PS5 Solutions

Jingle Fu

1 Problem 1

Solution (a).

We define the binary dependent variable y_i as:

$$y_i = \begin{cases} 1 & \text{if customer } i \text{ pays with cash,} \\ 0 & \text{if customer } i \text{ pays with card.} \end{cases}$$

We observe y_i as:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0, \\ 0 & \text{if } y_i^* \leq 0. \end{cases}$$

Thus, the probability that customer i pays with cash is:

$$P(y_i = 1 \mid x_i) = P(y_i^* > 0 \mid x_i) = P(\varepsilon_i > -x_i'\beta) = \Phi(x_i'\beta),$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution.

Interpretation of y_i^*:

The latent variable y_i^* reflects the unobservable utility difference between paying with cash and paying with a card. When $y_i^* > 0$, the customer prefers cash; otherwise, they prefer card payment.

Solution (b).

Let Age_i denote the age of customer i, and suppose it enters the model linearly:

$$x_i'\beta = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \mathbf{Z}_i,$$

where $\mathbf{Z}_i$ includes other explanatory variables.

Effect of Age Increasing by 5 Years:

The change in probability when age increases by 5 years is:

$$\Delta P = P(y_i = 1 \mid \text{Age}_i + 5, \mathbf{Z}_i) - P(y_i = 1 \mid \text{Age}_i, \mathbf{Z}_i).$$

Substituting the probit model:

$$\Delta P = \Phi\left(\beta_0 + \beta_1(\text{Age}_i + 5) + \beta_2\mathbf{Z}_i\right) - \Phi\left(\beta_0 + \beta_1\text{Age}_i + \beta_2\mathbf{Z}_i\right).$$

Simplify:

$$\Delta P = \Phi\left(x_i'\beta + 5\beta_1\right) - \Phi\left(x_i'\beta\right).$$

Dependence on Current Age and Other Variables:

- **Current Age (Age_i) and $\mathbf{Z}_i$:**

  The term $x_i'\beta$ depends on Age_i and $\mathbf{Z}_i$. Since Φ is a nonlinear function, the change ΔP depends on the initial value of $x_i'\beta$, which includes Age_i and other variables.

- **Other Variables ($\mathbf{Z}_i$):**

  Yes, the effect also depends on $\mathbf{Z}_i$ because they influence $x_i'\beta$.

Conclusion:

The effect of age increasing by 5 years on the probability of using cash depends on both the current age of the customer and the values of other explanatory variables.

Solution (c).

Yes, we could use a standard linear regression estimated via Ordinary Least Squares (OLS) to answer the question. This approach is known as the Linear Probability Model (LPM), where the binary dependent variable y_i is regressed on the explanatory variables:

$$E[y_i \mid x_i] = x_i'\beta.$$

However, the LPM has some limitations:

- **Predicted Probabilities:** The model may predict probabilities outside the $[0, 1]$ interval.

- **Heteroskedasticity:** The error term in the LPM is heteroskedastic, which affects the efficiency of OLS estimates.
- **Constant Marginal Effects:** The LPM assumes constant marginal effects, which may not capture the true relationship.

Despite these issues, the LPM can still provide consistent estimates of the average marginal effects under certain conditions.

Solution (d).

In the linear regression model, the expected value of y_i given x_i is:

$$E[y_i \mid x_i] = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \mathbf{Z}_i.$$

Effect of Age Increasing by 5 Years:

The change in expected probability when age increases by 5 years is:

$$\begin{aligned}\Delta E[y_i \mid x_i] &= E[y_i \mid \text{Age}_i + 5, \mathbf{Z}_i] - E[y_i \mid \text{Age}_i, \mathbf{Z}_i] \\ &= (\beta_0 + \beta_1(\text{Age}_i + 5) + \beta_2 \mathbf{Z}_i) - (\beta_0 + \beta_1 \text{Age}_i + \beta_2 \mathbf{Z}_i) \\ &= 5\beta_1.\end{aligned}$$

Dependence on Current Age and Other Variables:

- **Current Age (Age_i):**

  The change $\Delta E[y_i \mid x_i]$ does not depend on Age_i because β_1 is constant.

