Latent variable models

Latent variable models

Outline

In latent variable models, we specify an economic model at the level of an unobserved (latent) variable. This model implies a model for the observable variables. We then estimate the parameters in the economic model using the observable variables, and the implied model.

Components

For any latent variable model, we discuss:

1. Model for latent variables
2. Model for observables ("choice probabilities")
3. Identification
4. Estimation
5. Interpretation (through marginal effects)

Ordered choice

Model

whiteboard

Model probabilities

whiteboard

Likelihood

whiteboard

Marginal effects

whiteboard

Example

From Kleiber and Zeileis, *Applied Econometrics with R*, p.149

```
library(AER)
```

```
## Loading required package: car

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival
```

```
library(MASS)
```

Results

```
model <- polr(job ~ education + minority + gender, data=Bar
summary(model)
```

```
##
## Re-fitting to get Hessian

## Call:
## polr(formula = job ~ education + minority + gender, data
##
## Coefficients:
##                 Value Std. Error t value
## education      0.8861    0.08112  10.924
## minorityyes   -1.5249    0.39434  -3.867
## genderfemale   0.5111    0.28825   1.773
##
## Intercepts:
##                  Value   Std. Error t value
## custodial|admin  6.6052  0.8412     7.8517
```

Multinomial logit

Latent model (1)

There are J alternatives ("products"), labelled $j = 1, \cdots, J$.
Alternative j yields utility

$$y_{ij}^* = X_{ij}\beta + u_{ij}$$

where - X_{ij} are observable variables specific to the individual / alternative. - u_{ij} is individual i's unobservable preference for good j

Latent model (2)

The unobservables are assumed to be iid across alternatives, with a Type 1 Extreme Value distribution:

$$(u_{i1}, \cdots, u_{iJ})|X_i \ i.i.d.F$$

with

$$F(a) = exp(-exp(-a)).$$

Latent model (3)

Individual i chooses the alternative that maximizes their utility:

$$y_i = argmax_{j \in \{1, \cdots, J\}} y_{ij}^*$$

Observable model

Under the model above,

$$P(y_i = j|X_i) = \frac{exp(X_{ij}\beta)}{\sum_{h=1}^{J} exp(X_{ih}\beta)}$$

Wooldridge 16.2.2 gives - the McFadden (1974) reference, and - a discussion of the **independence of irrelevant alternatives** implication

For another derivation, see the notes by James Heckman in the required readings

Identification

The regression coefficient in this model is identified after the normalization $X_{ij} = X_{ij} - X_{i1}$.

- Compare to the binary choice model, where $y^*_{i0} = 0$. Here, $y^*_{i1} = u_{i1}$

Estimation

Under - the latent model above, - the normalization on X_{i1}, and - the functional independence assumption for conditional likelihood you can estimate β using maximum likelihood, with

$$\mathcal{L}_n(\beta) = \sum_i \sum_j 1\{y_i = j\} log\left(P(y_i = j | X_i)\right)$$

Marginal effects (1)

The effect of changing X_{ijk}, the $k-$th regressor for individual i and product j, on the conditional probability of choosing alternative j:

$$
\begin{aligned}
\frac{\partial P(y_i = j|X_i)}{\partial X_{ijk}} &= \frac{(\sum_h exp(X_{ih}\beta))\beta_k exp(X_{ij}\beta) - exp(X_{ij}\beta)\beta_k exp(X_{ij}\beta)}{(\sum_h exp(X_{ih}\beta))^2} \\
&= \frac{(\sum_h exp(X_{ih}\beta))\beta_k exp(X_{ij}\beta)}{(\sum_h exp(X_{ih}\beta))^2} - \frac{exp(X_{ij}\beta)\beta_k exp(X_{ij}\beta)}{(\sum_h exp(X_{ih}\beta))^2} \\
&= \beta_k \frac{exp(X_{ij}\beta)}{\sum_h exp(X_{ih}\beta)} - \beta_k \left(\frac{exp(X_{ij}\beta)}{\sum_h exp(X_{ih}\beta)}\right)^2 \\
&= \beta_k P(y_i = j|X_i)\,(1 - P(y_i = j|X_i))
\end{aligned}
$$

Same sign as β_k.

Example

Using the mlogit vignette, with a very nice example.

