Binary probit panel model

Definition

Net utility from taking action in t is $U_{nt} = V_{nt} + \varepsilon_{nt}$ and person takes action if $U_{nt} > 0$. This is the net utility, since it is difference between utility of taking action and that of not taking action. The errors are correlated over time, and the covariance matrix for $\varepsilon_n = (\varepsilon_{n1}, \dots, \varepsilon_{nT})$ is $\Omega \in \mathbb{R}^{T \times T}$. The sequence of actions that person n takes can be represented by a set of T dummy variables: $d_{nt} = 1$ if person n took action at t and $d_{nt} = -1$ otherwise. The probability of sequence if choices $d_n = (d_{n1}, \dots, d_{nT})$ is

$$\begin{aligned}
P(d_n|V_n) &= P(U_{nt}d_{nt} > 0 \; |V_n, \; \forall t) \\
&= P(V_{nt}d_{nt} + \varepsilon_{nt}d_{nt} > 0 \; |V_n \; \forall t) \\
&= \int_{\varepsilon_n \in B_n} \phi_\Omega(\varepsilon_n) d\varepsilon_n
\end{aligned}$$

where B_n is a set such that $V_{nt}d_{nt} + \varepsilon_{nt}d_{nt} > 0$ **for all** t and $\phi_\Omega(\varepsilon_n)$ is a joint normal density with covariance matrix Ω. Once again, mathematically:

$$B_n = \{\varepsilon_n = (\varepsilon_{n1}, \dots, \varepsilon_{nT}) \mid V_{nt}d_{nt} + \varepsilon_{nt}d_{nt} > 0 \quad \forall t = 1...T\}$$

Structure on error covariance

Structure can be placed on the covariance of the errors over time. Further we will suppose, that the error consists of a portion that is specific to the decision maker, reflecting his proclivity to take the action, and a part that varies over time for each decision maker: $\varepsilon_{nt} = \eta_n + \mu_{nt}$, where μ_{nt} is iid over time and people with a standard normal density, and η_n is iid over people with a normal density with zero mean and variance σ. And $V(\varepsilon_{nt}) = V(\eta_n + \mu_{nt}) = V(\eta_n) + V(\mu_{nt}) = \sigma + 1$. Covariance between two time periods s and t is

$$\mathrm{Cov}(\varepsilon_{nt}, \varepsilon_{ns}) = \mathrm{Cov}(\eta_n + \mu_{nt}, \eta_n + \mu_{ns}) = \sigma$$

Therefore covariance matrix Ω takes form of:

$$\Omega = \begin{pmatrix} \sigma + 1 & \sigma & \dots & \sigma \\ \sigma & \sigma + 1 & \dots & \sigma \\ \dots & \dots & \dots & \dots \\ \sigma & \dots & \dots & \sigma + 1 \end{pmatrix}$$

Only one parameter σ, enters the covariance matrix. its value indicates the variance in unobserved utility across indidividuals (the variance of η_n) relative to the variance across time for each individual (the variance of μ_{nt}).

Estimation of the parameters

In order to estimate parameters we need to simulate the choice probabilities. It is advised by *Train* to use error partioning approach. Conditional on η_n, the probability of not taking the action in period t is $P(V_{nt} + \eta_n + \mu_{nt} < 0) = P(\mu_{nt} < -(V_{nt} + \eta_n)) = \Phi(-(V_{nt} + \eta_n))$, where Φ is cdf of standard normal distribution. Probability of taking action is then: $1 - \Phi(-(V_{nt} + \eta_n)) = \Phi(V_{nt} + \eta_n)$. The probability of the sequence of choices $d_n = (d_{n1}, \ldots, d_{nT})$ is then:

$$P(d_n|\eta_n, V_n) = \prod_{t=1}^{T} \Phi((V_{nt} + \eta_n)d_{nt})$$

So far we have conditioned on η_n, when in fact η_n is random. The unconditional probability is the integral of the conditional probability $P(d_n|\eta_n, V_n)$ over all possible values of η_n:

$$P(d_n|V_n) = \int_{\mathbb{R}} P(d_n|\eta_n, V_n)\phi_\sigma(\eta_n)d\eta_n$$

where $\phi_\sigma(\eta_n)$ is normal density with zero mean and variance σ. Overall result:

$$P(d_n|V_n) = \int_{\mathbb{R}} \prod_{t=1}^{T} \Phi((V_{nt} + \eta_n)d_{nt})\phi_\sigma(\eta_n)d\eta_n$$

Once again $P(d_n|V_n)$ is unconditional (w.r.t η_n) probability that agent n will choose sequence of choices $d_n = (d_{n1}, \ldots, d_{nT})$, where each action d_{nt} is either 1 if action is taken and -1 otherwise. *Train* (ch 5, pg. 17) proposes to use the following algorithm for $P(d_n|V_n)$ estimation:

1. Take a draw of η_n from $\phi_\sigma(\eta_n)$
2. For this draw of η_n, calculate $P(d_n|\eta_n, V_n)$
3. Repeat steps 1–2 many times, and average the results. This average is a simulated approximation to $P(d_n|V_n)$

So far we have conditioned all of our probabilities on V_n, but we have data as inputs, namely agents' and attributes' characteristics which are described as $x_n = (x_{n1}, \ldots . x_{nT})$. Since we incorporate taste variation, $V_n = \beta_n' x_n$ becomes random, since β_n is a random vector. In order to estimate the model we should understand likelihood function. Since we have binary choice situation, lets stick to the following notation: probability of agent n choosing sequence of choices $d_n = (d_{n1}, \ldots, d_{nT}) \in \{-1, 1\}^T$ is $P(d_n|x_n)$, or in other words, conditional probability on given agent's n attributes and choice characteristics.

Lets also notice, that we have taste variation, which implies the following:

$$P(d_n|x_n) = \int_{\mathbb{R}^k} P(d_n|\beta_n, x_n)\phi_\Sigma(\beta_n)d\beta_n$$

The next step is to remember that $P(d_n|\beta_n, x_n) = P(d_n|V_n)$:

$$P(d_n|\beta_n, x_n) = \int_{\mathbb{R}} P(d_n|\eta_n, \beta_n, x_n)\phi_\sigma(\eta_n)d\eta_n$$

Having everything put together, we obtain:

$$\begin{aligned} P(d_n|x_n) &= \int_{\mathbb{R}^k} \left(\int_{\mathbb{R}} P(d_n|\eta_n, \beta_n, x_n)\phi_\sigma(\eta_n)d\eta_n \right) \phi_\Sigma(\beta_n)d\beta_n \\ &= \int_{\mathbb{R}^k} \left(\int_{\mathbb{R}} \prod_{t=1}^{T} \Phi((\beta_n' x_{nt} + \eta_n)d_{nt})\phi_\sigma(\eta_n)d\eta_n \right) \phi_\Sigma(\beta_n)d\beta_n \end{aligned}$$

where $\phi_\Sigma(\beta_n)$ is joint normal pdf with expectation $\boldsymbol{b}$ covariance matrix $\boldsymbol{\Sigma}$ and $\phi_\sigma(\eta_n)$ is normal pdf with zero expectation and variance σ