Discrete Choice Models

Binary Models

Mohammad Arshad Rahman
(webpage: https://www.arshadrahman.com)

Course: Econometrics
Chennai Mathematical Institute

Gaussian/Normal Distribution

Suppose $X \sim N(\mu, \sigma^2)$, then the probability density function (*pdf*) of X is given by,

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)}, \qquad -\infty < x, \mu, < \infty; \sigma > 0,$$

and the *cdf* is given by,

$$F_X(x|\mu,\sigma) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-\mu)^2/(2\sigma^2)}\, dt.$$

Gaussian/Normal Distribution

The *mgf*, mean and variance of X are as follows,

$$\begin{aligned} E(X) &= \mu, \\ V(X) &= \sigma^2 \\ M_X(t) &= E(e^{tX}) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}. \end{aligned}$$

We get a *standard normal distribution* when $\mu = 0$ and $\sigma = 1$.

Logistic Distribution

Suppose $X \sim \mathcal{L}(a, b)$, then the pdf of X is given by,

$$f_X(x|a, b) = \frac{1}{b} \frac{\exp\left\{-\frac{(x-a)}{b}\right\}}{\left[1 + \exp\left\{-\frac{(x-a)}{b}\right\}\right]^2}, \qquad -\infty < x, a, < \infty; b > 0,$$

and the *cdf* is given by,

$$F_X(x|a, b) = \frac{1}{\left[1 + \exp\left\{-\frac{(x-a)}{b}\right\}\right]}.$$

Logistic Distribution

The *mgf*, mean and variance of X are as follows,

$$\begin{aligned} E(X) &= a, \\ V(X) &= b^2\frac{\pi^2}{3} \\ M_X(t) &= E(e^{tX}) = e^{at}\Gamma(1-\beta t)\Gamma(1+\beta t), \quad |t| < 1/\beta. \end{aligned}$$

We have a *standard logistic distribution* when $a = 0$ and $b = 1$. The first two moments for $\mathcal{L}(0,1)$ are: $E(X) = 0$, $V(X) = \frac{\pi^2}{3}$.

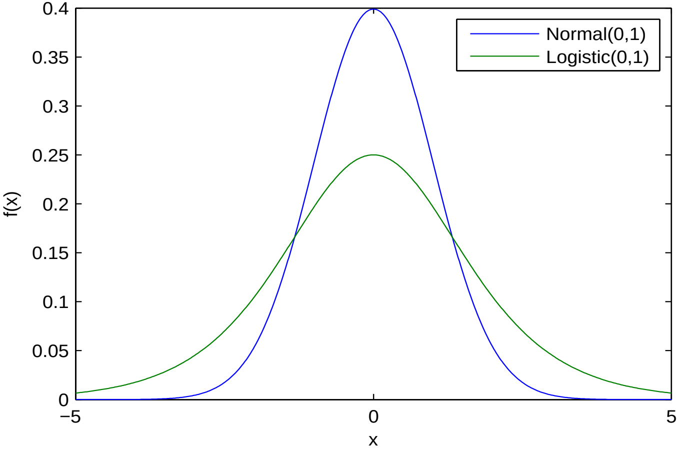

Figure 1: Pdf's for the Normal and Logistic distributions.

Binary Data Models

Binary models are designed to deal with situations where the outcome (dependent) variable is dichotomous i.e., takes only two values, typically coded as 1 for 'success' and 0 for 'failure'.

For example: when deciding to raise funds for firm expansion, we may model the outcome as 'success' if a firm chooses to raise funds by issuing new stocks and a 'failure' if it raises fund by issuing bonds.

Other examples: Choice of one of the parties in a two-party election, affiliation to union membership, and predicting bank failure.

We express a binary model in terms of a continuous latent random variable z_i as,

$$z_i = x_i'\beta + \varepsilon_i, \qquad \forall\, i = 1, \cdots, n, \tag{1}$$

where x_i is a $k \times 1$ vector of covariates, β is a $k \times 1$ vector of unknown parameters and n denotes the number of observations.

The error ε_i is stochastic and we assume that ε_i is independently and identically distributed (*i.i.d.*) from some distribution. Different distributional choices lead to different models.

The latent variable z_i is unobserved and relates to the observed binary outcome y_i as follows,

$$y_i = \begin{cases} 1 & \text{if } z_i > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where y_i denotes outcome of the i-th mortgage applicant in the context of the mortgage application.