- **Other Variables ($\mathbf{Z}_i$):**

  The effect does not depend on $\mathbf{Z}_i$ since they cancel out in the difference.

Conclusion:

In the linear regression model, the effect of age increasing by 5 years is constant and does not depend on the current age of the customer or the values of other variables.

Solution (e).

Comparison of Marginal Effects:

- **Under Probit Model:**

  The marginal effect of age is:

$$\frac{\partial P(y_i = 1 \mid x_i)}{\partial \text{Age}_i} = \phi(x_i'\beta)\beta_1,$$

  where ϕ is the standard normal probability density function (PDF).

- **Under Linear Regression (LPM):**

  The marginal effect of age is:

$$\frac{\partial E[y_i \mid x_i]}{\partial \text{Age}_i} = \beta_1.$$

Customers with Similar Effects:

- **Middle-Range Probabilities:**

  For customers where $x_i'\beta$ is near zero, $\phi(x_i'\beta)$ is at its maximum. In this region, the probit model's marginal effects are largest and the relationship between x_i and y_i is approximately linear. Therefore, the LPM and probit model yield similar marginal effects.

- **Extreme Probabilities:**

  For customers where $x_i'\beta$ is very positive or very negative (leading to predicted probabilities near 1 or 0), $\phi(x_i'\beta)$ is small. The probit model's marginal effects diminish, whereas the LPM continues to predict constant marginal effects, potentially outside the $[0, 1]$ interval.

Functional Form Comparison:

- **LPM:**

$$E[y_i \mid x_i] = x_i'\beta.$$

  Linear in parameters and explanatory variables, leading to constant marginal effects.

- **Probit Model:**

$$E[y_i \mid x_i] = \Phi(x_i'\beta).$$

Nonlinear CDF of the standard normal distribution, leading to variable marginal effects dependent on x_i.

Conclusion:

The linear regression model (LPM) is a reasonable approximation when dealing with average effects in a sample with moderate probabilities. However, for individual predictions or when the probability of the outcome is near the extremes, the probit model provides a more accurate representation due to its nonlinear nature and variable marginal effects.

2 Problem 2

Solution (a).

We are asked to derive $P[y_i = 0 \mid x_i]$ given the Tobit model with lower censoring at δ.

Step 1: Understanding the Censoring Mechanism

- The observed variable y_i is zero when the latent variable y_i^* is less than or equal to δ:

$$y_i = 0 \quad \text{if} \quad y_i^* \leq \delta.$$

- The latent variable y_i^* is distributed as:

$$y_i^* = x_i'\beta + u_i, \quad u_i \sim N(0, \sigma^2).$$

Step 2: Deriving the Probability

- The probability that $y_i = 0$ is:

$$P[y_i = 0 \mid x_i] = P[y_i^* \leq \delta \mid x_i] = P(u_i \leq \delta - x_i'\beta).$$

- Since $u_i \sim N(0, \sigma^2)$, we can standardize:

$$P[u_i \leq \delta - x_i'\beta] = \Phi\left(\frac{\delta - x_i'\beta}{\sigma}\right),$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution.

So,

$$P[y_i = 0 \mid x_i] = \Phi\left(\frac{\delta - x_i'\beta}{\sigma}\right).$$

Solution (b).

We are to derive $E[y_i^* \mid x_i]$.

Step 1: Understanding y_i^*

- The latent variable y_i^* is normally distributed:

$$y_i^* \mid x_i \sim N(x_i'\beta, \sigma^2).$$

Step 2: Calculating the Conditional Mean

- The expected value is:

$$E[y_i^* \mid x_i] = x_i'\beta.$$

Solution (c).

Step 1: Understanding y_i

- The observed variable y_i is zero when $y_i^* \leq \delta$ and equals y_i^* when $y_i^* > \delta$.

