ECO475H1 S
Applied Econometrics II

Tianyu Du

February 22, 2019

Github Page https://github.com/TianyuDu/Spikey_UofT_Notes
Note Page TianyuDu.com/notes

Contents

1 Topic 1 Binary Outcome Model

1.1 Lecture 1. Jan. 10 2019

1.1.1 Model

Interpretation Binary outcome models can be interpreted as rational individuals are making binary decisions by comparing the utilities resulting from each decision.

$$\begin{cases} u_1(\mathbf{x}, \varepsilon_1) & \text{utility from decision 1} \\ u_0(\mathbf{x}, \varepsilon_0) & \text{utility from decision 0} \end{cases} \tag{1.1}$$

and rationality of individuals suggests

$$y = \begin{cases} 1 & \text{if } u_1(\mathbf{x}_i, \varepsilon_1) \geq u_0(\mathbf{x}_i, \varepsilon_0) \\ 0 & \text{otherwise} \end{cases} \tag{1.2}$$

Assumption 1.1. By convention, the edge case that $u_0 = u_1$ is modelled and classified to be in the first case, but $\mathbb{P}[u_1 - u_0 = 0|\mathbf{x}] = 0$ for continuous utilities, so the edge does not really matter.

Assumption 1.2. The difference in utilities can be captured by a linear function, and we define the **latent variable** y^* as

$$y^*(\mathbf{x}) \equiv u_1(\mathbf{x}, \varepsilon_1) - u_0(\mathbf{x}, \varepsilon_0) \tag{1.3}$$
$$= \mathbf{x}'\beta + \varepsilon \tag{1.4}$$

Remark 1.1 (About shapes of variables)**.**

$$\mathbf{x},\ \beta \in \mathbb{M}_{k\times 1};\ y,\ d \in \mathbb{R} \tag{1.5}$$

Reduced Model Then the primary model (1.2) is reduced to

$$y = \mathbb{1}\{y^*(\mathbf{x}) \geq 0\} \tag{1.6}$$

1.1.2 Data Structure

For each individual, the observation contains two *primary variables*, $(\mathbf{x}_i, y_i)$. So the dataset would look like

$$\{(\mathbf{x}_i, y_i)\}_{i=1}^N \tag{1.7}$$

1.1.3 Linear Probability Model

Problems

1. Out-of-range prediction;
2. Homogeneous marginal effect;
3. (Potentially) heteroskedasticity.

1.1.4 Model with Binomial Distribution

Procedures

1. Construct distribution and density functions.
2. Construct likelihood from density function, assuming iid of observations.
3. (MLE) estimates model parameters.
4. Predictions, both individual and average effect.

Conditional Density

$$f_Y(y|\mathbf{x}) = \begin{cases} P(\mathbf{x}) & \text{if } y = 1 \\ 1 - P(\mathbf{x}) & \text{if } y = 0 \end{cases} \tag{1.8}$$

$$= P(\mathbf{x})^d[1 - P(\mathbf{x})]^{1-d} \tag{1.9}$$

Assumption 1.3 (Distributions of ε)**.** Generally, there are two assumptions regarding to the distribution of ε,

(i) **Standard Normal** $\varepsilon \sim \mathcal{N}(0, 1)$[1].

(ii) **Gumbel Distribution** $F_\varepsilon(x) = \frac{e^x}{1+e^x}$.

Note that, in either case, ε is symmetrically distributed, which means ε and $-\varepsilon$ have the identical density and distribution.

$$P(\mathbf{x}) \equiv \mathbb{P}[y^* \geq 0|\mathbf{x}] = \mathbb{P}[-\varepsilon \leq \mathbf{x}'\beta|\mathbf{x}] \equiv F_\varepsilon(\mathbf{x}'\beta) \tag{1.10}$$

$$\implies f_Y(y|\mathbf{x}) = (F_\varepsilon(\mathbf{x}'\beta))^y(1 - F_\varepsilon(\mathbf{x}'\beta))^{1-y} \tag{1.11}$$