The probability of binary outcomes and cut-points is pictorially represented in Figure 2.

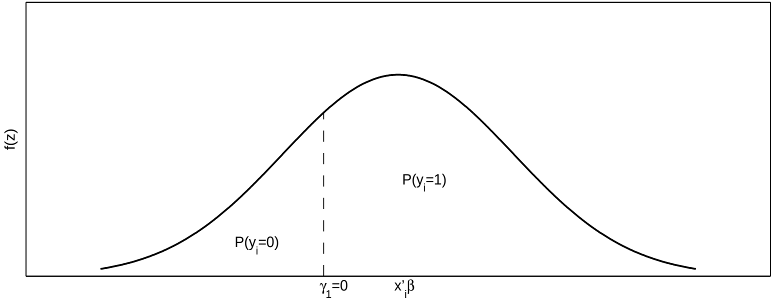

Figure 2: The cut-point γ_1 divides the area under the curve into two parts, the probability of failure and probability of success. Note that for each individual i the mean $x_i'\beta$ and hence the probabilities, $\mathrm{P}(y_i = 0)$ and $\mathrm{P}(y_i = 1)$, will be different.

Model Identification

Both location and scale restrictions are necessary to uniquely identify the parameters of the model.

Location restriction is achieved by fixing the first cut-point, for example $\gamma_1 = 0$.

Scale restriction requires the variance to be a constant.

Model Likelihood

The likelihood for a binary model expressed as a function of β is,

$$\ell(\beta; y) = \prod_{i=1}^{n} \left\{ P(y_i = 0|x_i'\beta) I(y_i = 0) + P(y_i = 1|x_i'\beta) I(y_i = 1) \right\}, \quad (3)$$

where $P(y_i = 1|x_i'\beta)$ denotes the probability of success and $I(\cdot)$ is an indicator function, which equals 1 if the condition within parenthesis is true and 0 otherwise.

Probit Model

Binary probit model arises when we assume that the errors are *i.i.d.* as a standard normal distribution i.e., $\varepsilon_i \sim N(0,1)$ for all $i = (1)n$. Hence, using equation (3), the binary probit model likelihood is,

$$\ell(\beta; y) = \prod_{i=1}^{n} \left\{ \Phi(-x_i'\beta) I(y_i = 0) + \Phi(x_i'\beta) I(y_i = 1) \right\}, \tag{4}$$

where $\Phi(\cdot)$ denotes the *cdf* of a standard normal distribution and $I(\cdot)$ is an indicator function defined earlier.

Probit Model

The probability of success used in equation (4) is obtained as,

$$
\begin{aligned}
P(y_i = 1|\beta) &= P(z_i > 0|\beta) = P(x_i'\beta + \varepsilon_i > 0|\beta) \\
&= P(\varepsilon_i > -x_i'\beta|\beta) \\
&= 1 - P(\varepsilon_i \leq -x_i'\beta|\beta) \\
&= 1 - \left[1 - P(\varepsilon_i \leq x_i'\beta|\beta)\right] \\
&= P(\varepsilon_i \leq x_i'\beta|\beta) \\
&= \Phi(x_i'\beta),
\end{aligned}
$$

where the fourth line follows from the property of a symmetric distribution.

Logit Model

Logit model arises when we assume that the errors are *i.i.d.* as a standard logistic distribution i.e., $\varepsilon_i \sim \mathcal{L}(0,1)$ for all $i = (1)n$. Hence, using equation (3), the logit model likelihood is,

$$\ell(\beta; y) = \prod_{i=1}^{n} \left\{ \Lambda(-x_i'\beta) I(y_i = 0) + \Lambda(x_i'\beta) I(y_i = 1) \right\}, \tag{5}$$

where $\Lambda(x_i'\beta) = e^{x_i'\beta}/(1+e^{x_i'\beta})$ denotes the *cdf* of a standard logistic distribution and $I(\cdot)$ is an indicator function defined earlier.

Estimation

The maximum likelihood estimate (MLE) for the probit model is obtained by maximizing the log-likelihood (4) with respect to β.

Similarly, MLE for logit model is obtained by maximizing the log-likelihood (5) with respect to β.

The MLE for probit and logit model do not have a closed form solution. So, the log-likelihood's are maximized using numerical techniques such as the Newton-Raphson method and BHHH algorithm.