Step 2: Expressing $E[y_i \mid x_i]$

- The expected value is:

$$E[y_i \mid x_i] = E[y_i^* \mid y_i^* > \delta, x_i] \cdot P[y_i^* > \delta \mid x_i] + 0 \cdot P[y_i^* \leq \delta \mid x_i].$$

- Simplify:

$$E[y_i \mid x_i] = E[y_i^* \mid y_i^* > \delta, x_i] \cdot \left[1 - \Phi\left(\frac{\delta - x_i'\beta}{\sigma}\right)\right] = E[y_i^* \mid y_i^* > \delta, x_i] \cdot \Phi\left(\frac{x_i'\beta - \delta}{\sigma}\right).$$

Step 3: Calculating $E[y_i^* \mid y_i^* > \delta, x_i]$

- For a truncated normal distribution:

$$E[y_i^* \mid y_i^* > \delta, x_i] = x_i'\beta + \sigma \frac{\phi(z_i)}{\Phi(z_i)},$$

where

$$z_i = \frac{x_i'\beta - \delta}{\sigma},$$

and ϕ is the standard normal probability density function (PDF).

Step 4: Combining Terms

- Multiply the expected value by the probability:

$$E[y_i \mid x_i] = \left[x_i'\beta + \sigma\frac{\phi(z_i)}{\Phi(z_i)}\right]\Phi(z_i).$$

- Simplify:

$$E[y_i \mid x_i] = x_i'\beta\Phi(z_i) + \sigma\phi(z_i).$$

Solution (d).

We are to derive the predicted effect of decreasing c_i by 10 percentage points on y_i^* and y_i.

Effect on y_i^*

- From part (b):

$$E[y_i^* \mid x_i] = x_i'\beta.$$

- The marginal effect with respect to c_i is:

$$\frac{\partial E[y_i^* \mid x_i]}{\partial c_i} = \beta_c,$$

where β_c is the coefficient of c_i in β. - A decrease of 10 percentage points ($\Delta c_i = -0.10$) results in:

$$\Delta E[y_i^* \mid x_i] = \beta_c \times (-0.10).$$

Effect on y_i

- From part (c):

$$E[y_i \mid x_i] = x_i'\beta\Phi(z_i) + \sigma\phi(z_i).$$

- The derivative with respect to c_i is:

$$\frac{\partial E[y_i \mid x_i]}{\partial c_i} = \beta_c\Phi(z_i) + x_i'\beta\left(\frac{\partial}{\partial c_i}\Phi(z_i)\right) + \sigma\frac{\partial\phi(z_i)}{\partial c_i}.$$

- Note that:

$$\frac{\partial\Phi(z_i)}{\partial c_i} = \phi(z_i)\frac{\partial z_i}{\partial c_i},$$

$$\frac{\partial\phi(z_i)}{\partial c_i} = \phi'(z_i)\frac{\partial z_i}{\partial c_i}.$$

- Since:

$$z_i = \frac{x_i'\beta - \delta}{\sigma},$$

$$\frac{\partial z_i}{\partial c_i} = \frac{\beta_c}{\sigma}.$$

- Therefore:

$$\frac{\partial E[y_i \mid x_i]}{\partial c_i} = \beta_c \Phi(z_i) + x_i'\beta \left(\phi(z_i) \frac{\beta_c}{\sigma} \right) + \sigma \left(\phi'(z_i) \frac{\beta_c}{\sigma} \right).$$

- Simplify:

$$\frac{\partial E[y_i \mid x_i]}{\partial c_i} = \beta_c \left[\Phi(z_i) + \frac{x_i'\beta}{\sigma} \phi(z_i) + \phi'(z_i) \right].$$

As

$$z_i = \frac{x_i'\beta - \delta}{\sigma} \sim N(0, 1),$$

we know that the PDF of z_i is:

$$\phi(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}},$$

so,

$$\phi'(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} \cdot (-z_i) = -z_i \phi(z_i),$$

and we have:

$$\frac{\partial E[y_i \mid x_i]}{\partial c_i} = \beta_c \left[\Phi(z_i) + \frac{x_i'\beta}{\sigma} \phi(z_i) - z_i \phi(z_i) \right].$$