Likelihood assuming observations are iid, the joint density, conditioned on model parameter β, of the collection of observations $\{(y_i)\}_{i=1}^N$ is

$$f_{\{(Y_i)\}_{i=1}^N}(\{(y_i)\}_{i=1}^N|\beta, \{(\mathbf{x}_i)\}_{i=1}^N) = \prod_{i=1}^N f_Y(y_i|\mathbf{x}_i, \beta) \tag{1.12}$$

$$= \prod_{i=1}^N F_\varepsilon(\mathbf{x}_i'\beta)^{y_i}(1 - F_\varepsilon(\mathbf{x}_i'\beta))^{1-y_i} \tag{1.13}$$

and the likelihood of collection of observations $\{(\mathbf{x}_i, y_i, d_i)\}_{i=1}^N$ is given parameter β is

$$\mathcal{L}(\beta|\{(\mathbf{x}_i, y_i, d_i)\}_{i=1}^N) = \prod_{i=1}^N F_\varepsilon(\mathbf{x}_i'\beta)^{y_i}(1 - F_\varepsilon(\mathbf{x}_i'\beta))^{1-y_i} \tag{1.14}$$

$$\implies \ln \mathcal{L} = \sum_{i=1}^N y_i \ln F_\varepsilon(\mathbf{x}_i'\beta) + (1 - y_i)\ln(1 - F_\varepsilon(\mathbf{x}_i'\beta)) \tag{1.15}$$

[1] In the normal distribution case, we cannot estimate the variance of $\hat{\beta}_{MLE}$ laster if we do not assume the variance to be 1.

Maximum Likelihood Estimation First order conditions

$$\frac{\partial \ln \mathcal{L}}{\partial \beta} = \sum_{i=1}^{N} y_i \frac{F_\varepsilon'(\mathbf{x}_i'\beta)}{F_\varepsilon(\mathbf{x}_i'\beta)}\mathbf{x}_i' - (1-y_i)\frac{F_\varepsilon'(\mathbf{x}_i'\beta)}{1-F_\varepsilon(\mathbf{x}_i'\beta)}\mathbf{x}_i' \tag{1.16}$$

$$= \sum_{i=1}^{N} \frac{y_i F_\varepsilon'(\mathbf{x}_i'\beta) - F_\varepsilon'(\mathbf{x}_i'\beta)F_\varepsilon(\mathbf{x}_i'\beta)}{F_\varepsilon(\mathbf{x}_i'\beta)(1-F_\varepsilon(\mathbf{x}_i'\beta))}\mathbf{x}_i' = 0 \tag{1.17}$$

Prediction - Marginal Effect With binary outcome model, according to mean of Binomial distribution,

$$\mathbb{E}[y|\mathbf{x}] = P(\mathbf{x}) \equiv F_\varepsilon(\mathbf{x}'\hat{\beta}_{MLE}) \tag{1.18}$$

Then the marginal effect is

$$\frac{\partial \mathbb{E}[y|\mathbf{x}]}{\partial x_k} = F_\varepsilon'(\mathbf{x}'\hat{\beta}_{MLE})\hat{\beta}_{j,MLE} \tag{1.19}$$

In contrast to the linear probability model, in which the marginal effect is assumed to be homogeneous. In BOM prediction, the $F_\varepsilon'(\mathbf{x}'\hat{\beta}_{MLE})$ term in the marginal effect acts as a source of **heterogeneity** marginal effect.

Remark 1.2. Given $f_\varepsilon \geq 0$, $\hat{\beta}_{MLE}$ reports the **sign** of marginal effect, but it provides no quantitative implication.

Prediction - Average Marginal Effect There are two methods to estimate the average marginal effect, these two methods generate different estimations unless the density function of ε is linear.