Newton Raphson Algorithm

Suppose $L(\beta)$ denotes the log-likelihood function where maximization will be done w.r.t. β.

To determine the value of β_{t+1}, we take a second-order Taylor series approximation of $L(\beta_{t+1})$ around $L(\beta_t)$:

$$L(\beta_{t+1}) \approx L(\beta_t) + (\beta_{t+1} - \beta_t)' g_t + \frac{1}{2}(\beta_{t+1} - \beta_t)' H_t (\beta_{t+1} - \beta_t).$$

The first order condition requires taking the partial derivative and equating to 0: $\frac{\partial L(\beta_{t+1})}{\partial \beta_{t+1}} = g_t + H_t(\beta_{t+1} - \beta_t) = 0$.

Newton Raphson Algorithm

Solving for β_{t+1} yields: $\beta_{t+1} = \beta_t + (-H_t^{-1})g_t$. This is the Newton-Raphson formula for updating β.

The step from the current value of β to the new value is $(-H_t)^{-1}g_t$ i.e., gradient vector pre-multiplied by the negative of inverse of Hessian.

In the BHHH algorithm, the Hessian is calculated as outer product of the gradient i.e., $H_t = g_t g_t'$.

Stopping rule: The difference $|L(\beta_{t+1}) - L(\beta_t)| < \varepsilon$, where ε is a small positive number.

Covariate Effects

The link functions for probit and logit models are non-linear. So, the β's do not give the covariate effects.

Covariate effects for a continuous variable x_j can be calculated as

$$\text{Probit:} \quad \frac{\partial P_i}{\partial x_{ij}} = \frac{\partial \Phi(\beta_1 + x_{i2}\beta_2 + \cdots + x_{ik}\beta_k)}{\partial x_{ij}} = \beta_j \phi(x_i'\beta),$$

$$\text{Logit :} \quad \frac{\partial P_i}{\partial x_{ij}} = \frac{\partial \Lambda(\beta_1 + x_{i2}\beta_2 + \cdots + x_{ik}\beta_k)}{\partial x_{ij}} = \beta_j \Lambda'(x_i'\beta),$$

where $P_i = P(y_i = 1 | x_i'\beta)$.

Covariate Effects

Covariate effects for an indicator variable x_m are calculated as,

$$\Pr(y_i = 1 | x_{i,-m}, x_{i,m} = 1) - \Pr(y_i = 1 | x_{i,-m}, x_{i,m} = 0),$$

where the probability of success can be calculated using the cumulative distribution function of the assumed distribution and $x_{i,-m}$ denotes all the variables except the m-th variable.

Model Selection

To compare models, one may compute the Akaike information criterion (AIC) and Bayesian information criterion (BIC).

$$\begin{aligned} \text{log-like} &= \log \ell(\hat{\theta}; y) \\ \text{AIC} &= -2 \,\text{log-like} + 2k, \\ \text{BIC} &= -2 \,\text{log-like} + k \log(n), \end{aligned}$$

where k is the number of estimated model parameters and n is the number of observations. Both measures are based on information theoretic approach.

Information theoretic (IT) approach: When a statistical model is used to represent the process that generated the data, the representation is typically not exact and some information is lost. IT measures (e.g., AIC and BIC) estimate the relative amount of information lost by a given model: the less information a model loses, higher the quality of that model.

In estimating the amount of information lost by a model, IT measures deal with the trade-off between the goodness of fit of the model and the simplicity of the model. In other words, AIC and BIC deals with both the risk of overfitting and the risk of underfitting.

LR Statistic

Some non-IT measures include the likelihood ratio (LR) statistic, McFadden's R-squared, and the hit-rate.

The LR test statistic λ_{LR} for $H_0 : \beta_2 = \ldots = \beta_k = 0$, is:

$$\lambda_{LR} = -2[\ln L_0 - \ln L_{\text{fit}}] \quad \overset{H_0}{\sim} \quad \chi^2_{k-1},$$

where $\ln L_{\text{fit}}$ ($\ln L_0$) is the log-likelihood of fitted (intercept-only) model. Under H_0, λ_{LR} follows a chi-square distribution with *d.f.* equaling $k-1$, i.e., the number of restrictions under H_0.

