The Probit & Logit Models

Quantitative Microeconomics

R. Mora

Department of Economics
Universidad Carlos III de Madrid

Outline

1. Motivation: The Labor Market Participation Decision
2. Estimation of $\Pr(work = 1|x)$ for women at working age
3. The Probit and Logit Models

The Consumption-Leisure Choice

Utility Function

- $U = U(C, L)$
- C: consumption
- L: leisure

- Consumption: $U_C = \left.\frac{\partial U}{\partial C}\right|_L > 0$: more consumption gives more utility...
  - $\left.\frac{\partial U_C}{\partial C}\right|_L < 0$: but at a decreasing rate
- Leisure: $U_L = \left.\frac{\partial U}{\partial L}\right|_C > 0$: additional leisure gives additional utility...
  - $\left.\frac{\partial U_L}{\partial L}\right|_C < 0$: but at a decreasing rate

By how much can I reduce my consumption without losing utility if I increase my leisure?

Marginal Rate of Substitution

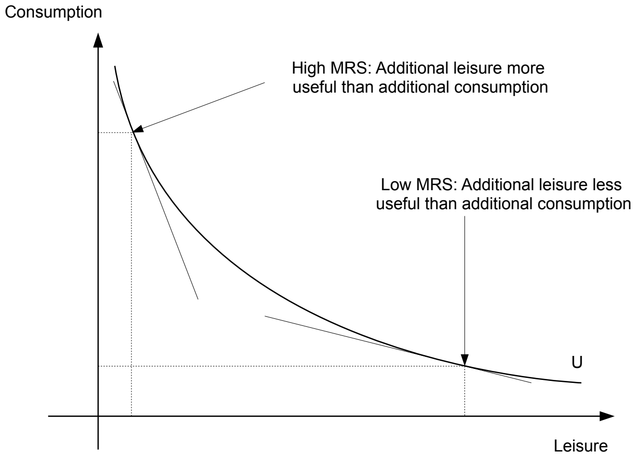
- $MRS = \left.\frac{\partial C}{\partial L}\right|_U = -\frac{U_L}{U_C}$

- The *MRS* gives the individual's value of leisure in terms of consumption.

Cobb-Douglas: $U = C^\alpha L^\beta \rightarrow MRS = \left(\frac{\alpha}{\beta}\right)\left(\frac{C}{L}\right)$

- Increasing in consumption
- Decreasing in leisure

A Graphical Interpretation



Time and Budget Constraints

Constraints

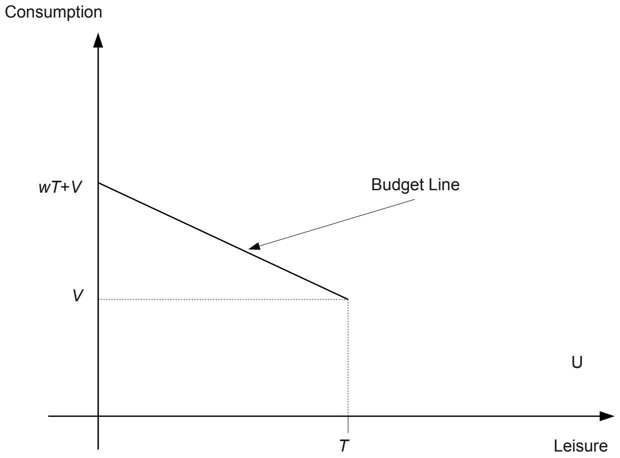
- Time constraint: $L + h = T$, h: hours of work, T: total hours available
- Budget constraint: $C = w * h + V$, w: hourly wage, V: non-labour income

- Replacing $h = T - L$ in the budget constraint, we get

$$C + wL = wT + V$$

where $wT + V$ (time and non-labor income) equals consumption plus the cost of leisure (w is the market opportunity cost of leisure in terms of consumption)

