Maximum Likelihood and Binary Dependent Variable Models

Matteo Paradisi
(EIEF and NBER)

Applied Micro - Lecture 9
Limited Dependent Variable

- So far Y, the dependent variable, was continuous.
- However, dependent variables could be *dichotomous* (dummy variables) or *categorical*.
- Hence, we study non-linear estimation with dichotomous Y vars.
- Some examples:
  - Votes (Left vs Right)
  - Labor force participation (extensive margin)
  - College dropout
Example and Intuition

- Suppose you want to relate the share of blacks in the population to whether a black representative is elected.
- Dependent variable: dummy = 1 if elected.
- Now plot the data to see the relationship.
Example and Intuition
Example and Intuition - Linear Fit
Example and Intuition

- A line does not fit the data well
- We need something better
- AND something that will predict values between 0 and 1
- What can we do?
Example and Intuition

- Think of dependent variable as a **probability of the event**
- We need a function that takes continuous values and provides something in the \([0, 1]\) interval
- Which function does this?
- Example: the CDF of a Normal distribution!

\[
Y = \Phi(X\beta + \epsilon)
\]

- Must be careful about interpretation of \(\beta\)s (we’ll see it later)
- This model fits the data better!
Example and Intuition - Non-Linear Fit
Estimating the Model

- This model allows us to better fit the data.
- However, how do we estimate it?
- We need to introduce the concept of maximum likelihood estimation.
Maximum Likelihood Estimation
What is maximum likelihood?

Estimation method: find values of parameters that maximize the likelihood of observing the sample at hand

Two steps:
1. Write a closed-form of the likelihood
   - function of data and parameters
2. Maximize it to find estimates

These methods are particularly useful in models where the dependent variable is a discrete choice
Probit Model

- We start from the simplest model
- Let’s assume that our theory gives us a latent variable \( y^* \) determined by some \( x \)
  \[ y^* = x\beta + u \]
- However, \( y^* \) is not observed. It could be for instance the utility from a choice
- We instead observe a binary choice
  \[ y = \begin{cases} 
  1 & \text{if } y^* > 0 \\
  0 & \text{if } y^* \leq 0
  \end{cases} \]
- Notice that 0 is just a normalization
- Example: if \( y \) is choice about entering the labor force, then you enter if utility from entering is greater than alternative (normalized to 0)
Probit Model

- Suppose that \((y_i, x_i)_{i \in \mathbb{N}}\) are i.i.d.
- We need to find a way to write the probability of observing a vector of choices and characteristics.
- Assume that \(u \sim \mathcal{N}(0, 1)\) then

\[
\Pr(y_i = 1 | x_i, \beta) = \Pr(x_i\beta + u_i > 0 | x_i) \\
= \Pr(u_i > -x_i\beta | x_i, \beta) \\
= 1 - \Phi(-x_i\beta) = \Phi(x_i\beta)
\]

where \(\Phi(\cdot)\) is the cdf of a Normal distribution.

- Analogously

\[
\Pr(y_i = 0 | x_i, \beta) = 1 - \Phi(x_i\beta)
\]
Likelihood Function

We can then write the likelihood of observing \((y_i, x_i)\)

\[ p(y_i | x_i, \beta) = [\Phi(x_i \beta)]^{y_i} [1 - \Phi(x_i \beta)]^{1-y_i} \]

also called "likelihood contribution" of i

Hence, the likelihood for the entire sample is

\[ L(\beta) = \prod_{i=1}^{N} p(y_i | x_i, \beta) \]

since \(y_i\) and \(x_i\) are observed, \(\beta\) is the only unknown
Maximizing the Likelihood

- In most cases, algorithms maximize a monotonic transformation of $L$

$$\hat{\beta}_{ML} = \arg \max_{\beta \in \Theta} \log L(\beta) = \arg \max_{\beta \in \Theta} \sum_{i=1}^{N} \ln p(y_i | x_i, \beta)$$

- the transformation takes the log of $L$ and it is referred to as log-likelihood

- Notice that $\Theta$ is a generic set, so that one can add additional constraints on $\beta$
We can apply maximum likelihood to linear regressions too.