We calculate λ_{LR} and compare it with χ^2_{k-1} for a given level of significance. If $\lambda_{LR} > \chi^2_{k-1}$, then we reject H_0; else we fail to reject H_0.

McFadden's R-squared

The McFadden's R-squared (pseudo R-squared or likelihood ratio index) is defined as follows,

$$R_M^2 = 1 - \frac{\ln L_{\text{fit}}}{\ln L_0}. \tag{6}$$

The R_M^2 is bounded between 0 and 1, similar to the coefficient of determination (R^2) in linear regression models.

When all slope coefficients are zero, the R_M^2 equals zero; but in discrete choice models R_M^2 can never equal 1, although it can come close to 1. Higher values of R_M^2 implies better fit, but it has no natural interpretation like R^2 that denotes the proportion of variation in the dependent variable explained by the covariates.

Hit-rate

The hit-rate (HR) is defined as the percentage of correct predictions i.e., percentage of observations for which the model correctly assigns the highest probability to the observed response category. Mathematically, the HR is defined as,

$$HR = \frac{1}{n} \sum_{i=1}^{n} I\left(\left(\max_{j} \{\hat{p}_{ij}\}_{j=1}^{J} \right) = y_i \right), \tag{7}$$

where $\hat{p}_{ij}$ is the predicted probability that individual i selects outcome j, and $I(\cdot)$ is the indicator function as defined earlier.

Application:
Mortgage Lending in Boston

Mortgage Data

The Boston HMDA data set was collected by researchers at the Federal Reserve Bank of Boston. The data set combines information from mortgage applications and a follow-up survey of the banks and other lending institutions that received these mortgage applications.

The data pertain to mortgage applications made in 1990 in the greater Boston metropolitan area. The full data set has 2925 observations, consisting of all mortgage applications by blacks and Hispanics plus a random sample of mortgage applications by whites.

Mortgage Data

To narrow the scope of the analysis, we use a subset of the data for single-family residences only (thereby excluding data on multi-family homes) and for black and white applicants only (thereby excluding data on applicants from other minority groups). This leaves us with 2380 observations.

The description and summary of the data is provided in Table 1.

Data Summary

Table 1: Variable definitions and data summary.

VARIABLES	DESCRIPTION	MEAN/COUNT	STD/PROP
`deny`	mortgage application, = 1 if denied, 0 otherwise	285	0.12
`pirat`	payments-to-income ratio	0.33	0.12
`hirat`	inhouse expense-to-total-income ratio	0.25	0.01
`lvrat`	loan-to-value ratio	0.74	0.18
`chist`	consumer credit score, values = 1 to 6		
`mhist`	mortgage credit score, values = 1, 2		
`phist`	public bad credit record, = 1 if yes, 0 otherwise	175	0.07
`unemp`	unemployment rate	3.77	2.03
`selfemp`	self employed, = 1 if yes, 0 otherwise	277	0.12
`insurance`	denied mortgage insurance, = 1 if yes, 0 otherwise	48	0.02
`condomin`	condominium/single family residence, = 1 if yes, 0 otherwise	686	0.29
`afam (black)`	family is black, = 1 if yes, 0 otherwise	339	0.14
`single`	marital status is single, = 1 if yes, 0 otherwise	936	0.39
`hschool`	high school and above, = 1 if yes, 0 otherwise	2341	0.98

Linear Probability Model (LPM): Model 1

We are interested in modeling the binary variable "`deny`", an indicator for whether an applicant's mortgage application has been denied (`deny` = yes) or accepted (`deny` = no).

A regressor that ought to have power in explaining whether a mortgage application has been denied is "`pirat`", the size of the anticipated total monthly loan payments relative to the applicants income.

As a first step, we ignore the binary nature of the variable `deny` and estimate the model as a linear regression model:

$$\texttt{Deny} = \beta_1 + \beta_2 * \texttt{(P/I) ratio} + \varepsilon$$

This is known as linear probability model.