(i) $\overline{ME}_j = \frac{1}{N}\sum_{i=1}^{N} F_\varepsilon'(\mathbf{x}_i\hat{\beta})\hat{\beta}_j$

(ii) $\overline{ME}_j = F_\varepsilon'(\overline{\mathbf{x}}_i\hat{\beta})\hat{\beta}_j$

2 Lecture 3. Jan. 24 2019

2.1 Two Side Censoring MLE

Consider the latent dependent variable

$$Y^* = \mathbf{x}'\boldsymbol{\beta} + \epsilon \tag{2.1}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$[2].
Therefore, given fixed $\mathbf{x}$,

$$Y^* \sim \mathcal{N}(\mathbf{x}'\boldsymbol{\beta}, \sigma^2) \tag{2.2}$$

Define parameter set

$$\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \sigma) \tag{2.3}$$

[2] In general, we can assume the error variance to be σ^2 when the dependent variable is *quantitive*, but with *qualitative* dependent variables, we assume $\varepsilon \sim \mathcal{N}(0,1)$ since we don't have sufficient information to estimate the error variance.

The observable variable is

$$Y = \begin{cases} U & \text{if } Y^* \geq U \\ Y^* & \text{if } Y^* \in (L, U) \\ L & \text{if } Y^* \leq L \end{cases} \tag{2.4}$$

Let $f_Y(y|\mathbf{x}, \boldsymbol{\beta}) : [L, U] \to [0, 1]$ be the probability measure of Y.
Let $y \in [L, U]$,

$$f_Y(y|\mathbf{x}, \boldsymbol{\beta}) = \begin{cases} \mathbb{P}(Y^* \geq U|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \geq U \\ f_{Y^*}(y|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \in (L, U) \\ \mathbb{P}(Y^* \leq L|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \leq L \end{cases} \tag{2.5}$$

$$= \begin{cases} 1 - F_{Y^*}(U|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \geq U \\ f_{Y^*}(y|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \in (L, U) \\ F_{Y^*}(L|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \leq L \end{cases} \tag{2.6}$$

Define indicator $(d_1(y), d_2(y), d_3(y))$ as

$$d_1(y) \equiv \mathcal{I}(y \geq U) \tag{2.7}$$
$$d_2(y) \equiv \mathcal{I}(y \in (L, U)) \tag{2.8}$$
$$d_3(y) \equiv \mathcal{I}(y \leq L) \tag{2.9}$$

Then the probability measure of Y can be expressed as

$$f_Y(y|\mathbf{x}, \boldsymbol{\beta}) = (1 - F_{Y^*}(U|\mathbf{x}, \boldsymbol{\beta}))^{d_1} \times f_{Y^*}(y|\mathbf{x}, \boldsymbol{\beta})^{d_2} \times F_{Y^*}(L|\mathbf{x}, \boldsymbol{\beta})^{d_3} \tag{2.10}$$

Suppose samples are i.i.d., the joint density is

$$f_{Y_1,\ldots,Y_N}(y_1, \ldots, y_N|\mathbf{X}, \boldsymbol{\beta}) = \prod_{i=1}^{N} f_Y(y_i|\mathbf{x}_i, \boldsymbol{\beta}) \tag{2.11}$$

The log-likelihood is

$$\mathcal{L}_N(\boldsymbol{\theta}|\mathbf{X}) = \sum_{i=1}^{N} \Big\{ d_{1,i} \times \ln(1 - F_{Y^*}(U|\mathbf{x}_i, \boldsymbol{\beta})) + d_{2,i} \times \ln(f_{Y^*}(y|\mathbf{x}_i, \boldsymbol{\beta})) + d_{3,i} \times \ln(F_{Y^*}(L|\mathbf{x}_i, \boldsymbol{\beta})) \Big\} \tag{2.12}$$

Finally, solving

$$\hat{\boldsymbol{\theta}}_{MLE} = (\hat{\boldsymbol{\beta}}_{MLE}, \hat{\sigma}_{MLE}) = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \, \mathcal{L}_N(\boldsymbol{\theta}) \tag{2.13}$$

2.2 Two Side Truncated MLE

Suppose the observations are truncated with lower and upper bounds L and U.
Let the latent dependent variable be

$$Y^* = \mathbf{x}'\boldsymbol{\beta} + \epsilon \tag{2.14}$$

and

$$\epsilon \sim \mathcal{N}(0, \sigma^2) \tag{2.15}$$

which implies, for given $\mathbf{x}$,

$$Y^* \sim \mathcal{N}(\mathbf{x}'\boldsymbol{\beta}, \sigma^2) \tag{2.16}$$

Define parameter set

$$\boldsymbol{\theta} \equiv \{\boldsymbol{\beta}, \sigma\} \tag{2.17}$$

Observable random variable Y is

$$Y = \begin{cases} Y^* & \text{if } Y^* \in (L, U) \\ -- & \text{if } Y^* \notin (L, U) \end{cases} \tag{2.18}$$