Censoring and truncation

Why?

- We only observe wages for individiduals who choose to work
  - selection / truncation
- Variables such as taxes paid cannot fall below zero
  - censoring

Model (Probit)

A linear model for the latent variable:

$$Y_i^* = X_i\beta + u_i$$
$$u_i|X_i \sim \mathcal{N}(0, \sigma^2)$$

1. Censored:

$$Y_i = \begin{cases} 0 & \text{if } Y_i^* \leq 0 \\ Y_i^* & \text{if } Y_i^* > 0 \end{cases}$$

2. Truncated:

$$Y_i = \begin{cases} \text{not observed} & \text{if } Y_i^* \leq 0 \\ Y_i^* & \text{if } Y_i^* > 0 \end{cases}$$

Failure of OLS

What can go wrong when you apply OLS to censored data?

whiteboard

Likelihood (censored)

The likelihood contribution for the censored model is

$$L_i(\beta, \sigma) = \left(\frac{1}{\sigma} \phi \left(\frac{Y_i - X_i\beta}{\sigma} \right) \right)^{1\{Y_i > 0\}} \left(\Phi \left(- \frac{X_i\beta}{\sigma} \right) \right)^{1\{Y_i = 0\}}$$

Likelihood (truncated)

For the truncated model:

$$
\begin{aligned}
L_i = P(Y_i|X_i) &= P(Y_i = Y_i^* | Y_i^* > 0, X_i) \\
&= \frac{f_{Y_i^*}(Y_i^*|X_i)}{P(Y_i^* > 0|X)} \\
&= \frac{\frac{1}{\sigma}\phi\left(\frac{Y_i^* - X_i\beta}{\sigma}\right)}{\Phi\left(\frac{-X_i\beta}{\sigma}\right)}
\end{aligned}
$$

Estimation

MLE based on the likelihood contributions derived above.

Conditional means

For partial effects, we can look at three different conditional means.

1. Latent variable.
$$E[Y_i^*|X_i] = X_i\beta$$
2. Truncated mean.
$$E[Y_i|X_i, Y_i > 0] = X_i\beta + E[u_i|Y_i > 0, X_i]$$
3. Censored mean.
$$E[Y_i|X_i] = P(Y_i > 0|X_i)E(Y_i|X_i, Y_i > 0) + 0$$

Inverse Mills ratio

The quantity

$$E[u_i|Y_i > 0, X_i] = E[u_i|u_i > -X_i\beta, X_i]$$

plays an important role.

For $u \sim \mathcal{N}(0,1)$,

$$\begin{aligned} E[u|u > c] &= \int_c^\infty uf(u|u > c)du \\ &= \int_c^\infty u\frac{\phi(u)}{1-\Phi(c)}du \\ &= \frac{1}{1-\Phi(c)}\int_c^\infty u\phi(u)du \end{aligned}$$

Mills (2)

Using normality:

$$\frac{\partial \phi(u)}{\partial u} = \frac{\partial (2\pi)^{-1/2} exp(-u^2/2)}{\partial u}$$
$$= -u\phi(u)$$

Mills (3)

With that expression, the integral simplifies to

$$
\begin{aligned}
E[u|u > c] &= \frac{1}{1-\Phi(c)} \int_c^\infty u\phi(u)du \\
&= \frac{1}{1-\Phi(c)} [-\phi(\infty) + \phi(c)] \\
&= \frac{\phi(c)}{1-\Phi(c)}
\end{aligned}
$$

and, using symmetry,

$$
E[u|u > -c] = \frac{\phi(c)}{\Phi(c)} \equiv \lambda(c),
$$

which is referred to as the *inverse mills ratio*

Truncated expectation

The truncated conditional mean can then be expressed as:

$$\begin{aligned} E[Y_i|X_i, Y_i > 0] &= X_i\beta + E[u_i|Y_i > 0, X_i] \\ &= X_i\beta + \lambda\left(-\frac{X_i\beta}{\sigma}\right) \end{aligned}$$

From here, you can work out the marginal effects, which gives you an interpretation for β.

Structural model 1: The Roy Model

Roy model: setup

Chris Taber's slides (#1-20, suggested readings) do a great job of setting up the Roy model and walking you through its implications.