- Recognizing that $x_i'\beta - \sigma z_i = \delta$, we have:

$$\frac{x_i'\beta}{\sigma} \phi(z_i) - z_i \phi(z_i) = \frac{\delta}{\sigma} \phi(z_i).$$

- Therefore:

$$\frac{\partial E[y_i \mid x_i]}{\partial c_i} = \beta_c \left[\Phi(z_i) + \frac{\delta}{\sigma} \phi(z_i) \right].$$

- The change in $E[y_i \mid x_i]$ is:

$$\Delta E[y_i \mid x_i] = \frac{\partial E[y_i \mid x_i]}{\partial c_i} \times (-0.10).$$

Final Answer:

- Effect on y_i^*:

$$\Delta E[y_i^* \mid x_i] = -0.10 \times \beta_c.$$

- Effect on y_i:

$$\Delta E[y_i \mid x_i] = -0.10 \times \beta_c \left[\Phi\left(\frac{x_i'\beta - \delta}{\sigma}\right) + \frac{\delta}{\sigma}\phi\left(\frac{x_i'\beta - \delta}{\sigma}\right) \right].$$

Solution (e).

We are to determine the effect of decreasing c_i on y_i using the linear regression on uncensored data and compare it to the Tobit model.

Effect in Linear Regression

- In the regression:

$$y_i = x_i'\gamma + v_i, \quad i \in U,$$

the effect of decreasing c_i is:

$$\Delta E[y_i \mid x_i, i \in U] = \gamma_c \times \Delta c_i.$$

- Assuming $\gamma = \beta$:

$$\Delta E[y_i \mid x_i, i \in U] = \beta_c \times \Delta c_i.$$

Comparison with Tobit Model

- From part (d), the effect under the Tobit model is:

$$\Delta E[y_i \mid x_i] = \beta_c \left[\Phi\left(\frac{x_i'\beta - \delta}{\sigma}\right) + \frac{\delta}{\sigma}\phi\left(\frac{x_i'\beta - \delta}{\sigma}\right) \right] \times \Delta c_i.$$

- The linear regression effect matches the Tobit effect when:

$$\Phi\left(\frac{x_i'\beta - \delta}{\sigma}\right) + \frac{\delta}{\sigma}\phi\left(\frac{x_i'\beta - \delta}{\sigma}\right) \approx 1.$$

- This occurs when $\frac{x_i'\beta-\delta}{\sigma}$ is significantly positive (i.e., $x_i'\beta$ is much larger than δ), so $\Phi\left(\frac{x_i'\beta-\delta}{\sigma}\right)$ is close to 1 and $\phi\left(\frac{x_i'\beta-\delta}{\sigma}\right)$ is small.

Conclusion

- The predicted effect under linear regression is close to the Tobit model for cities with high expected concentrations (large $x_i'\beta$).
- It differs significantly for cities where $x_i'\beta$ is close to or less than δ.

Solution (f).

Objective: Determine whether the OLS estimator $\hat{\gamma}$ from the regression on uncensored data ($i \in U$) consistently estimates β from the true model, and under what conditions OLS works better or worse.

True Model:

$$y_i^* = x_i'\beta + u_i, \quad u_i \sim N(0, \sigma^2).$$

Censoring Mechanism:

$$y_i = \begin{cases} 0, & \text{if } y_i^* \leq \delta, \\ y_i^*, & \text{if } y_i^* > \delta. \end{cases}$$

OLS Regression on Uncensored Data ($i \in U$):

$$y_i = x_i'\gamma + v_i, \quad i \in U \equiv \{i : y_i > \delta\}. \tag{2}$$

Goal: Assess whether $\hat{\gamma} \xrightarrow{p} \beta$ as $n_u \to \infty$.