Consider the model

\[ y_i = x_i \beta + u_i \]

We add assumption to standard OLS assumptions:

\[ u \sim N(0, \sigma^2) \]
Linear Regression and Maximum Likelihood

- Because data is continuous in this case, we cannot write a probability function.
- We write a distribution function instead:
  \[ f(y_i | x_i, \beta, \sigma) = f(u_i = y_i - x_i \beta | x_i) \]
  \[ = \phi \left( \frac{y_i - x_i \beta}{\sigma} \right) \]

- Hence, log-likelihood is:
  \[ L(\beta, \sigma) = \sum_{i=1}^{N} \ln \phi \left( \frac{y_i - x_i \beta}{\sigma} \right) \]
Linear Regression and Maximum Likelihood

Rewrite the likelihood as

\[ L(\beta, \sigma) = \sum_{i=1}^{N} \ln \phi \left( \frac{y_i - x_i \beta}{\sigma} \right) \]

\[ = \sum_{i=1}^{N} \ln \left[ \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{y_i - x_i \beta}{\sigma} \right)^2 \right) \right] \]

\[ = \sum_{i=1}^{N} \left[ -\ln \sigma - \frac{1}{2} \ln 2\pi - \frac{1}{2\sigma^2} (y_i - x_i \beta)^2 \right] \]

\[ = -N \ln \sigma - \frac{N}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - x_i \beta)^2 \]
Linear Regression and Maximum Likelihood

\[ L(\beta, \sigma) = -N \ln \sigma - \frac{N}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - x_i\beta)^2 \]

▶ FOCs are

\[ \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i' (y_i - x_i\beta) = 0 \]

\[-\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{N} (y_i - x_i\beta)^2 = 0 \]

▶ Combining them

\[ \hat{\beta}_{ML} = \left[ \sum_{i=1}^{N} x_i' x_i \right]^{-1} \left[ \sum_{i=1}^{N} x_i' y_i \right] \]

\[ \hat{\sigma}_{ML}^2 = N^{-1} \sum_{i=1}^{N} (y_i - x_i\beta)^2 \]
Linear Regression and Maximum Likelihood

\[ \hat{\beta}_{ML} = \left[ \sum_{i=1}^{N} x_i' x_i \right]^{-1} \left[ \sum_{i=1}^{N} x_i' y_i \right] \]

\[ \hat{\sigma}^2_{ML} = N^{-1} \sum_{i=1}^{N} (y_i - x_i \beta)^2 \]

▶ These are OLS formulas!

▶ However, the result depends on the assumption on \( y \)'s distribution

▶ With a different distribution, OLS would not be a maximum likelihood estimator for this linear model
Formal Characterization of Maximum Likelihood

Let’s be a little more formal, derive the maximum likelihood, and discuss some properties

Assumption 1: $y|\mathbf{x} \sim \text{i.i.d. } F(\cdot|\theta)$

Conditioning on $\theta$ emphasizes the fact that $F$ is a function of the parameters to be estimated
Formal Characterization of Maximum Likelihood

- Log likelihood contribution of i
  \[ \ell_i (\theta) = \ln f (y_i | x_i, \theta) \]

- Total log likelihood is
  \[ L (\theta) = \sum_{i=1}^{N} \ell_i (\theta) = \sum_{i=1}^{N} \ln f (y_i | x_i, \theta) \]

- The maximum likelihood estimator is
  \[ \hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} L (\theta) \]

- We must add assumptions to make \( \hat{\theta}_{ML} \) consistent and asymptotically efficient
Additional Assumptions

- Additional assumptions:
  - a) $\Theta$ is closed and bounded
  - b) $f$ is continuous and twice differentiable over $\Theta$
  - c) $f(y|\theta)$ is such that $f(y|\theta_1) = f(y|\theta_2)$ if and only if $\theta_1 = \theta_2$