The Budget Line



The Optimal Allocation of Leisure

max $U(C, L)$ s.t. $C + wL = wT + V$

- $MRS > w$: a small increase in leisure will increase utility
- $MRS < w$: a small increase in work will increase utility (via higher consumption)

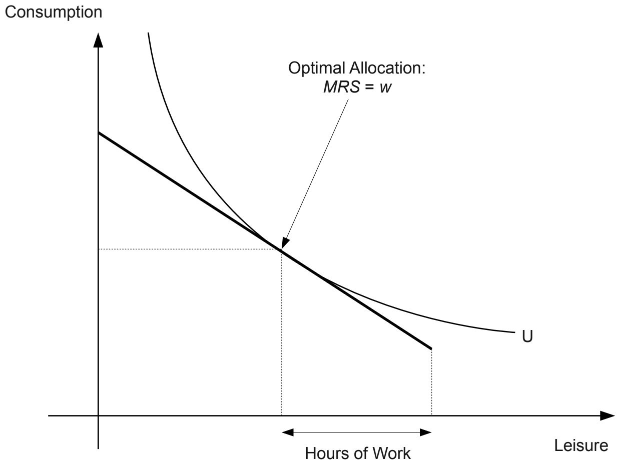
- Internal Solution: the individual's value of leisure equals its market value

$$MRS = w \qquad L^* < T \qquad h^* > 0$$

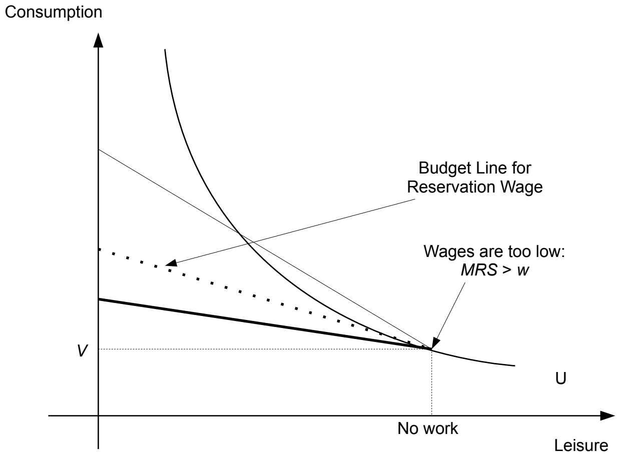
- Corner Solution: the individual's value of leisure is larger than the market value

$$MRS > w \qquad L^* = T \qquad h^* = 0$$

Internal Solution



Corner Solution



The Reservation Wage

$$w_R = MRS(T, V)$$

Individuals work if the wage is larger than their reservation wage

- For any $w > w_R$: Internal Solution ($h^* > 0$)
- For any $w \leq w_R$: Corner solution ($h^* = 0$)

- An increase in non-labor income cannot increase h^* if leisure is a normal good.
- An increase in the market wage:
  - Increases the opportunity cost of leisure (substitution effect).
  - Expands the budget constraint (income effect).

The Decision to Participate

- We assume two options
  - Working full time ($h^* > 0$ and $work = 1$)
  - Leisure full time (could include working full-time in the household) ($h^* = 0$ and $work = 0$)

If $w > w_R$ then the individual works ($work = 1$)

- We can think of w as the value of choosing to work in the market.
- We can think of w_R as the value of choosing not to work in the market.

What are the factors that affect the probability of participation, $\Pr(work = 1|x)$?

Everything that changes the probability of $(w > w_R)$

- If substitution effect > income effect, $\uparrow w \rightarrow \uparrow \Pr(work = 1|x)$
  - more education $\rightarrow \uparrow w \rightarrow \uparrow \Pr(work = 1|x)$
- If leisure is normal, $\uparrow V \rightarrow \downarrow \Pr(work = 1|x)$
  - husband becomes unemployed $\rightarrow \downarrow V \rightarrow \uparrow \Pr(work = 1|x)$
- $\uparrow MRS \rightarrow \downarrow \Pr(work = 1|x)$:
  - If the individual needs a lot of time to care for/help members of her family (children, elderly), then her personal value of leisure will be large (high *MRS*).
  - An additional kid in the family may increase *MRS* for the woman and make market work undesirable for her.