Step 1: Analyze the OLS Estimator

For observations $i \in U$:

$$y_i = y_i^* = x_i'\beta + u_i.$$

Therefore, the error term in the regression is:

$$v_i = y_i - x_i'\gamma = x_i'\beta + u_i - x_i'\gamma = x_i'(\beta - \gamma) + u_i.$$

OLS Assumption: For OLS to yield consistent estimates, the error term v_i must be uncorrelated with x_i, i.e.,

$$E[v_i \mid x_i] = 0.$$

Step 2: Check the OLS Assumption

However, since u_i is observed only when $y_i > \delta$, and $y_i > \delta$ implies $u_i > \delta - x_i'\beta$, the distribution of u_i is truncated from below at $\delta - x_i'\beta$. Thus, u_i depends on x_i, and:

$$E[u_i \mid x_i, y_i > \delta] \neq 0.$$

Step 3: Compute $E[u_i \mid x_i, y_i > \delta]$

Given that $u_i \sim N(0, \sigma^2)$, we can know from the truncated normal distribution that:

$$E[u_i \mid u_i > c] = \mu + \sigma\lambda\left(\frac{c-\mu}{\sigma}\right) = \sigma\lambda\left(\frac{c}{\sigma}\right).$$

Take $c = \delta - x_i'\beta$, we can compute:

$$E[u_i \mid u_i > \delta - x_i'\beta] = \sigma\lambda_i,$$

where:

$$\lambda_i = \frac{\phi\left(\frac{x_i'\beta - \delta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta - \delta}{\sigma}\right)}$$

is the **inverse Mills ratio**.

Step 4: Show that $E[v_i \mid x_i] \neq 0$

Because $E[u_i \mid x_i, y_i > \delta] \neq 0$, the error term v_i is correlated with x_i:

$$E[v_i \mid x_i] = x_i'(\beta - \gamma) + E[u_i \mid x_i, y_i > \delta].$$

If γ is consistent with β, we have $E[v_i \mid x_i] = 0$, it must be that:

$$x_i'(\beta - \gamma) + E[u_i \mid x_i, y_i > \delta] \xrightarrow{p} E[u_i \mid x_i, y_i > \delta] = 0.$$

But since $E[u_i \mid x_i, y_i > \delta] \neq 0$, we get a contradiction. So, γ will be biased from β.

Conclusion on Consistency

Since $E[x_i u_i \mid y_i > \delta] \neq 0$, the bias does not vanish even as the sample size increases. Therefore, the OLS estimator $\hat{\gamma}$ is biased and inconsistent for estimating β.

Method 2

Step 1: Express the OLS Estimator

For the uncensored observations ($i \in U$), we have:

$$y_i = y_i^* = x_i'\beta + u_i.$$

The OLS estimator for γ is:

$$\hat{\gamma} = \left(\sum_{i \in U} x_i x_i'\right)^{-1} \sum_{i \in U} x_i y_i.$$

Normalizing by n_U (the number of observations in U):

$$\hat{\gamma} = \left(\frac{1}{n_U}\sum_{i \in U} x_i x_i'\right)^{-1} \left(\frac{1}{n_U}\sum_{i \in U} x_i y_i\right).$$

Step 2: Compute Probability Limits Using the Hint

Using the hint provided:

$$\frac{1}{n_U}\sum_{i \in U} z_i \xrightarrow{p} E[z_i \mid y_i^* > \delta],$$

for any random variable z_i. Therefore:

- **Denominator Converges To:**

$$\frac{1}{n_U}\sum_{i \in U} x_i x_i' \xrightarrow{p} E[x_i x_i' \mid y_i^* > \delta].$$

- **Numerator Converges To:**

$$\frac{1}{n_U}\sum_{i \in U} x_i y_i \xrightarrow{p} E[x_i y_i^* \mid y_i^* > \delta].$$

Thus, the probability limit of $\hat{\gamma}$ is:

$$\hat{\gamma} \xrightarrow{p} \gamma = \left(E[x_i x_i' \mid y_i^* > \delta]\right)^{-1} E[x_i y_i^* \mid y_i^* > \delta].$$

