Estimation of Binary Response Models

(GB: Chapter 15.5-15.7)

Isabel Casas

SYDDANSK UNIVERSITET

- Reporting results
  - Percent of correctly predicted probabilities
  - Pseudo- R^2
  - Wald, LM, LR tests
- Model misspecifications
  - Omitted variables
  - Misspecified distribution function
  - Heteroskedasticity
  - Endogeneity

Goodness of fit

- Inference: Wald test, LR test, LM test
- Goodness of fit
  - Percent of observations "correctly" predicted (problem with the decision rule)
  - The ML estimates are not chosen as the ones that fit the sample best
  - A method less fitted might describe (partial effects) the problem better
  - Pseudo R^2:

$$R^2 = 1 - \frac{\ell_{ur}}{\ell_0}$$

Computer work at home

File: Marginal effects.R

- Find APE of variable $nwifeinc$
- Find the effect in an average person of going from $kidlt6 = 0$ to $kidlt6 = 1$
- How about from $kidlt6 = 1$ to $kidlt6 = 2$?
- How would you do if for $kidlt6 = 0$ to $kidlt6 = 2$?

Marginal effects exercise

```
The APE of nwifeinc with the probit model:  -0.003616175
The APE of nwifeinc with the logit model:  -0.003811813
The effect of nwifeinc with the LPM:  -0.003405169

The marginal effect of the average person with the probit model:  -0.00418526
The marginal effect of the average person with the logit model:  -0.004457575

The effect of one 1 extra kid over the probability of a woman working is:
 kidsl6   LPM     Probit    Logit
 ------------------------------------------------------------
 0-1    -0.262   -0.335   -0.344
 1-2    -0.262   -0.252   -0.242
```

Computer work at home

File: goodnessoffit.R

- Find the pseudo-R^2 value
- How many $y = 1$ are well predicted by the model?
- and $y = 0$?

```
The pseudo-R2 is :  0.2196814
The percentage of 1s predicted right  0.8130841
The percentage of 0s predicted right  0.6307692
```

Computer work at home

File: LM_goodnessoffit.R

- Using LM, test the restriction $H_0 : \beta_{kidslt6} = \beta_{kidsge6} = 0$
  - Get standarised residuals of the restricted model
  - Regress the standarised residuals on standarised X
  - Test $nR^2 \sim \chi^2_2$?

```
The R2 is  0.1005117 so the LM statistics is:  75.68535
The number of restrictions is 2, the p-value  3.673971e-17
```

What is the conclusion of the test?

Model Misspecification

- Reminder of identification problem
- Endogeneity
  - Neglected heterogeneity
  - Continuous variables
- Misspecified distribution of the latent model error term
- Heteroskedasticity in the latent error term

Parameter Identification Problem

Latent variable model

$$\mathbf{y} = G(\mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\epsilon}$$

where:

- $E(\boldsymbol{\epsilon}) = 0$
- $\mathbf{X}$ is exogenous, independent of $\boldsymbol{\epsilon}$
- The c.d.f G is from an exponential family and symmetric around zero $\Rightarrow G(\mathbf{X}\boldsymbol{\beta}) = 1 - G(-\mathbf{X}\boldsymbol{\beta})$.

Then:

$$\begin{aligned} P(\mathbf{y} = 1 \,|\mathbf{X}) =& P(\mathbf{y}^* > 0 \,|\mathbf{X}) = P(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} > 0 \,|\mathbf{X}) \\ =& P(\boldsymbol{\epsilon} > -\mathbf{X}\boldsymbol{\beta} \,|X) = P(\boldsymbol{\epsilon} < \mathbf{X}\boldsymbol{\beta} \,|\mathbf{X}) \\ =& G(\mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

Latent variable model

- If ϵ follows either a standard logistic distribution or a standard normal distributions, then the logit and probit estimates of β are consistent.
- However, if the variance of ϵ is not 1, then we are estimating β/σ instead.

