Advanced Applied Econometrics
Static discrete choice labour supply

Summer 2024

Stochastic discrete choice labour supply models are very popular in empirical research and have frequently been used as a tool to perform tax policy analysis. In this problem set we will examine a basic example of this class of model, and for simplicity we will abstract from any demographic heterogeneity (although it is straightforward to incorporate this).

Suppose that individuals have the choice to work $h \in [0, 10, 20, 30, 40]$ hours per week. Preferences over these discrete alternatives may be described by a parametric utility function:

$$U(c, h) = \gamma \left[\frac{c^{\theta}}{\theta} - \alpha \, h \right] + \varepsilon_h$$

where the state-specific errors ε_h are assumed to follow a Type-I extreme value distribution.

Consumption is given by $c = y + wh - T(wh)$, where y is non-labour income, $T()$ is the tax system, and w is the gross hourly wage rate generated by the following log-linear relationship: $\log w = \mu_w + \varepsilon_w$, and where the unobserved component of wages ε_w is Normally distributed with mean 0 and standard deviation σ_w.

Suppose that the parameter values are:
$\Omega = \mu_w = 1, \sigma_w = 0.55, \theta = 0.3, \alpha = 0.1, \gamma = 2$ and that $y \sim$Uniform$[10, 100]$. Suppose also that earnings below €80 per week are not taxed; any earnings greater than €80 are taxed at the constant marginal tax rate $t = 0.3$. Non-labour income is not taxed. Under these assumptions, simulate a dataset of 1000 observations.

1. What is the distribution of work hours in your simulated dataset?

2. How does the distribution of non-labour income and wages vary with work hours?

3. Write down the log-likelihood function for this model. Note, the likelihood function differs between workers and non-workers.

4. Write your own code to estimate the model using the dataset you generated above (which comprises work hours, non-labour income, and wages only for workers). Having estimated the model, check that the parameter estimates are close to those you used to generate the dataset.