Binary Choice Models

Afees A. Salisu

Centre for Econometric & Allied Research (CEAR)

University of Ibadan

adebare1@yahoo.com

08034711769

Outline

- Introduction
- Logit Regressions
  - Specification & Assumptions
  - Estimation
- Probit Regressions
  - Specification & Assumptions
  - Estimation
- Comparing Logit & Probit Coefficients
- Diagnostics

Introduction

- Consider a linear regression model of the type:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$$

- Where **Y is continuous** while the X's can be purely continuous, dichotomous (dummy variables) or both.
- Recall that a **continuous variable** is one that can take any value between two numbers.

Intro...

- Examples of continuous variables include interest rate, inflation rate, weight, height, gdp, money supply, exchange rate, etc.
- In general, most economic and financial variables are continuous variables.

Intro...

- If the specified model is linear, so far the dependent variable is continuous, we can estimate with a linear estimator such as the Ordinary Least Squares (OLS) and do model checking, visualize results, etc.

Intro...

- ❑If the model is non-linear (but the dependent variable is still continuous), we can transform or add variables to get the equation to be linear.
- ❑For instance,
- We can take logs of Y and/or the X's depending on the source(s) of non-linearity.
- We can also add squared terms as well as interaction terms

Intro...

- However, it is also possible to have a **dichotomous** dependent variable.
- Dichotomous variables are nominal **variables** which have only two categories or levels. In econometrics, dummy variables are used to represent dichotomous variables. Why?

Intro...

- A **dummy variable** is one that takes the value 0 or 1.
- **Why is it used?** Regression analysis treats all independent (X) **variables** in the analysis as **numerical** but most dichotomous (and in fact categorical) variables are **non-numerical** (involving the use of alphabets).

Intro...

- The use of dummy variables is prominent in Social Sciences/Arts & Humanities where human behaviour is studied.
- Examples of binary outcomes:

➢ Modelling Labour Force Participation: YES/NO

The respondent here can either say "yes" if he/she is in the labour force or "no" if he/she is not. Thus, we can assign "1" to "yes" and "0" to "no".

Intro...

- Modelling the 2015 presidential election in Nigeria. Here, we may be interested in the factors that influence whether a political candidate wins the presidential election or not. The outcome (response) variable is binary (0/1); win or lose.

Intro...

- Modelling the Demand for Satellite TV Facilities: The dependent variable here represents the decision of the household to subscribe for satellite facilities or not. Thus, the response variable is “1”if the household subscribes and “0” other wise.

Intro...

- Modelling student academic performance in econometrics. The outcome (response) variable is binary (0/1); pass or fail.
- Modelling admission into graduate school. The outcome (response) variable is binary (0/1); admit or don't admit.

Intro...

- In general, when dealing with a dichotomous dependent variable (Y), it is your responsibility to convert the non-numerical information to binary.

Intro...

- The following terms are used interchangeably to describe models with binary dependent variables.

➢ Qualitative Response Models (QRM)

➢ Discrete Response Models (DRM)

➢ Binary Choice Models (BCM)

➢ Binary Response Models (BRM)

Intro...

- There are three essential features of a binary choice model:

➢ The dependent variable is qualitative in nature

➢ The response is binary in nature

➢ It involves non-linear estimation. Why?

Intro...

- The linear estimators rely on normality assumption. However, binary response models are non-normal and therefore, the linear estimators such as OLS may not applicable particularly when dealing with small samples.

Intro...

- The error terms of binary response models tend to exhibit heteroscedasticity . Thus, the application of OLS will bias the standard errors and hence inferential statistics using the standard errors such as the t-values will be invalid.

Intro...

- Binary response models are usually expressed as linear functions of a set of regressors. The estimates of Y given X are conditional probabilities of the event Y occurring (i.e. When it is 1). Therefore, the conditional probabilities are expected to lie between 0 and 1. However, if OLS is used, conditional probabilities are more likely to lie outside the (0,1) range.

Intro...

- As we shall see in subsequent slides, the Probit and Logit models can be employed to resolve the highlighted problems inherent in the use of linear estimators.

Generalization

- The traditional way of introducing probits and logits in econometrics is not as a response to a functional problem. Instead, probits and logits are traditionally viewed as models suitable for estimating parameters of interest when the dependent variable is not fully observed.
- This we illustrate as follows:

Generalization

- Let y^* be a continuous variable that we do not observe – a latent variable – and assume y^* is determined by the model:

$$y^* = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k + e$$

$$y^* = X\alpha + e$$

Generalization

- where e is a residual term, assumed uncorrelated with X (i.e. X is not endogenous). While we do not observe y^*, we do observe the discrete choice made by the individual, according to the following choice rule:

$$y = 1 \text{ if } \; y^* > 0$$

$$y = 0 \text{ if } \; y^* \leq 0$$

Generalization

- We can view y^* as representing net utility of, say, owning a house.
- The individual undertakes a cost-benefit analysis and decides to purchase a house if the net utility is positive.
- We do not observe the amount of net utility; all we observe is the actual outcome of whether or not the individual does buy a house

Generalization

- In fact, if we had data on y^*, we could estimate the latent variable model with OLS as usual.
- We can now model the probability that a 'positive' choice is made (e.g. buying as distinct from not buying a house).