Latent variable model

Subcase 1: If $\boldsymbol{\epsilon} \sim N\left(0, \sigma^2\right)$ where $\sigma \neq 1 \Rightarrow \boldsymbol{\epsilon}/\sigma$ is standard normal, and:

$$\begin{aligned} P\left(\mathbf{y}=1\left|\mathbf{X}\right.\right) &= P\left(\boldsymbol{\epsilon}<\mathbf{X}\boldsymbol{\beta}\left|\mathbf{X}\right.\right) \\ &= P\left(\frac{\boldsymbol{\epsilon}}{\sigma}<\mathbf{X}\frac{\beta}{\sigma}\left|X\right.\right) \\ &= \Phi\left(\mathbf{X}\frac{\boldsymbol{\beta}}{\sigma}\right)=\Phi\left(\mathbf{X}\tilde{\boldsymbol{\beta}}\right) \end{aligned}$$

Again, we have the probit, but this time with parameter $\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}/\sigma$. Hence, in this case, only $\boldsymbol{\beta}/\sigma$ is identified – not $\boldsymbol{\beta}$

Latent variable model

Subcase 2: If $\boldsymbol{\epsilon} \sim \Lambda\left(\mu = 0, s = 1\right)$ where $Var(\boldsymbol{\epsilon}) = \frac{\pi^2 s^2}{3}$:

$$\begin{aligned} P\left(\mathbf{y} = 1 \mid X\right) &= P\left(\boldsymbol{\epsilon} < \mathbf{X}\boldsymbol{\beta} \,|\mathbf{X}\right) \\ &= P\left(\frac{\sqrt{3}\boldsymbol{\epsilon}}{\pi s} < \mathbf{X}\frac{\sqrt{3}\beta}{\pi s} \,|\mathbf{X}\right) \\ &= \Lambda\left(\mathbf{X}\tilde{\boldsymbol{\beta}}; 0, 1\right) \end{aligned}$$

Again, we have the logit, but this time with parameter $\tilde{\beta} = \sqrt{3}\beta/(\pi s)$.
Hence, in this case, only $\tilde{\boldsymbol{\beta}}$ is identified – not $\boldsymbol{\beta}$

Latent variable model

- Parameters β are not identified so their estimation is inconsistent.
- However, we are not so much interested in β as in the partial effects of certain variable.
- The APE is consistently estimated.

Endogeneity

Omitted variables

Latent model:

$$y^* = \mathbf{X}\beta + \gamma c + \epsilon \quad \epsilon|\mathbf{X}, c \sim N(0,1)$$

where c is an unobserved variable with mean zero and variance τ^2.

The probit model: $P(y = 1|\mathbf{X}, c) = \Phi(\mathbf{X}\beta + \gamma c)$

Case 1: c and $\mathbf{X}$ are independent $\Rightarrow$ neglected heterogeneity

Case 2: c and $\mathbf{X}$ are dependent $\Rightarrow$ endogeneity

Case 2: $c, \mathbf{X}$ dependent

- If c is correlated with $\mathbf{X}$ or dependent in any other way: omission of c is a serious problem ($\Rightarrow$ Endogeneity).
- We can estimate the LPM by 2SLS finding instruments of c.
- We can use the 2-step approach by Rivers and Vuong's, also known as control function approach.

Simultaneous variables

We have a simultaneous model (Tobit model):

$$
\begin{aligned}
y_1^* &= \delta_1 \; z_1 + \alpha_1 \; y_2 + \boldsymbol{\epsilon}_1 \text{ unobserved} \\
y_2 &= \delta_{21} \; z_1 + \delta_{22} \; z_2 + \boldsymbol{\epsilon}_2 = \delta_2 \; z + \boldsymbol{\epsilon}_2 \\
y_1 &= 1[y_1^* > 0]
\end{aligned}
$$

- $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2)$ is a bivariate normal, with mean zero and independent of z
- If $\boldsymbol{\epsilon}_2$ is correlated with $\boldsymbol{\epsilon}_1 \Rightarrow y_2$ endogeneous (Problem!!!)
- If $\boldsymbol{\epsilon}_2$ is independent of $\boldsymbol{\epsilon}_1 \Rightarrow y_2$ exogenous (No problem)
- Assume $y_2|z$ is normally distributed
- y_2 is a continuous r.v

Tobit model

Let us first set the model.