- The last assumption makes sure that $L$ is never flat and there is a unique maximizer
Conditional VS Full Maximum Likelihood

- So far the formal model started from a conditional distribution of $y|x$
- But the probit model presented before relied on a joint distribution of $x$ and $y$
- Notice that
  $$f(y, x|\theta) = f(y|x, \theta) f(x|\theta)$$

- If the distribution of $x$ does not depend on $\theta$ (i.e. $f(x|\theta) = f(x)$) then the $\theta$ maximizing $f(y|x, \theta)$ also maximizes $f(y, x|\theta)$
- If this is not the case, a full maximum likelihood is needed
What is the asymptotic distribution of $\hat{\theta}_{ML}$?

$$\hat{\theta}_{ML} \sim N \left[ \theta, -E \left( \frac{\partial^2 L(\theta)}{\partial \theta \partial \theta'} \right)^{-1} \right]$$

First, the mean is $\theta$, so the estimator is **consistent**.

Second, the variance is the negative of the inverse of the Hessian matrix. This is called **Cramer-Rao lower bound** and it is the smallest possible variance estimator. It follows that the estimator is also efficient!
Discussion on Maximum Likelihood VS OLS

- ML is **consistent and efficient**, so this is like the best we could ask for.

- However, we have to put very strong assumptions to derive these estimators.

- In particular, we have to take a stance on data distribution.

- OLS only makes assumption on the conditional mean of the distribution ($E(u|X) = 0$), not on the entire conditional distribution.

- However, the mild assumption of the OLS does not guarantee efficiency.

- ML is attractive, but if wrong assumption on distribution then we lose efficiency AND consistency.
Binary Choice Models
Binary Choice Models Formally

Let’s analyze binary choice models in greater details

Let’s go back to probit model

\[ y^* = X\beta + u \]

We observe

\[
y = \begin{cases} 
1 & \text{if } y^* > 0 \\
0 & \text{if } y^* \leq 0 
\end{cases}
\]

To derive ML we need assumption on the distribution of \( u \)

We assume it is normally distributed, this time with some variance \( \sigma^2 \)

\[ u \sim N \left( 0, \sigma^2 \right) \]
Probit Model

- Probability of observing $y = 1$ and $y = 0$

\[
\Pr (y = 1 | X, \beta) = \Pr (X\beta + u > 0 | X) \\
= \Pr \left( \frac{u}{\sigma} > -\frac{X\beta}{\sigma} | X, \beta \right) \\
= \Phi \left( \frac{X\beta}{\sigma} \right)
\]

\[
\Pr (y = 0 | X, \beta) = 1 - \Phi \left( \frac{X\beta}{\sigma} \right)
\]

- Log likelihood is

\[
L (\beta, \sigma) = \sum_{i=1}^{N} \left\{ y_i \ln \Phi \left( \frac{X\beta}{\sigma} \right) + (1 - y_i) \ln \left[ 1 - \Phi \left( \frac{X\beta}{\sigma} \right) \right] \right\}
\]
Cannot identify both $\beta$ and $\sigma$

$$L(\beta, \sigma) = \sum_{i=1}^{N} \left\{ y_i \ln \Phi \left( \frac{X \beta}{\sigma} \right) + (1 - y_i) \ln \left[ 1 - \Phi \left( \frac{X \beta}{\sigma} \right) \right] \right\}$$

- Notice that $\beta$ and $\sigma$ always appear as a ratio
- Cannot identify both since if $\hat{\beta}$ and $\hat{\sigma}$ maximize $L$, then also $\tilde{\beta} = c \hat{\beta}$ and $\tilde{\sigma} = c \hat{\sigma}$ do
- For this reason, any assumption on $\sigma$ is irrelevant
- Hence, as a convention probit models assume $\sigma = 1$

$$\hat{\beta}_{ML} \in \arg\max_{\beta} L(\beta, 1)$$
Interpreting $\beta$s

- In OLS, $\hat{\beta}_k = \frac{\partial y}{\partial x_k}$, is the effect of $x_k$ on $y$ (partial derivative interpretation)

- But remember how we setup the probit: $\hat{\beta}_{ML}$ is the effect on the latent variable $y^*$, NOT the effect on the probability that $y = 1$.