Constructing the distribution for Y, note that F_Y is only defined on $y \in (L, U)$,

$$F_Y(y|\mathbf{x}, \boldsymbol{\theta}) = \mathbb{P}(Y < y|\mathbf{x}, \boldsymbol{\theta}) \tag{2.19}$$

$$= \frac{\mathbb{P}(Y^* < y \wedge Y^* \in (L, U)|\mathbf{x}, \boldsymbol{\theta})}{\mathbb{P}(Y^* \in (L, U)|\mathbf{x}, \boldsymbol{\theta})} \tag{2.20}$$

$$= \frac{\mathbb{P}(Y^* \in (L, y)|\mathbf{x}, \boldsymbol{\theta})}{\mathbb{P}(Y^* \in (L, U)|\mathbf{x}, \boldsymbol{\theta})} \tag{2.21}$$

$$= \frac{F_{Y^*}(y|\mathbf{x}, \boldsymbol{\theta}) - F_{Y^*}(L|\mathbf{x}, \boldsymbol{\theta})}{F_{Y^*}(U|\mathbf{x}, \boldsymbol{\theta}) - F_{Y^*}(L|\mathbf{x}, \boldsymbol{\theta})} \tag{2.22}$$

Then construct the density of Y

$$f_Y(y|\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial F_Y(y|\mathbf{x}, \boldsymbol{\theta})}{\partial y} \tag{2.23}$$

$$= \frac{f_{Y^*}(y|\mathbf{x}, \boldsymbol{\theta})}{F_{Y^*}(U|\mathbf{x}, \boldsymbol{\theta}) - F_{Y^*}(L|\mathbf{x}, \boldsymbol{\theta})} \tag{2.24}$$

The sample log-likelihood is

$$\mathcal{L}_N(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln(f_{Y^*}(y_i|\mathbf{x}_i, \theta)) - \ln(F_{Y^*}(U|\mathbf{x}_i, \boldsymbol{\theta}) - F_{Y^*}(L|\mathbf{x}_i, \boldsymbol{\theta})) \tag{2.25}$$

and the estimator is given by

$$\hat{\boldsymbol{\theta}}_{MLE} = \{\hat{\boldsymbol{\beta}}_{MLE}, \hat{\sigma}_{MLE}\} = \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} \mathcal{L}_N(\boldsymbol{\theta}) \tag{2.26}$$

3 Lecture 4. Jan. 31 2019

3.1 Tobit and Sample Selection

Model the *observable* variables in Tobit model with sample selection are determined by both **outcome equation** and **selection equation**.

$$y_i = \begin{cases} \mathbf{x}_i'\beta + \epsilon_i & \text{if } \mathbf{w}_i'\gamma + v_i > 0 \\ \text{✗} & \text{otherwise} \end{cases} \tag{3.1}$$

where **unmeasurable errors** are assumed to follow joint normal distribution,

$$\begin{pmatrix} \epsilon_i \\ v_i \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \begin{pmatrix} \sigma_\epsilon^2 & \rho\sigma^2 \\ \rho\sigma^2 & 1 \end{pmatrix}) \tag{3.2}$$