- $Var(\epsilon_1) = 1$ and $Var(\epsilon_2) = \tau_2^2$
- We write $\epsilon_1 = \theta\epsilon_2 + \eta_1 \quad \Rightarrow \quad Cov(\epsilon_1, \epsilon_2) = \theta\tau_2^2 \quad \rho_1 = \theta\tau_2$
- η_1 is independent of z and ϵ_2 and it is normal
- $E(\eta_1) = 0$,
  $Var(\eta_1) = Var(\epsilon_1) + \theta^2 Var(\epsilon_2) - 2\theta Cov(\epsilon_1, \epsilon_2) = 1 - \rho_1^2$
- $\eta_1 \sim N(0, 1 - \rho_1^2)$
- $y_1^* = \delta_1 z_1 + \alpha_1 y_2 + \theta\epsilon_2 + \eta_1$ (unobserved)
- So, $y_1 = 1[y_1^* > 0]$
- Calculate $P(y_1 = 1|z, y_2, \epsilon_2)$ (3 minutes)

Tobit model

$$P(y = 1|z, y_2, \boldsymbol{\epsilon}_2) = \Phi\left(\frac{\delta_1 z_1 + \alpha_1 y_2 + \theta \boldsymbol{\epsilon}_2}{\sqrt{1 - \rho_1^2}}\right)$$

$\delta_1, \alpha_1, \theta$ are not identify. Our estimates:

- $\tilde{\delta}_1 = \delta_1 / \sqrt{1 - \rho_1^2}$
- $\tilde{\alpha}_1 = \alpha / \sqrt{1 - \rho_1^2}$
- $\tilde{\theta} = \theta / \sqrt{1 - \rho_1^2}$
- As $0 < 1 - \rho_1^2 < 1 \Rightarrow 1/\sqrt{1 - \rho_1^2} > 1$ so $\tilde{\delta}_1 > \delta_1$
- We haven't got an estimate of δ_2 in this step. So we must estimate it with OLS.

$$y_2 = \delta_2 z + \boldsymbol{\epsilon}_2$$

Tobit model

- Probit estimates of a simultaneous system with endogeneity are not identified.
- We have to first test for endogeneity
- If endogeneity exists, then we use Rivers and Vuong 2 steps estimator

Test of Endogeneity: Rivers and Vuong

$$y_1^* = \delta_1 \; z_1 + \alpha_1 \; y_2 + \epsilon_1 \text{ unobserved}$$
$$y_2 = \delta_{21} \; z_1 + \delta_{22} \; z_2 + \epsilon_2 = \delta_2 \; z + \epsilon_2$$
$$y_1 = 1[y_1^* > 0]$$

$$H_0 : y_2 \textit{ exogenous}$$

Step 1 : Estimate δ_2 on z with OLS and obtain $\hat{\epsilon}_2$

Step 2 : Run de probit $y_1 = \delta_1 z_1 + \alpha_1 \; y_2 + \theta \hat{\epsilon}_2 + \eta_1$

- The t-test on $\hat{\theta}$ is equivalent to $H_0 : \theta = 0$.
- This test is valid even for heteroskedastic ϵ_2 and binary y_2

Look at Example 15.3 page 587 Wooldridge.

Estimates of Rivers and Vuong

If $H_0 : y_2$ *exogenous* is rejected, i.e $\theta = 0$, the two step procedure provides consistent estimates of δ_1 and α_1

- From the first step we have consistent estimators of δ_2, τ_2^2
- From the second step we have consistent estimators of $\tilde{\delta}_1, \tilde{\alpha}_1$ and $\tilde{\theta}$
- It can be shown:

$$\hat{\delta}_1 = \frac{\tilde{\delta}_1}{\sqrt{(1 + \tilde{\theta}^2 \hat{\tau}_2^2)}}$$

$$\hat{\alpha}_1 = \frac{\tilde{\alpha}_1}{\sqrt{(1 + \tilde{\theta}^2 \hat{\tau}_2^2)}}$$

Estimates with endogeneity

The endogenous variable is binary:

$$y_1 = \mathbf{1}[\delta_1 \; z_1 + \alpha_1 \; y_2 + \boldsymbol{\epsilon}_1 > 0]$$
$$y_2 = \mathbf{1}[\delta_2 \; z + \boldsymbol{\epsilon}_2 > 0]$$

- $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2)$ is a bivariate normal, with mean zero and independent of z
- If $\boldsymbol{\epsilon}_2$ is correlated with $\boldsymbol{\epsilon}_1 \Rightarrow$ inconsistent estimator δ_1, α_1
- If $\boldsymbol{\epsilon}_2$ is independent of $\boldsymbol{\epsilon}_1 \Rightarrow y_2$ exogenous (No problem)
- Assume $y_2|z$ is a binary random variable

Estimates with endogeneity

- We need the joint distribution of (y_1, y_2) given z to obtain the log-likelihood function and find the parameters that maximise it.
- The Rivers and Vuong procedure can be used to test for endogeneity but not to estimate the parameters
- See Wooldridge Chapter 15.7.3 or Rivers and Vuong (1988) for details