LPM: Model 1

```
# load ‘AER‘ package and attach HMDA
datalibrary(AER)
library(ggplot2)
library("writexl")

data(HMDA) # inspect the data
head(HMDA)
summary(HMDA)

# convert 'deny' to numeric
HMDA$deny <- as.numeric(HMDA$deny) - 1

# estimate a simple linear regression,
# aka, linear probabilty model (LPM)

denymod1 <- lm(deny ~ pirat, data = HMDA)
summary(denymod1)
```

LPM: Model 1 Output

```
Call:lm(formula = deny ~ pirat, data = HMDA)
Residuals:
Min       1Q        Median    3Q       Max
-0.73070 -0.13736  -0.11322  -0.07097  1.05577

Coefficients:
             Estimate   Std. Error  t value  Pr(>|t|)
(Intercept) -0.07991    0.02116    -3.777    0.000163***
pirat        0.60353    0.06084     9.920    < 2e-16***
---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3183 on 2378 degrees of freedom
Multiple R-squared:  0.03974,Adjusted R-squared:  0.03933
F-statistic: 98.41 on 1 and 2378 DF,  p-value: < 2.2e-16
```

LPM: Model 1

Table 2: Estimates from linear probability model.

	Coef	Std Err	t-stat
Intercept	−0.080	0.021	−3.777
pirat	0.603	0.061	9.920

Residual standard error $\hat{\sigma}$: 0.3183 on 2378 degrees of freedom. R-squared: 0.03974, Adj. R-squared: 0.03933. F-statistic: 98.41 on 1 and 2378 DF, p-value: $<$ 2.2e-16.

As per the estimated model, there is a positive relation between `(P/I) ratio` and the probability of a denied mortgage application. A 1 percentage point increase in `(P/I) ratio` leads to an increase in the probability of a loan denial by $0.604 \times 0.01 = 0.00604 \simeq 0.6\%$.

Scatter Plot and Regression Line

```
# plot the data
plot(x = HMDA$pirat,
 y = HMDA$deny,
 main = "Scatterplot and Estimated Regression Line",
 xlab = "P/I ratio",
 ylab = "Deny",
 pch = 20,
 ylim = c(-0.4, 1.4),
 cex.main = 0.8)

# add horizontal dashed lines and text
abline(h = 1, lty = 2, col = "darkred")
abline(h = 0, lty = 2, col = "darkred")
text(2.5, 0.9, cex = 0.8, "Mortgage denied")
text(2.5, -0.1, cex= 0.8, "Mortgage approved")
# add the estimated regression line
abline(denymod1, lwd = 1.8, col = "steelblue")
text(1.5, 0.5, cex= 0.8, "deny.hat = -0.07991 + 0.60353*(P/I)")
```

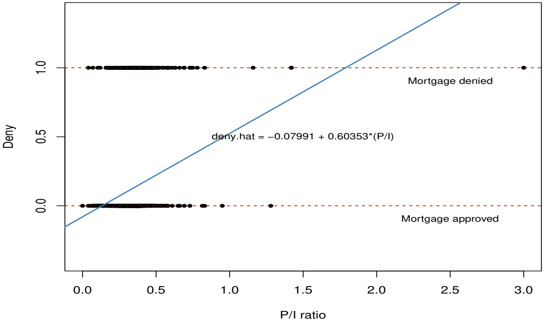

Figure 3: Scatter plot and estimated regression line.

According to the estimated model, a payment-to-income ratio of 1 is associated with an expected probability of mortgage application denial of roughly 50%.

Is there racial bias?

We augment our previous model by an additional regressor `black` which equals 1 if the applicant is an African American and equals 0 otherwise.

Such a specification is the baseline for investigating if there is racial discrimination in the mortgage market: if being black has a significant (positive) influence on the probability of a loan denial when we control for factors that allow for an objective assessment of an applicant's creditworthiness. This is an indicator for discrimination.