Roy model: parameters of interest

We are interested in

- the joint distribution of skills (F, H)
- prices

Issue: we only observe

- W_F for those who chose to fish
- W_H for those who chose to hunt

and the direction of the selection bias depends on the distribution of skills!

Labor version

In the labor market version,

- “fishing”: “market production” and
- “hunting”: becomes “non-market production”

We observe:

- who fishes and who hunts
- wages for those who fish
- regressors

Roy model (labor)

Under Normality:

$$Y_{fi} = X_{0i}\gamma_{0f} + X_{fi}\gamma_{ff} + \epsilon_{fi}$$
$$Y_{hi} = X_{0i}\gamma_{0h} + X_{hi}\gamma_{hh} + \epsilon_{hi}$$
$$\begin{pmatrix} \epsilon_{fi} \\ \epsilon_{hi} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_f^2 & \sigma_{fh} \\ \sigma_{fh} & \sigma_h^2 \end{pmatrix}\right)$$

Who fishes?

- Collect all regressors in X_i
- Denote $\Delta\epsilon_i = \epsilon_{hi} - \epsilon_{fi}$ with conditional standard deviation

$$\sigma^* = \sqrt{\sigma_f^2 + \sigma_h^2 - 2\sigma_{fh}}$$

An individual chooses to fish if $Y_{fi} > Y_{hi}$. Model probability:

$$\begin{aligned} P(Y_{fi} - Y_{hi} > 0|X_i) &= P\left(\Delta\epsilon_i < X_{0i}(\gamma_{0f} - \gamma_{0h}) + X_{fi}\gamma_{ff} - X_{hi}\gamma_{hh}|X_i\right) \\ &= \Phi(X_i\gamma^*) \end{aligned}$$

where

$$\gamma^* = \left(\frac{\gamma_{0f} - \gamma_{0h}}{\sigma^*}, \frac{\gamma_{ff}}{\sigma^*}, \frac{-\gamma_{hh}}{\sigma^*}\right)$$

Can estimate γ^* using probit.

Expected wage

We only observe wages for those who chose to work: selection.

$$\begin{aligned} E(Y_{fi}|X_i, Y_{fi} > Y_{hi}) &= E(X_{0i}\gamma_{0f} + X_{fi}\gamma_{ff} + \epsilon_{fi}|X_i) \\ &= X_{0i}\gamma_{0f} + X_{fi}\gamma_{ff} + E(\epsilon_{fi}|X_i, Y_{fi} > Y_{hi}) \end{aligned}$$

Last term requires reviewing properties of joint normal.

Joint normality and linearity

For any joint normal

$$\begin{pmatrix} \epsilon_{fi} \\ \Delta\epsilon_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_f^2 & \sigma_{f,\Delta} \\ \sigma_{f,\Delta} & (\sigma^*)^2 \end{pmatrix} \right)$$

with

$$\sigma_{f,\Delta} = \sigma_f^2 - \sigma_{fh}$$

There exist $\alpha_0, \alpha_1, \xi_i$ such that

$$\epsilon_{fi} = \alpha_0 + \alpha_1 \Delta\epsilon_i + \xi_i$$

such that ξ_i is zero-mean normal, independent of $\Delta\epsilon_i$. Furthermore,

$$\alpha_0 = E(\epsilon_{fi}) - E(\alpha_1 \Delta\epsilon_i + \xi_i) = 0$$

and

$$\begin{aligned} \alpha_1 &= Cov(\epsilon_{fi}, \Delta\epsilon_i)/Var(\Delta\epsilon_i) \\ &= \frac{\sigma_f^2 - \sigma_{fh}}{(\sigma^*)^2} \end{aligned}$$

Mills (4)

Define

$$c = X_{0i}(\gamma_f - \gamma_h) + X_{fi}\gamma_{ff} - X_{hi}\gamma_{hh}$$

so that we can use the Mills ratio:

$$\begin{aligned} E(\epsilon_{fi}|X_i, Y_{fi} > Y_{hi}) &= E(\epsilon_{fi}|\Delta\epsilon_i > -c, X_i) \\ &= \alpha_1 E(\Delta\epsilon_i|\Delta\epsilon_i > -c, X_i) \\ &= \frac{\alpha_1}{\sigma^*}\lambda(X'\gamma^*) \end{aligned}$$