Computer work

File: Vuong.R

- Test for exogeneity of $educ$. Example 15.3 of Wooldridge.

$$\begin{aligned}
\mathbf{y}_1^* =& \beta_0 + \beta_1\ nwifeinc + \beta_2\ educ + \beta_3\ exper + \beta_4\ expersq \\
&+ \beta_5\ age + \beta_t\ kidslt6 + \boldsymbol{\epsilon}_1 \\
educ =& \delta_0 + \delta_1\ nwifeinc + \delta_2\ exper + \delta_3\ expersq \\
&+ \delta_4\ age + \delta_5\ kidslt6 + \delta_6\ kidsge6 + \delta_7\ motheduc + \delta_8\ fatheduc \\
&+ \delta_9 houseduc + \boldsymbol{\epsilon}_2
\end{aligned}$$

Misspecified distribution

Misspecified error distribution

Probit:

- We assume that $\boldsymbol{\epsilon} \sim N(0,1)$
- Instead $\boldsymbol{\epsilon} \sim N(0,\sigma^2)$ then we cannot estimate β consistently but we can estimate the APE consistently
- What if $\boldsymbol{\epsilon}$ does not follow a normal distribution? Then:

$$P(\mathbf{y}=1|X) = P(\mathbf{y}^* > 0|X) = P(\boldsymbol{\epsilon} < \mathbf{X}\boldsymbol{\beta}|X) = F(\mathbf{X}\boldsymbol{\beta}) \neq \Phi(\mathbf{X}\boldsymbol{\beta})$$

- Then the log-lik function of the probit is wrong because we are using Φ instead of F
- The estimates are inconsistent

Misspecified error distribution

Logit:

- We assume that $\boldsymbol{\epsilon} \sim \Lambda(0, 1)$
- Instead $\boldsymbol{\epsilon} \sim \Lambda(0, s)$ then we cannot estimate β consistently but we can estimate the APE consistently
- What if $\boldsymbol{\epsilon}$ does not follow a logistic distribution? Then:

$$P(\mathbf{y} = 1|X) = P(\mathbf{y}^* > 0|X) = P(\boldsymbol{\epsilon} < \mathbf{X}\boldsymbol{\beta}|X) = F(\mathbf{X}\boldsymbol{\beta}) \neq \Lambda(\mathbf{X}\boldsymbol{\beta})$$

- Then the log-lik function of the logit is wrong because we are using Λ instead of F
- The estimates are inconsistent

Misspecified error distribution

- Even if the estimates are inconsistent, the estimates of the partial effects might be very good.
- For example: $\boldsymbol{\epsilon} \sim N(0,1)$ but we use the log-lik function of a logit model.
- The estimates $\hat{\boldsymbol{\beta}}^{logit} \approx 1.6\hat{\boldsymbol{\beta}}^{probit}$
- However the partial effects of both models are comparable.

Heteroskedasticity

- We have assume that ϵ_j are identically distributed (G)
- Assume that ϵ_j/σ_j are identically distributed (G)
- For $\sigma_j = \exp(Z_j\gamma)$ and $\mathbf{Z}$ some observed variables
- $$P(\mathbf{y} = 1|X) = G\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\exp(Z\gamma)}\right)$$
- Null hypothesis of homoskedasticity $H_0 : \gamma = 0$ (restricted model)
- The unrestricted model is $P(\mathbf{y} = 1|X) = G\left(\frac{\mathbf{X}\boldsymbol{\beta}}{\exp(Z\gamma)}\right)$
- The restricted $P(\mathbf{y} = 1|X) = G(\mathbf{X}\boldsymbol{\beta})$
- We can use LR test, the Wald test or the LM test.

LM: test for heteroskedasticity

Step 1 : Estimate restricted model, i.e the homoskedastic model by ML (probit or logit) and obtain the standarised residuals $\tilde{\mathbf{r}}$ and $\tilde{\boldsymbol{\beta}}$

Step 2 : Regress $\tilde{r}_j$ on

$$\frac{G'(\mathbf{x}_j\tilde{\boldsymbol{\beta}})}{\sqrt{\hat{G}(1-\tilde{G})}}\mathbf{x}_j \quad \text{and} \quad \frac{G'(\mathbf{x}_j\tilde{\boldsymbol{\beta}})X\tilde{\boldsymbol{\beta}}}{\sqrt{\tilde{G}(1-\tilde{G})}}Z_j$$

and obtain the R^2

Step 3 : $LM = nR^2 \sim \chi^2_q$ where q is the number of parameters in γ.