- For an economic interpretation we want to know the effect on the probability that $y = 1$

- Hence

$$\Pr(y = 1 | X) = F(X\beta) = F(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K)$$

- The effect we are interested in is

$$\frac{\partial \Pr(y = 1 | X)}{\partial x_k} = f(X\beta) \beta_k$$

- These are called marginal effects
The effect is small for small X
Marginal Effects - Graphical Intuition

The effect is larger for average X
Marginal Effects

- Marginal effects depend on $F(X\beta)$ and therefore are not constant.

- Normally people look for meaningful points at which presenting marginal effects.

- One option is the mean of $X$s

\[
\frac{\partial \Pr(y = 1|X)}{\partial x_k} = f(\bar{X}\beta) \beta_k
\]

- Where $\bar{X}$ is the vector of means for $X$s.

- If $x_k$ is a dummy variable (e.g. gender), marginal effects should show

\[
\Pr(y = 1|\bar{X}_{-k}, x_k = 1) - \Pr(y = 1|\bar{X}_{-k}, x_k = 0)
\]
More on Marginal Effects

▶ The ML coefficients however already contain some useful info
▶ First, their sign shows us the direction of the effect of $x_k$
▶ Second, they tell us the relative importance of marginal effects

$$\frac{\partial \Pr(y = 1 \mid X)}{\partial x_k} / \frac{\partial \Pr(y = 1 \mid X)}{\partial x_j} = \frac{\beta_k}{\beta_j}$$

▶ If $\beta_k > \beta_j$, then $x_k$ has a greater marginal effect than $x_j$
The Linear Probability Model

- An alternative to the probit is a simple linear probability model
- Linear model with dichotomous dependent variable y
- Hence
  \[
  \Pr(y = 1) = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K = X\beta
  \]

- What is the problem here?
- Nothing constraints \( \Pr(y = 1) \) to be between 0 and 1
- Indeed
  \[
  \mathbb{E}(y|X) = 1 \times \Pr(y = 1|X) + 0 \times \Pr(y = 0|X) = \Pr(y = 1|X) = X\beta
  \]
The Linear Probability Model

- The error term is also a dichotomous variable
  \[ u = \begin{cases} 
  1 - X\beta & \text{if } y = 1 \text{ with probability } \Pr(y = 1|X) = X\beta \\
  -X\beta & \text{if } y = 0 \text{ with probability } \Pr(y = 0|X) = 1 - X\beta 
  \end{cases} \]

- The mean is zero
  \[
  E(u|X) = (1 - X\beta) \times \Pr(y = 1|X) - X\beta \times \Pr(y = 0|X) \\
  = (1 - X\beta) X\beta - X\beta (1 - X\beta) = 0 
  \]

- and variance
  \[
  \text{Var}(u|X) = E\left(u^2|X\right) = (1 - X\beta)^2 X\beta - (X\beta)^2 (1 - X\beta) \\
  = (1 - X\beta) X\beta 
  \]

- since the variance is not constant the model is necessarily heteroskedastic
The Linear Probability Model

- We can deal with heteroskedasticity the usual way (GLS)
- For consistency we only need $E(X'u) = 0$
- The big advantage of this model however is that it requires milder assumptions than the probit
- Also, easier to interpret: $\beta$ is already the effect on the probability
- Last but not least, it can easily be used with panel data unlike the probit
The logit model is another alternative to the probit

It works the same way, but with a different assumption on the distribution of $y$

Logit assumes that $u \sim F(u) = e^{-e^{-u}}$

This is called Type I extreme distribution

It seems ugly, but has nice properties that we will see in multinomial models
Some Applications
The Effects of Different Political Campaigns

- **Wantchekon (2003)** studies the effects of clientelism in political campaigns

- The focus is on developing countries

- Research question: is a purely clientelist political platform more effective than a purely public policy one?