Step 3: Expand $E[x_i y_i^* \mid y_i^* > \delta]$

Since $y_i^* = x_i'\beta + u_i$:

$$\begin{aligned} E[x_i y_i^* \mid y_i^* > \delta] &= E[x_i x_i'\beta \mid y_i^* > \delta] + E[x_i u_i \mid y_i^* > \delta] \\ &= E[x_i x_i' \mid y_i^* > \delta]\beta + E[x_i u_i \mid y_i^* > \delta]. \end{aligned}$$

Substituting back into the expression for γ:

$$\hat{\gamma} = \beta + \left(E[x_i x_i' \mid y_i^* > \delta]\right)^{-1} E[x_i u_i \mid y_i^* > \delta].$$

Step 5: Compute $E[x_i u_i \mid y_i^* > \delta]$

We need to compute $E[x_i u_i \mid y_i^* > \delta]$. Note that:

- In the full sample, u_i is independent of x_i: $E[x_i u_i] = 0$.
- However, in the censored sample ($y_i^* > \delta$), u_i is truncated based on x_i:

$$y_i^* > \delta \implies u_i > \delta - x_i'\beta.$$

This truncation induces a correlation between x_i and u_i in the censored sample.

Step 6: Derive $E[x_i u_i \mid x_i, y_i^* > \delta]$

Given that $u_i \sim N(0, \sigma^2)$, we can know from the truncated normal distribution that:

$$E[u_i \mid u_i > c] = \mu + \sigma\lambda\left(\frac{c-\mu}{\sigma}\right) = \sigma\lambda\left(\frac{c}{\sigma}\right).$$

Take $c = \delta - x_i'\beta$, we can compute:

$$E[u_i \mid u_i > \delta - x_i'\beta] = \sigma\lambda_i,$$

where:

$$\lambda_i = \frac{\phi\left(\frac{x_i'\beta - \delta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta - \delta}{\sigma}\right)}$$

is the **inverse Mills ratio**.

Now, we can compute:

$$E[x_i u_i \mid y_i^* > \delta] = E\left[x_i E[u_i \mid x_i, y_i^* > \delta]\right] = E\left[x_i \sigma\lambda_i\right].$$

Substitute $E[x_i u_i \mid y_i^* > \delta] = \sigma E[x_i \lambda_i]$ into the expression for γ:

$$\gamma = \beta + (E[x_i x_i' \mid y_i^* > \delta])^{-1} \sigma E[x_i \lambda_i].$$

Step 7: Analyze the Bias Term

The second term:

$$(E[x_i x_i' \mid y_i^* > \delta])^{-1} \sigma E[x_i \lambda_i],$$

represents the **bias** in the OLS estimator due to the correlation between x_i and u_i in the censored sample.

Conclusion on Consistency:

Since $E[x_i \lambda_i] \neq 0$ (because λ_i depends on x_i), the OLS estimator $\hat{\gamma}$ is generally **inconsistent** for β.

Circumstances Affecting OLS Performance

OLS Works Better When:

- **Low Censoring Probability:**
  - If $x_i'\beta$ is much larger than δ for most observations, $x_i'\beta - \delta$ is positive and large in magnitude.
  - This makes λ_i small because:

$$\lambda_i = \frac{\phi\left(\dfrac{x_i'\beta - \delta}{\sigma}\right)}{\Phi\left(\dfrac{x_i'\beta - \delta}{\sigma}\right)} \approx 0.$$

  - Hence, $E[x_i \lambda_i] \approx 0$, and the bias is minimal.
- **Small Error Variance (σ^2):**
  - A smaller σ reduces the spread of u_i, decreasing the impact of censoring.
  - This leads to a smaller λ_i, reducing the bias.

OLS Works Worse When:

- **High Censoring Probability:**

  - If $x_i'\beta$ is close to δ for many observations, $x_i'\beta - \delta$ is near zero.
  - λ_i becomes significant, leading to a larger bias.

- **Large Error Variance (σ^2):**
  - A larger σ increases the spread of u_i, increasing the chance that $y_i^* \leq \delta$.
  - This results in more censoring and a larger λ_i, amplifying the bias.