Truncated mean

Replace this in the truncated expectation of market wages:

$$\begin{aligned} E(Y_{fi}|X_i, Y_{fi} > Y_{hi}) &= X_{0i}\gamma_{0f} + X_{fi}\gamma_{ff} + E(\epsilon_{fi}|X_i, Y_{fi} > Y_{hi}) \\ &= X_{0i}\gamma_{0f} + X_{fi}\gamma_{ff} + \frac{\alpha_1}{\sigma^*}\lambda(X'\gamma^*) \end{aligned}$$

1. γ^* from the probit
2. Regression of Y_{fi} on X_i and $\lambda(X_i\gamma^*)$ gives us
   - γ_{0f}
   - γ_{0h}
   - $\frac{\alpha_1}{\sigma^*}$
3. Together, γ^* and γ give us σ^*, and $\sigma_f^2 - \sigma_{fh}$
   - French and Taber: re-run probit; "structural probit"

Talent distributions

1. We obtained
$$(\sigma^*)^2 = \sigma_f^2 + \sigma_h^2 - 2\sigma_{fh}$$
2. We obtained $\alpha_1 = \sigma_f^2 - \sigma_{fh}$

See French and Taber (2010) / Heckman and Honore (1990) who show that σ_f^2 can be estimated from the residuals of Step 2, which pins down the entire distribution of skills.

Why not ML?

We could have written down the likelihood contribution in one go, and used MLE. Why not?

- Cannot write down the likelihood for models without normality assumption
- Multi-step estimation above generalizes to semiparametric version of the model

Details: See French and Taber (2010)

Application

Use Mroz data and the example described in the sampleSelection vignette

```
library(stargazer)
```

```
##
## Please cite as:
##  Hlavac, Marek (2015). stargazer: Well-Formatted Regress
##  R package version 5.2. http://CRAN.R-project.org/packag
```

```
library(sampleSelection)
```

```
## Loading required package: maxLik
## Loading required package: miscTools
##
## Please cite the 'maxLik' package as:
## Henningsen, Arne and Toomet, Ott (2011). maxLik: A packa
```

Application (2)

Compute several estimators

```
# Two-step procedure discussed in slides.
twostep <- selection( lfp ~ age + I( age^2 ) + faminc + kic
              wage ~ exper + I( exper^2 ) + educ + city,
              data = Mroz87, method = "2step" )
# Maximum likelihood
mle <- selection( lfp ~ age + I( age^2 ) + faminc + kids +
                          wage ~ exper + I( exper^2 ) + educ +
                          data = Mroz87, method = "ml" )
# Regression ignoring censoring
ignore_censoring <- lm(wage ~ exper + I(exper^2) + educ + c
# Regression on >0 wages, ignoring selection
ignore_selection <- lm(wage ~ exper + I(exper^2) + educ + c
                           subset=wage>0)
```

Application (3)

Table 1:

	Dependent variable:			
	wage			
	selection		*OLS*	
	(1)	(2)	(3)	(4)
exper	0.021	0.028	0.188***	0.032
	(0.062)	(0.062)	(0.039)	(0.062)
I(exper^2)	0.0001	−0.0001	−0.003***	−0.0003
	(0.002)	(0.002)	(0.001)	(0.002)
educ	0.417***	0.457***	0.415***	0.481***
	(0.100)	(0.073)	(0.049)	(0.067)
city	0.444	0.447	0.073	0.449
	(0.316)	(0.316)	(0.229)	(0.318)
Constant	−0.971	−1.963	−4.183***	−2.561***
	(2.059)	(1.198)	(0.614)	(0.929)

Application (4)

- Columns 1 and 2 give similar results. MLE (column 2) produces smaller standard errors: MLE is asymptotically efficient, the two-stage procedure is not
- Column 3 ignores censoring, and gives qualitatively different results
- Column 4 ?

Structural modelling

In the structural model, we identify all parameters of interest for counterfactual analyses:

- What happens to wages when we change the skill distribution?
- What happens when we change the prices?
- What is the loss of exogenously assigning people to jobs

Structural model 2: Marriage market

Model

Choo and Siow

Non-logit version

Galichon and Salanie

Structural model 3: Network effects

The reflection problem

- Manski

Binary choice

[brock and durlauf]