- Experimental design in Benin to answer the question
Experimental Design

- Two interventions: present the same policies with a clientelist framing vs a "public policy" framing
- You need three groups: two treatment arms and one control
- Each group is composed of different villages to avoid spillovers
## Experimental Design

<table>
<thead>
<tr>
<th>District</th>
<th>Exp. Candidate</th>
<th>Exp. Villages</th>
<th>Treatment</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kandi</td>
<td>Kerekou</td>
<td>Kassakou</td>
<td>clientelism</td>
<td>Bariba (92%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Keferi</td>
<td>public policy</td>
<td>Bariba (90%)</td>
</tr>
<tr>
<td>Nikki</td>
<td>Kerekou</td>
<td>Ouenou</td>
<td>clientelism</td>
<td>Bariba (89%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kpawolou</td>
<td>public policy</td>
<td>Bariba (88%)</td>
</tr>
<tr>
<td>Bembereke</td>
<td>Saka Lafia</td>
<td>Bembereke Est</td>
<td>clientelism</td>
<td>Bariba (86%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wannarou</td>
<td>public policy</td>
<td>Bariba (88%)</td>
</tr>
<tr>
<td>Perere</td>
<td>Saka Lafia</td>
<td>Tisserou</td>
<td>clientelism</td>
<td>Bariba (93%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alafiarou</td>
<td>public policy</td>
<td>Bariba (94%)</td>
</tr>
<tr>
<td>Abomey-Bohicon</td>
<td>Soglo</td>
<td>Agnangnan</td>
<td>clientelism</td>
<td>Fon (99%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gnidzoun</td>
<td>public policy</td>
<td>Fon (99%)</td>
</tr>
<tr>
<td>Ouidah-Pahou</td>
<td>Soglo</td>
<td>Acadjame</td>
<td>clientelism</td>
<td>Fon (99%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ahozon</td>
<td>public policy</td>
<td>Fon (99%)</td>
</tr>
<tr>
<td>Aplahoue</td>
<td>Amoussou</td>
<td>Boloume</td>
<td>clientelism</td>
<td>Adja (99%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Avetuime</td>
<td>public policy</td>
<td>Adja (96%)</td>
</tr>
<tr>
<td>Dogbo-Toviklin</td>
<td>Amoussou</td>
<td>Dekandji</td>
<td>clientelism</td>
<td>Adja (99%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Avedjin</td>
<td>public policy</td>
<td>Adja (99%)</td>
</tr>
<tr>
<td>Parakou</td>
<td>Ker./Lafia</td>
<td>Guema</td>
<td>competition</td>
<td>Bariba (80%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thiam</td>
<td>competition</td>
<td>Bariba (82%)</td>
</tr>
<tr>
<td>Come</td>
<td>Am./Soglo</td>
<td>Kande</td>
<td>competition</td>
<td>Adja (90%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tokan</td>
<td>competition</td>
<td>Adja (95%)</td>
</tr>
</tbody>
</table>
Probit Analysis

- Estimate the following probit model:

  \[ y_{ik}^1 = \alpha + \beta X_i + \lambda y_{ik}^0 + \gamma CL_k + \delta PB_k + \varepsilon_i \]

- We observe \( y_{ik} = 1 \) if \( y_{ik}^* > 0 \) and \( y_{ik} = 0 \) if \( y_{ik}^* \leq 0 \)