LPM: Model 2

```
# rename the variable 'afam' for consistency
colnames(HMDA)[colnames(HMDA) == "afam"] <- "black"

# estimate the model
denymod2 <- lm(deny ~ pirat + black, data = HMDA)
coeftest(denymod2, vcov. = vcovHC)

t test of coefficients:
             Estimate   Std. Error   t value   Pr(>|t|)
(Intercept) -0.090514   0.033430    -2.7076    0.006826**
pirat        0.559195   0.103671     5.3939    7.575e-08***
blackyes     0.177428   0.025055     7.0815    1.871e-12

***---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

LPM: Model 2

Table 3: Estimates from linear probability model.

	Coef	Std Err	t-stat
Intercept	−0.090	0.033	−2.708
pirat	0.559	0.104	5.394
black	0.177	0.025	7.081

Residual standard error $\hat{\sigma}$: 0.3123 on 2377 degrees of freedom. R-squared: 0.076, Adj. R-squared: 0.07523. F-statistic: 97.76 on 2 and 2377 DF, p-value: $< 2.2e\text{-}16$.

The intercept, `pirat`, and `black` are statistically siginificant at 5% level since the absolute value of `t-statistic` is greater than 2 for all the variables.

LPM: Model 2

The coefficient of black is positive and significantly different from zero at the 0.01% level. The interpretation is that, holding constant the `P/I ratio`, being `black` increases the probability of a mortgage denial by about 17.7 percentage points.

This finding is compatible with racial discrimination. However, it might be distorted by omitted variable bias so discrimination could be a premature conclusion.

The linear probability model (LPM) has a major flaw, it predicts $P(y = 1|x_1, \ldots, x_k)$ to lie outside 0 and 1. However probabilities cannot lie between 0 and 1. So, the LPM is not appropriate to model binary responses.

Probit Model 1

We now estimate a simple Probit model of the probability of a mortgage denial.

```
# estimate a simple probit model

denyprobit <- glm(deny ~ pirat, family = binomial(link = "probit"),
data = HMDA)
coeftest(denyprobit, vcov. = vcovHC, type = "HC1")

# Output
z test of coefficients:
             Estimate   Std. Error   z value   Pr(>|z|)
(Intercept) -2.19415    0.18901     -11.6087   < 2.2e-16***
pirat        2.96787    0.53698       5.5269   3.259e-08***

---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Probit Model 1

Table 4: Estimates from Probit regression model.

	Coef	Std Err	z-stat
Intercept	−2.194	0.138	−15.927
pirat	2.968	0.386	7.694

Null deviance: 1744.2 on 2379 degrees of freedom. Residual deviance: 1663.6 on 2378 degrees of freedom. AIC: 1667.6

The coefficient for `P/I ratio` is significant, which implies that applicants with a high P/I ratio face a higher risk of being rejected.

Probit Model 1: Scatter Plot and Regression Line

```
# plot dataplot(x = HMDA$pirat, y = HMDA$deny,
main = "Probit Model of the Probability of Denial, given P/I Ratio",
xlab = "P/I ratio", ylab = "Deny", pch = 20,
ylim = c(-0.4, 1.4), cex.main = 0.85)

# add horizontal dashed lines and text
abline(h = 1, lty = 2, col = "darkred")
abline(h = 0, lty = 2, col = "darkred")
text(2.5, 0.9, cex = 0.8, "Mortgage denied")
text(2.5, -0.1, cex= 0.8, "Mortgage approved")
# add estimated regression line
x <- seq(0, 3, 0.01)
y <- predict(denyprobit, list(pirat = x), type = "response")
lines(x, y, lwd = 1.5, col = "steelblue")
```

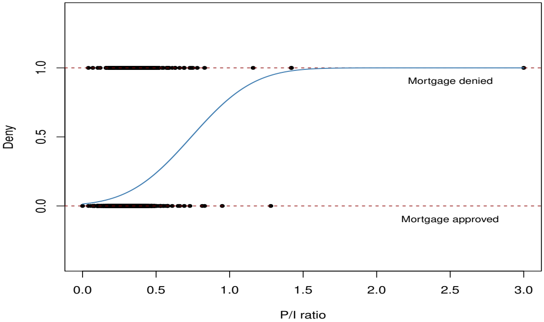

Figure 4: Probit Model of the Probability of Denial, given P/I Ratio

The estimated regression function has a stretched “S”-shape which is typical for the CDF of a continuous random variable with symmetric PDF like that of a normal random variable.

The function is nonlinear and flattens out for large and small values of P/I ratio. The functional form thus also ensures that the predicted conditional probabilities of a denial lie between 0 and 1.

Covariate Effect

Probit (and Logit) are non-linear models. So, the coefficients do not give the covariate effect and we need to compute it.

Let's suppose `P/I ratio` is increased from 0.3 to 0.4. What is the predicted change in the denial probability? This change in probability is computed as follows:

$$\Phi(-2.194 + 2.968 * 0.4) - \Phi(-2.194 + 2.968 * 0.3) = 0.061.$$

We find that an increase in the payment-to-income ratio from 0.3 to 0.4 is predicted to increase the probability of denial by approximately 6.08%.