Lemma 3.1. If (ϵ, v) follows joint normal distribution, then there exists $e \perp v$ and $e \sim \mathcal{N}(0, 1)$ such that

$$\frac{\epsilon}{\sigma_\epsilon} = \rho v + e \tag{3.3}$$

Expectation Define $\tilde{\mathbf{x}}_i \equiv [\mathbf{x}_i, \mathbf{w}_i]$, then the expected *observed* dependent variable is [3]

$$\mathbb{E}[y|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{3.4}$$
$$= \mathbb{E}[\mathbf{x}'\beta + \epsilon|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{3.5}$$
$$= \mathbf{x}'\beta + \mathbb{E}[\epsilon|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{3.6}$$
$$= \mathbf{x}'\beta + \mathbb{E}[\rho v \sigma_\epsilon + e\sigma_\epsilon|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{3.7}$$
$$= \mathbf{x}'\beta + \rho\sigma_\epsilon\mathbb{E}[v|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] + \sigma_\epsilon\mathbb{E}[e|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{3.8}$$
$$= \mathbf{x}'\beta + \rho\sigma_\epsilon\mathbb{E}[v|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{3.9}$$

Remark 3.1. If $\rho = 0$ in equation (2.9), there is no sample selection problem and we can use OLS to estimate the outcome equation.

Lemma 3.2. If $X \sim \mathcal{N}(\mu, \sigma^2)$ then

$$\mathbb{E}[X|X > \alpha] = \mu + \sigma\frac{\phi(\frac{x-\mu}{\sigma})}{1 - \Phi(\frac{x-\mu}{\sigma})} \tag{3.10}$$

(continue)

$$\cdots = \mathbf{x}'\beta + \rho\sigma_\epsilon\mathbb{E}[v|v > -\mathbf{w}'\gamma, \tilde{\mathbf{x}}] \tag{3.11}$$
$$= \mathbf{x}'\beta + \rho\sigma_\epsilon\frac{\phi(-\mathbf{w}'\gamma)}{1 - \Phi(-\mathbf{w}'\gamma)} \tag{3.12}$$
$$= \mathbf{x}'\beta + \rho\sigma_\epsilon\frac{\phi(\mathbf{w}'\gamma)}{\Phi(\mathbf{w}'\gamma)} \tag{3.13}$$
$$= \mathbf{x}'\beta + \rho\sigma_\epsilon\lambda(\mathbf{w}'\gamma) \tag{3.14}$$

where $\lambda(x)$ is the **inverse Mill's ratio** of standard normal at x.

Marginal Effect Consider the case

$$\exists x_k \in \mathbf{x} \cap \mathbf{w} \tag{3.15}$$

for instance, x_k can be *wage taxation.* The marginal effect of x_k is

$$\frac{\partial\mathbb{E}[y|\mathbf{w}'\gamma + v > 0, \tilde{\mathbf{x}}]}{\partial x_k} = \frac{\partial\mathbf{x}'\beta + \rho\sigma_\epsilon\lambda(\mathbf{w}'\gamma)}{\partial x_k} \tag{3.16}$$
$$= \beta_k + \rho\sigma_\epsilon\lambda'(\mathbf{w}'\gamma)\gamma_k \tag{3.17}$$
$$\tag{3.18}$$

where β_k measures the **direct effect** and $\lambda'(\mathbf{w}'\gamma)\gamma_k$ measures the **indirect effect** of x_k.

3.2 Heckman Estimation (Two-Step Procedure)

Step 1 Run a *probit* estimation on the selection equation.
MLE gives

(i) An estimation $\hat{\gamma}_{MLE}$ captures the *indirect effect* of regressors in $\mathbf{w}$ on y through the selection equation.

And compute

$$\hat{\lambda}(\mathbf{w}'\hat{\gamma}_{MLE}) \equiv \frac{\phi(\mathbf{w}'\hat{\gamma}_{MLE})}{\Phi(\mathbf{w}'\hat{\gamma}_{MLE})} \tag{3.19}$$

[3] For each variable, the i subscript is omitted in the derivation

Step 2 Run OLS

$$y = \mathbf{x}'\beta + \rho\sigma_\epsilon\hat{\lambda} + \eta \text{ where } \mathbb{E}[\eta|\mathbf{x}, \hat{\lambda}] = 0 \tag{3.20}$$

OLS gives

(i) An estimation $\hat{\beta}_{OLS}$ measures the *direct effect* of regressors in $\mathbf{x}$ on y through the outcome equation.

(ii) An estimation of $\widehat{\rho\sigma_\epsilon}$, given $\sigma_\epsilon > 0$, we can estimate the *sign* of ρ.