- \( CL_k \): clientist treatment; \( PB_k \): public policy treatment

- \( y_{ik}^0 \): past vote
### Table 5

**Probit Analysis of Vote for Type of Candidates in Treatment Villages**

<table>
<thead>
<tr>
<th></th>
<th>Southern</th>
<th>Northern</th>
<th>Local</th>
<th>National</th>
<th>Incumbent</th>
<th>Opposition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>-0.946**</td>
<td>-0.513</td>
<td>-0.367</td>
<td>-0.741</td>
<td>-0.186</td>
<td>0.222</td>
</tr>
<tr>
<td></td>
<td>(0.395)</td>
<td>(0.374)</td>
<td>(0.306)</td>
<td>(0.469)</td>
<td>(0.415)</td>
<td>(0.271)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>-0.513*</td>
<td>-0.516***</td>
<td>-0.424</td>
<td>-0.828**</td>
<td>-0.415</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.200)</td>
<td>(0.194)</td>
<td>(0.330)</td>
<td>(0.332)</td>
<td>(0.370)</td>
<td>(0.231)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>0.006</td>
<td>-0.003</td>
<td>-0.009*</td>
<td>0.011*</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.005)</td>
</tr>
<tr>
<td><strong>Past</strong></td>
<td>2.139***</td>
<td>.865***</td>
<td>1.555***</td>
<td>2.057***</td>
<td>1.893***</td>
<td>0.966***</td>
</tr>
<tr>
<td></td>
<td>(0.203)</td>
<td>(0.235)</td>
<td>(0.201)</td>
<td>(0.271)</td>
<td>(0.180)</td>
<td>(0.215)</td>
</tr>
<tr>
<td><strong>Public policy</strong></td>
<td>0.309**</td>
<td>-0.372***</td>
<td>-0.594*</td>
<td>0.429</td>
<td>-0.287</td>
<td>0.512*</td>
</tr>
<tr>
<td></td>
<td>(0.333)</td>
<td>(0.365)</td>
<td>(0.318)</td>
<td>(0.427)</td>
<td>(0.387)</td>
<td>(0.290)</td>
</tr>
<tr>
<td><strong>Clientelist</strong></td>
<td>1.004**</td>
<td>0.264</td>
<td>0.444</td>
<td>0.550</td>
<td>0.344</td>
<td>0.754**</td>
</tr>
<tr>
<td></td>
<td>(0.447)</td>
<td>(0.391)</td>
<td>(0.342)</td>
<td>(0.457)</td>
<td>(0.468)</td>
<td>(0.319)</td>
</tr>
<tr>
<td><strong>Sex*Client.</strong></td>
<td>-0.502</td>
<td>-0.191</td>
<td>-0.348</td>
<td>0.489</td>
<td>0.208</td>
<td>-0.324</td>
</tr>
<tr>
<td></td>
<td>(0.505)</td>
<td>(0.435)</td>
<td>(0.379)</td>
<td>(0.548)</td>
<td>(0.539)</td>
<td>(0.364)</td>
</tr>
<tr>
<td><strong>Sex*Public Pol.</strong></td>
<td>0.167</td>
<td>-1.050**</td>
<td>0.147</td>
<td>-0.572</td>
<td>-0.111</td>
<td>-0.773**</td>
</tr>
<tr>
<td></td>
<td>(0.402)</td>
<td>(0.414)</td>
<td>(0.358)</td>
<td>(0.482)</td>
<td>(0.450)</td>
<td>(0.345)</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>524</td>
<td>543</td>
<td>596</td>
<td>510</td>
<td>472</td>
<td>602</td>
</tr>
<tr>
<td><strong>log-L</strong></td>
<td>-145.250</td>
<td>-208.538</td>
<td>-284.050</td>
<td>-115.986</td>
<td>-146.161</td>
<td>-244.583</td>
</tr>
</tbody>
</table>
Changes in Production Inputs

- Conley and Udry (2010) study how information changes choices of inputs.
- They study the effect of good and bad news for information neighbors.
- See whether farmers adjust to align with those of neighbors.
- Study adoption of fertilizer (new technology) in Ghana.
Changes in Production Inputs: Logit

They run the following specification:

\[
\Pr ( \Delta x_{it} \neq 0 ) = \Lambda \left[ \alpha_1 s ( \text{good}, x = x_{it} ) + \alpha_2 s ( \text{good}, x \neq x_{it} ) + \alpha_3 s ( \text{bad}, x = x_{it} ) + \alpha_4 s ( \text{bad}, x \neq x_{it} ) + \alpha_5 \tilde{I}_{it} + z'_{it} \alpha_6 \right]
\]