The available data

- Population: All women at working age.
- The dependent variable is whether the woman works ($work = 1$) or not ($work = 0$)
- Controls:
  - Family characteristics: Number of kids, employment status of the husband, non-labor income, a relative suffers a long-term disability...
  - Personal characteristics: Level of education, work experience, measures of ability and skills,...(Note that wages are only observed for those women who choose to participate.)
  - Market and Economic characteristics: local market unemployment rates, local wages, ...

$\Pr(work = 1|x) = \beta_0 + x_1\beta_1 + ... + x_k\beta = x'\beta$

- This is the Linear Probability Model.
- As *work* is binary: $\Pr(work = 1|x) = \mathrm{E}(work|x)$
  - OLS is consistent and inference can be carried out as usual if using robust standard errors.
- Fundamental problem: The linear assumption is impossible if x can take any value (like *income*) because the probability must lie between 0 and 1.
- Practical problem: the estimated model may predict negative probabilities or probabilities larger than 1.
- Solution: Non-linear model.
  - linear with kinks: difficult to estimate, beyond the goal of this course.
  - non-linear random utility model (this is the most popular solution).

The Random Utility Model

If $w > w_R$ then the individual works in the market ($work = 1$)

- We can think of $w = U_m$ as the value of choosing to work, and of $w_R = U_h$ as the value of choosing not to work

- The value of each alternative depends on many factors:

$$U_m = x_m' \beta_m + \varepsilon_m$$
$$U_h = x_h' \beta_h + \varepsilon_h$$

where $\varepsilon_m, \varepsilon_h$ are effects on utility unobserved to the econometrician.

$$\text{If } x_m' \beta_m + \varepsilon_m \geq x_h' \beta_h + \varepsilon_h \text{ then } work = 1$$

if $U_m = U_h$ there is indecision, but this happens with zero probability if ε_m and ε_h are continuous random variables

$x_m'\beta_m + \varepsilon_m \geq x_h'\beta_h + \varepsilon_h \rightarrow work = 1$

Assumption

$$\varepsilon_m - \varepsilon_h = \varepsilon \sim F_\varepsilon$$

where F_ε is symmetric.

Let $x'\beta = x_m'\beta_m - x_h'\beta_h$. Thus

$$\Pr(work = 1|x) = \Pr\left(x'\beta + \varepsilon \geq 0|x\right) = \Pr\left(\varepsilon \geq -\left(x'\beta\right)|x\right)$$

Given symmetry,

$$\Pr(work = 1|x) = \Pr\left(\varepsilon \leq x'\beta|x\right) = F_\varepsilon\left(x'\beta\right)$$

Probit Model: $\varepsilon_h, \varepsilon_m \sim N(0, \Sigma)$ such that $\varepsilon \sim N(0,1)$

$$\Pr(work = 1|x) = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt.$$

Logit Model: $\varepsilon_h, \varepsilon_m$ such that $\varepsilon \sim$Logistic distribution

$$\Pr(work = 1|x) = \Lambda(x'\beta) = \frac{exp(x'\beta)}{1+exp(x'\beta)}$$

- In both models,
  - Probabilities lie between 0 and 1 by construction.
  - β can be consistently estimated by ML.
- The standard normal has variance 1 while the Logistic has variance $\frac{\pi^2}{3}$.

Summary

- The probability of participating in the labor market depends on personal characteristics, family characteristics, and market characteristics.
- OLS techniques are usually not the most appropriate because the conditional expectation equals the probability that the dependent variable takes value 1, and probabilities must lie between 0 and 1.
- Two appropriate models are the Probit and the Logit models.
- Both can be estimated using ML techniques .