Special Case (i) Consider the special case where

$$\mathbf{w} = \mathbf{x} \tag{3.21}$$
$$\lambda(x) \text{ is linear} \tag{3.22}$$

then (2.14) and regression (2.20) can be written as

$$y = \mathbf{x}'\beta + \rho\sigma_\epsilon\mathbf{x}'\lambda(\gamma) + \eta \tag{3.23}$$
$$= \mathbf{x}'[\beta + \rho\sigma_\epsilon\lambda(\gamma)] + \eta \tag{3.24}$$

where $\beta + \rho\sigma_\epsilon\lambda(\gamma)$ represents the *mixed and non-separable* effect.

Special Case (ii) If

$$\mathbf{w} = [\mathbf{x}, z] \tag{3.25}$$
$$\lambda(x) \text{ is linear} \tag{3.26}$$
$$\tag{3.27}$$

Let the coefficients of $\mathbf{w}$ be $[\gamma, \theta]$, then

$$\lambda(\mathbf{w}[\gamma, \theta]) = \lambda(\mathbf{x}\gamma) + \lambda(z\theta) \tag{3.28}$$
$$= \mathbf{x}\lambda(\gamma) + z\lambda(\theta) \tag{3.29}$$

Then the regression can be rewritten as

$$y = \mathbf{x}'[\beta + \rho\sigma_\epsilon\lambda(\gamma)] + \rho\sigma_\epsilon z\lambda(\theta) + \eta \tag{3.30}$$

Remark 3.2. Therefore, if λ is linear, we need at least one exclusion variable to identify the direct and indirect effects. If λ is non-linear, it's *probably* fine.

4 Binary Outcome with Continuous Endogenous Regressors: Control Function Approach

4.1 Model

In ordinary binary outcome models, like Probit models, we assumed all regressors are *exogenous* ($Cov(x, \varepsilon) = 0$). But in many cases, we have some of the explanatory variables are endogenous. In this section, we are going to consider the case where the endogenous regressors are continuous.

Outcome Equation

$$y = \mathbb{I}\{\mathbf{x}_y\theta + \mathbf{w}\gamma + \varepsilon > 0\} \tag{4.1}$$

where

(i) $\mathbf{x}_y$: exogenous observable characteristics.

(ii) $\mathbf{w}$: endogenous observable regressors, which are continuous.

Similarly to the IV approach, we use another "auxiliary equation" to estimate $\mathbf{w}$:

$$\mathbf{w} = \mathbf{x}_w \eta + \sigma_w v \tag{4.2}$$

where the error terms in (3.1) and (3.2) follows

$$\begin{pmatrix} \varepsilon \\ v \end{pmatrix} \sim \mathcal{N}\Big(\mathbf{0},\ \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\Big) \tag{4.3}$$

Define $\tilde{\mathbf{x}} \equiv [\mathbf{x}_y, \mathbf{x}_w]$ as the set of usable regressors.

4.2 Maximum Likelihood Estimator

To estimate the model using MLE, we need to construct the likelihood function. By Bayesian Theorem,

$$f(y|\mathbf{w}, \tilde{\mathbf{x}}) = \frac{f(y, \mathbf{w}|\tilde{\mathbf{x}})}{f(\mathbf{w}|\tilde{\mathbf{x}})} \tag{4.4}$$

$$\iff f(y, \mathbf{w}|\tilde{\mathbf{x}}) = f(y|\mathbf{w}, \tilde{\mathbf{x}}) f(\mathbf{w}|\tilde{\mathbf{x}}) \tag{4.5}$$

By equation (3.2)

$$w|_{\tilde{\mathbf{x}}} \sim \mathcal{N}\Big(\mathbf{x}_w \eta, \sigma_w^2\Big) \tag{4.6}$$

$$\implies f(w|\tilde{\mathbf{x}}) = \frac{1}{\sqrt{2\pi}\sigma_w} e^{\frac{-(w - \mathbf{x}_w \eta)^2}{2\sigma_w^2}} \tag{4.7}$$

and to compute $f(y|w, \mathbf{x}_w \eta)$, since y is binary, we are going to compute $\mathbb{P}[y = 1|w, \tilde{\mathbf{x}}]$ first.