Probit Model 1: Covariate Effect

```
# 1. compute predictions for P/I ratio = 0.3, 0.4
predictions1 <- predict(denyprobit,
newdata = data.frame("pirat" = c(0.3, 0.4)), type = "response")

# 2. Compute difference in probabilities
diff(predictions1)

# Output
0.06081433
```

Probit Model 2: Is there racial bias?

We continue by using an augmented Probit model to estimate the effect of race on the probability of a mortgage application denial.

```
denyprobit2 <- glm(deny ~ pirat + black,
family = binomial(link = "probit"), data = HMDA)
coeftest(denyprobit2, vcov. = vcovHC, type = "HC1")

# Output
z test of coefficients:
             Estimate   Std. Error   z value  Pr(>|z|)
(Intercept) -2.258787   0.176608    -12.7898  < 2.2e-16***
pirat        2.741779   0.497673      5.5092  3.605e-08***
blackyes     0.708155   0.083091      8.5227  < 2.2e-16***

---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Probit Model 2: Is there racial bias?

Table 5: Estimates from Probit regression model.

	Coef	Std Err	z-stat
Intercept	−2.259	0.137	−16.525
pirat	2.742	0.380	7.206
black	0.709	0.083	8.496

Null deviance: 1744.2 on 2379 degrees of freedom. Residual deviance: 1594.3 on 2377 degrees of freedom. AIC: 1600.3

```
# Estimated model:
P(deny=1|x's) = \Phi(-2.259 + 2.742*(P/I) ratio + 0.709*black)
```

Probit Model 2

All coefficients are statistically significant. The coefficient for African Americans is positive and indicates that African Americans have a higher probability of denial than White Americans, ceteris paribus. Also, applicants with a high P/I ratio face a higher risk of being rejected.

Question: How big is the estimated difference in denial probabilities between two hypothetical applicants with the same payments-to-income ratio?

Probit Model 2: Covariate Effect

We continue by using an augmented Probit model to estimate the effect of race on the probability of a mortgage application denial.

```
predictions2 <- predict(denyprobit2, newdata
= data.frame("black" = c("no", "yes"), "pirat" = c(0.3,0.3)),
type = "response")

# 2. compute difference in probabilities
diff(predictions2)

# Output
0.1578117
```

We find that the white applicant faces a denial probability of only 7.546%, while the African American is rejected with a probability of 23.327%, a difference of 15.781 percentage points.

Logit Model

As for Probit regression, there is no simple interpretation of the model coefficients and it is best to consider predicted probabilities or differences in predicted probabilities.

However, for the logit model the coefficients represent the log-odds (or logarithm of odds ratio) of success. As such, it is preferred in many applications.

We now analyze the mortgage lending problem using the Logit model.

Logit Model 1

We continue by using an augmented Probit model to estimate the effect of race on the probability of a mortgage application denial.

```
# Estimating a logit model
denylogit <- glm(deny ~ pirat, family = binomial(link =
"logit"), data = HMDA)
coeftest(denylogit, vcov. = vcovHC, type = "HC1")

# Output
z test of coefficients:
             Estimate   Std. Error   z value   Pr(>|z|)
(Intercept) -4.02843    0.35898    -11.2218    < 2.2e-16***
pirat        5.88450    1.00015      5.8836    4.014e-09 ***

---Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Logit Model 1

Table 6: Estimates from Logit regression model.

	Coef	Std Err	z-stat
Intercept	−4.028	0.269	−15.000
pirat	5.884	0.734	8.021

Null deviance: 1744.2 on 2379 degrees of freedom. Residual deviance: 1660.2 on 2378 degrees of freedom. AIC: 1664.2

Both models produce very similar estimates of the probability that a mortgage application will be denied depending on the applicants payment-to-income ratio. This is shown in Figure 5 in the next slide.

