begin{pmatrix}\hat{\gamma}\\ \hat{\sigma}\end{pmatrix}-\begin{pmatrix}\gamma\\ \sigma\end{pmatrix}\right]\xrightarrow{d} N[0,\Sigma],$$

where:

$$\Sigma=\sigma_v^2 Q_0^{-1}+\sigma^2 Q_0^{-1}\left[\begin{array}{l} Q_1 R_1^{-1}(\beta)Q_1' \\ -Q_1 R_1^{-1}(\beta)Q_2' - Q_2 R_1^{-1}(\beta)Q_1' \end{array}\right]Q_0^{-1}$$

with:

$$W\equiv[z\lambda],\;\; R_1(\beta)=-E\left[\frac{\partial^2 f(x;\beta)}{\partial\beta\partial\beta'}\right]^{-1},\;\; Q_0=\text{plim}\frac{1}{n}[W'W],$$

$$Q_1=\text{plim}\frac{1}{n}\left[W'\left[\frac{\partial\lambda}{\partial c}x\right]\right],\;\; Q_2=\text{plim}\frac{1}{n}\sum_{i=1}^{N}W_i'\frac{\partial f}{\partial\beta}(x;\beta)$$