- \( s ( \text{good}, x = x_{it} ) \): share of neighbors with same technology with good news
- \( \tilde{I}_{it} \): difference with growing conditions nearby
Results of Logit Analysis

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable: Indicator for change between zero and positive</td>
<td>-0.94 (1.24)</td>
<td>-0.08 (0.95)</td>
<td>-0.34 (0.84)</td>
</tr>
<tr>
<td></td>
<td>s(good, x = x_{previous})</td>
<td>[-0.04]</td>
<td>[-0.01]</td>
</tr>
<tr>
<td>Good news at previous input use</td>
<td>1.15 (0.81)</td>
<td>1.64 (0.78)</td>
<td>2.35 (1.80)</td>
</tr>
<tr>
<td></td>
<td>s(good, x ≠ x_{previous})</td>
<td>[0.03]</td>
<td>[0.09]</td>
</tr>
<tr>
<td>Good news at alternative fertilizer use</td>
<td>6.38 (2.86)</td>
<td>4.32 (1.93)</td>
<td>4.16 (1.80)</td>
</tr>
<tr>
<td></td>
<td>s(bad, x = x_{previous})</td>
<td>[0.15]</td>
<td>[0.20]</td>
</tr>
<tr>
<td>Bad news at lagged fertilizer use</td>
<td>-6.72 (3.04)</td>
<td>-5.90 (2.57)</td>
<td>-3.05 (1.85)</td>
</tr>
<tr>
<td></td>
<td>s(bad, x ≠ x_{previous})</td>
<td>[-0.09]</td>
<td>[-0.15]</td>
</tr>
<tr>
<td>Ave. abs. dev. from geog. neighbors’ fertilizer use</td>
<td>0.09 (0.10)</td>
<td>0.15 (0.07)</td>
<td>0.08 (0.04)</td>
</tr>
<tr>
<td></td>
<td>Γ_{ij}</td>
<td>[0.07]</td>
<td>[0.24]</td>
</tr>
<tr>
<td>Novice farmer</td>
<td>2.32 (0.75)</td>
<td>1.97 (0.89)</td>
<td>1.22 (0.92)</td>
</tr>
<tr>
<td></td>
<td>[0.26]</td>
<td>[0.43]</td>
<td>[0.30]</td>
</tr>
<tr>
<td>Talks with extension agent</td>
<td>-0.48 (0.61)</td>
<td>-1.35 (0.67)</td>
<td>-1.38 (0.76)</td>
</tr>
<tr>
<td></td>
<td>[-0.05]</td>
<td>[-0.29]</td>
<td>[-0.34]</td>
</tr>
<tr>
<td>Wealth (million cedis)</td>
<td>0.20 (0.10)</td>
<td>0.18 (0.13)</td>
<td>0.10 (0.12)</td>
</tr>
<tr>
<td></td>
<td>[0.06]</td>
<td>[0.10]</td>
<td>[0.06]</td>
</tr>
<tr>
<td>Clan 1</td>
<td>1.62 (1.14)</td>
<td>1.59 (1.10)</td>
<td>2.15 (1.03)</td>
</tr>
<tr>
<td></td>
<td>[0.18]</td>
<td>[0.35]</td>
<td>[0.54]</td>
</tr>
<tr>
<td>Clan 2</td>
<td>4.54 (1.45)</td>
<td>2.15 (1.23)</td>
<td>2.51 (0.99)</td>
</tr>
<tr>
<td></td>
<td>[0.51]</td>
<td>[0.47]</td>
<td>[0.63]</td>
</tr>
<tr>
<td>Church 1</td>
<td>1.84 (0.93)</td>
<td>-0.29 (0.73)</td>
<td>-0.24 (0.77)</td>
</tr>
<tr>
<td></td>
<td>[0.21]</td>
<td>[-0.06]</td>
<td>[-0.06]</td>
</tr>
</tbody>
</table>