Probit & Logit: Probability of Denial

```
plot(x = HMDA$pirat, y = HMDA$deny,
#main = "Probit and Logit Models of the Probability of Denial,
Given P/I Ratio", xlab = "P/I ratio", ylab = "Deny",
pch = 20,     ylim = c(-0.4, 1.4), cex.main = 0.9)

# add horizontal dashed lines and text
abline(h = 1, lty = 2, col = "darkred")
abline(h = 0, lty = 2, col = "darkred")
text(2.5, 0.9, cex = 0.8, "Mortgage denied")
text(2.5, -0.1, cex= 0.8, "Mortgage approved")
```

Probit & Logit: Probability of Denial

```
# add estimated regression line of Probit and Logit models
x <- seq(0, 3, 0.01)
y_probit <- predict(denyprobit, list(pirat = x),
type = "response")
y_logit <- predict(denylogit, list(pirat = x), type = "response")
lines(x, y_probit, lwd = 1.5, col = "steelblue")
lines(x, y_logit, lwd = 1.5, col = "black", lty = 2)

# add a legend
legend("topleft", horiz = TRUE,  legend = c("Probit",
"Logit"), col = c("steelblue", "black"), lty = c(1, 2))
```

Probit & Logit: Probability of Denial

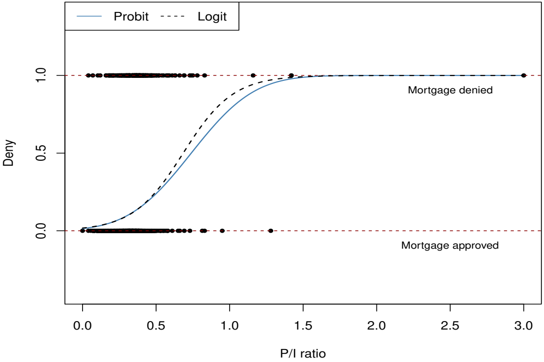

Figure 5: Probit and Logit models of the probability of denial, given the P/I ratio.

Logit Model 2: Is there racial bias? Revisited

We now extend the Logit model of mortgage denial with the additional regressor black.

```
# estimate a Logit regression with multiple regressors
denylogit2 <- glm(deny ~ pirat + black, family
= binomial(link = "logit"), data = HMDA)
coeftest(denylogit2, vcov. = vcovHC, type = "HC1")

# Output
z test of coefficients:
             Estimate   Std. Error   z value   Pr(>|z|)
(Intercept) -4.12556    0.34597     -11.9245   < 2.2e-16***
pirat        5.37036    0.96376       5.5723   2.514e-08***
blackyes     1.27278    0.14616       8.7081   < 2.2e-16***
---Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Logit Model 2

Table 7: Estimates from Logit regression model.

	Coef	Std Err	z-stat
Intercept	−4.126	0.268	−15.370
pirat	5.370	0.728	7.374
black	1.279	0.146	8.706

Null deviance: 1744.2 on 2379 degrees of freedom. Residual deviance: 1591.4 on 2377 degrees of freedom. AIC: 1597.4

```
# Estimated model:
P(deny=1|x's) = F(-4.126 + 5.370*(P/I) ratio + 1.279*black)
```

Logit Model 2

Similar to the Probit model, all model coefficients are highly significant and we obtain positive estimates for the coefficients on `P/I ratio` and `black`.

Question: How big is the estimated difference in denial probabilities between two hypothetical applicants with the same payments-to-income ratio?

Logit Model 2

```
predictions3 <- predict(denylogit2, newdata =
data.frame("black" = c("no", "yes"),"pirat"=c(0.3,0.3)),
type = "response")

predictions3

# 2. Compute difference in probabilities
diff(predictions3)

# Output
0.1492945
```

We find that the white applicant faces a denial probability of only 7.485%, while the African American is rejected with a probability of 22.414%, a difference of 14.929 percentage points.

Model Comparison

We compute the McFadden's R-squared or pseudo R-squared for the Probit and Logit models augmented with the additional regressor `black`.

```
# compute the null Probit model
denyprobit_null <- glm(formula = deny ~ 1, family
= binomial(link = "probit"), data = HMDA)

# compute the pseudo-R2 using 'logLik'
1 - logLik(denyprobit2)[1]/logLik(denyprobit_null)[1] #> [1]

# Output
0.08594259
```

Model Comparison

```
# compute the null Logit model
denylogit_null <- glm(formula = deny ~ 1, family
= binomial(link = "logit"), data = HMDA)

# compute the pseudo-R2 using 'logLik'
1 - logLik(denylogit2)[1]/logLik(denylogit_null)[1]

# Output
0.08759475
```

The pseudo R-squared for the Probit model is 0.0859, while that for the Logit model is 0.0875. So, the Logit model provides a marginally better fit. This is also evident from a lower AIC value (1597.4 vs 1600.3) for the Logit model.

Thank you!