$$\mathbb{P}[y = 1|w, \tilde{\mathbf{x}}] = \mathbb{P}[-\varepsilon < \mathbf{x}_y \theta + w\gamma | w, \tilde{\mathbf{x}}] \tag{4.8}$$

$$= \mathbb{P}[-\varepsilon < \mathbf{x}_y \theta + w\gamma | v, \tilde{\mathbf{x}}] \tag{4.9}$$

Lemma 4.1. Given joint normal variables (ε, v) conditioned on $\tilde{\mathbf{x}}$ following

$$\begin{pmatrix} \varepsilon \\ v \end{pmatrix} |_{\tilde{\mathbf{x}}} \sim \mathcal{N}\Big(\mathbf{0},\ \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\Big) \tag{4.10}$$

then

$$\varepsilon|_{v, \tilde{\mathbf{x}}} \sim \mathcal{N}(\rho v,\ 1 - \rho^2) \tag{4.11}$$

which implies

$$\frac{\varepsilon - \rho v}{\sqrt{1 - \rho^2}} \sim \mathcal{N}(0,\ 1) \tag{4.12}$$

$$\implies \frac{-\varepsilon + \rho v}{\sqrt{1 - \rho^2}} \sim \mathcal{N}(0,\ 1) \tag{4.13}$$

by the symmetry of standard normal distribution.

Therefore,

$$\mathbb{P}[-\varepsilon < \mathbf{x}_y\theta + w\gamma | v, \tilde{\mathbf{x}}] \tag{4.14}$$
$$= \mathbb{P}[-\varepsilon + \rho v < \mathbf{x}_y\theta + w\gamma + \rho v | v, \tilde{\mathbf{x}}] \tag{4.15}$$
$$= \mathbb{P}[\frac{-\varepsilon + \rho v}{\sqrt{1-\rho^2}} < \frac{\mathbf{x}_y\theta + w\gamma + \rho v}{\sqrt{1-\rho^2}} | v, \tilde{\mathbf{x}}] \tag{4.16}$$
$$= \Phi(\frac{\mathbf{x}_y\theta + w\gamma + \rho v}{\sqrt{1-\rho^2}}) \tag{4.17}$$

4.3 Control Function

Step 1 Run OLS on $w = \mathbf{x}_w\eta + \sigma_w v$, Obtain estimations $\hat{\eta}_{OLS}$, $\hat{\sigma_w}_{OLS}$.

Step 2 Obtain estimation of v using the error terms and standard deviation in OLS results.

$$\hat{v} = \frac{w - \mathbf{x}_w\hat{\eta}_{OLS}}{\hat{\sigma}_{OLS}} \tag{4.18}$$

Step 3 Plug in $\hat{v}$ and run **probit** model in (3.17),

$$\Phi(\frac{\mathbf{x}_y\theta + w\gamma + \rho v}{\sqrt{1-\rho^2}}) \tag{4.19}$$
$$= \Phi(\frac{\mathbf{x}_y\theta}{\sqrt{1-\rho^2}} + \frac{w\gamma}{\sqrt{1-\rho^2}} + \frac{\rho v}{\sqrt{1-\rho^2}}) \tag{4.20}$$

Define

$$\theta^* \equiv \frac{\theta}{\sqrt{1-\rho^2}} \tag{4.21}$$
$$\gamma^* \equiv \frac{\gamma}{\sqrt{1-\rho^2}} \tag{4.22}$$
$$\alpha^* \equiv \frac{\alpha}{\sqrt{1-\rho^2}} \tag{4.23}$$

So the probit model can be written as

$$y = \mathbb{I}\{-\tilde{u} < \mathbf{x}_y\theta^* + w\gamma^* + v\alpha^*\} \tag{4.24}$$

where $\tilde{u} \sim \mathcal{N}(0,1)$.

Once we have an estimation on α^*, ρ can be calculated with

$$\rho = \pm\sqrt{\frac{\alpha^*}{1+\alpha^{*2}}} \tag{4.25}$$