Compute work

File: LM_heteroskedasticity.R

The ϵ_j of the latent model might be heteroskedastic, $\epsilon_j \sim N(0, \sigma_j^2)$ where $\sigma_j^2 = \exp(\gamma \mathbf{x}_j)$. Basically $Z = X$. Test for heteroskedasticity in:

$$\begin{aligned}\mathbf{y}^* =& \beta_0 + \beta_1\ nwifeinc + \beta_2\ educ + \beta_3\ exper + \beta_4\ expersq \\ &+ \beta_5\ age + \beta_6\ kidslt6 + \beta_7\ kidsge6 + \boldsymbol{\epsilon}\end{aligned}$$

Summary

- Misspecification issues
  - Omitted variables which are independent of $\mathbf{X}$
  - Omitted variables which are dependent of $\mathbf{X}$
  - Misspecified error distribution
  - Heteroskedasticity of the error
- The last three all change the functional form $\Rightarrow$ inconsistent $\hat{\boldsymbol{\beta}}$

Testing against more general models (LM)

- Assume that the latent model has an error such that $\boldsymbol{\epsilon} \sim N(0, exp(2\mathbf{x}_1\delta))$
- Where $\mathbf{x}_1$ is a part of $\mathbf{X}$.
- Basically there is heteroskedasticity in the latent model
- $P(y = 1|X) = \Phi(\exp(\mathbf{x}_1\delta)\mathbf{X}\boldsymbol{\beta})$
- The partial effects of $\mathbf{x}_j$ depends on β and δ and are difficult to interpret
- If $\delta = 0$ then we have the probit model
- We want to test $H_0 : \ \delta = 0$

Testing against more general models (LM)

- LM is convenient to test the null hypothesis presented above
- We need to estimate only the restricted model, i.e. with $\delta = 0$
- We can consider a more general test $H_0 : \delta = \delta_0 = 0$
- Under the null hypothesis $G(\mathbf{X}\boldsymbol{\beta}) = m(\mathbf{X}\boldsymbol{\beta}, X, \delta_0)$
- In the example, $G(\mathbf{X}\boldsymbol{\beta}) = \Phi(\exp(\mathbf{x}_1\delta_0)\mathbf{X}\boldsymbol{\beta}) = \Phi(\mathbf{X}\boldsymbol{\beta})$

Testing against more general models (LM)

- Estimate the model without heteroskedasticity and get the residuals $\tilde{\eta}_j = y_j - G(\mathbf{x}_j\tilde{\beta})$
- Construct the standarised residual

$$\tilde{r}_j = \frac{\tilde{\eta}_j}{\sqrt{\tilde{G}_j(1-\tilde{G}_j)}} \quad \text{where } \tilde{G}_j = G(\mathbf{x}_j\tilde{\beta})$$

- Until now, it is all the same than for the LM test

Testing against more general models (LM)

- Find gradients (first partial derivatives) of $m()$ wrt β and δ
- Evaluate these at restricted estimates

$$\nabla_\beta \tilde{m}_j = G'(\mathbf{x}_j\tilde{\beta})\mathbf{x}_j$$
$$\nabla_\delta \tilde{m}_j = \nabla_\delta m(\mathbf{x}_j\tilde{\beta}, \mathbf{x}_j, \delta_0)$$

- Regress standarised residuals on:

$$\frac{G'(\mathbf{x}_j\tilde{\beta})\mathbf{x}_j}{\sqrt{\tilde{G}_j(1-\tilde{G}_j)}} \quad \text{and} \quad \frac{\nabla_\delta \tilde{m}_j}{\sqrt{\tilde{G}_j(1-\tilde{G}_j)}}$$

- $LM = n * R^2 \sim \chi^2_q$

Testing against more general models (LM)

In the example,

$$\nabla_\beta \tilde{m}_j = \phi(\mathbf{x}_j\tilde{\beta})\mathbf{x}_j$$
$$\nabla_\delta \tilde{m}_j = \phi(\exp(\mathbf{x}_1\delta)\mathbf{x}_j\tilde{\beta})\mathbf{x}_j\beta\mathbf{x}_1\exp(\mathbf{x}_1\delta)$$

Under the null hypothesis

$$\nabla_\delta \tilde{m}_j = \phi(\mathbf{x}_j\tilde{\beta})\mathbf{x}_j\tilde{\beta}\mathbf{x}_1$$