Generalization

- We can show that:

$$\Pr(y = 1|X) = \Pr(y^* > 0|X)$$

$$\Pr(y = 0|X) = \Pr(y^* \leq 0|X)$$

$$= 1 - \Pr(y = 1|X)$$

The error term (e) can follow either a logistic distribution or standard normal distribution depending on the choice of probability model being estimated.

Generalization

The logistic distribution is for Logit models

While standard normal distribution is for Probit models

The Logistic Regression

The Logit ...

- Consider a class of BRM of the form:

$$\Pr\left(Y=1 \middle| X\right)=G\left(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_k X_k\right)$$

- The equation above defines the conditional probabilities of Y=1 (i.e. Y occurring) given X.
- For a more compact representation:

$$\Pr\left(y=1 \middle| X\right)=G\left(X\beta\right)$$

The Logit ...

- In the Logit model, $\Lambda(X\beta)$ is given as:

$$\Lambda(X\beta) = \frac{\exp(X\beta)}{1+\exp(X\beta)}$$

- The equation above is the cumulative (logistic) distribution function (cdf) and it ranges between zero and one for all values of $X\beta$.

The Logit ...

- Note the following about the equation:
  - Λ is a non-linear function of $X\beta$ and hence, we cannot use OLS.
  - The errors follow standard logistic distribution.
  - The estimator is the Maximum Likelihood estimator.

The Logit ...

- When dealing with Logistic regressions, the following parameters are usually estimated:

➢ Odds ratios

➢ Log odds

➢ Marginal effects & Conditional Probability

- Let us now take each in turn.

The Logit ...

- **Odds Ratio**: It is the ratio of probability of Y=1 to the probability that Y=0.
- This is given as:

$$\frac{\Lambda\left(X\beta\right)}{1-\Lambda\left(X\beta\right)}=\exp\left(X\beta\right)$$

- How?
- Note that the numerator denotes the prob. that Y=1 and the denominator is for the prob. that Y=0.

The Logit ...

- Recall: $$\Lambda(X\beta)=\frac{\exp(X\beta)}{1+\exp(X\beta)}$$
- As a consequence,
$$1-\Lambda(X\beta)=\frac{1}{1+\exp(X\beta)}$$
- Given these representations, it is easy to show that:
$$\frac{\Lambda(X\beta)}{1-\Lambda(X\beta)}=\exp(X\beta)$$

The Logit ...

- Let us consider a simple illustration.

The Logit ...

- The Problem:
- A researcher is interested in how variables, such as GPA (grade point average) and GQES (graduate qualifying exam scores) affect admission into graduate school (GSA).
- The response variable (GSA), admit/don't admit, is a binary variable.

The Logit ...

- The model can be expressed as:

$$\Pr\left(\text{GSA}=1 \mid \text{GQES},\text{GPA}\right)=\frac{\exp\left(\beta_0+\beta_1\text{GQES}+\beta_2\text{GPA}\right)}{1+\exp\left(\beta_0+\beta_1\text{GQES}+\beta_2\text{GPA}\right)}$$

- If after estimation we obtain the following for the relevant parameters:

$$\hat{\beta}_0=-2.3868$$
$$\hat{\beta}_1=0.0014$$
$$\hat{\beta}_2=0.4777$$

- Not to worry, the estimation process will be demonstrated later using Stata software.

The Logit ...

- Interpreting the odds ratios :
- For GQES: The results show that for a unit increase in GQES, the odds ratio in favour of gaining admission into graduate school increases by 1.0014 [i.e. exp(0.0014)] or about 0.14%.
- The percent change is obtained by:

  [exp(0.0014)-1]* 100 %

The Logit ...

- For GPA: The results show that for a unit increase in GPA, the odds in favour of gaining admission into graduate school increases by 1.6124 [i.e. exp(0.4777) or about 61.24% [i.e. (exp(0.4777)-1)* 100 %]

The Logit ...

- **Log odds**: This is obtained by taking the natural log of the odds ratio. This gives:

$$L = \text{In}\left(\frac{\Lambda(X\beta)}{1-\Lambda(X\beta)}\right) = \ln(\exp(X\beta)) = X\beta$$

$$X\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

The Logit ...

- In relation to our example, the log odds can be expressed as:

$$L = X\beta = \beta_0 + \beta_1 \text{GQES} + \beta_2 \text{GPA}$$

- Given our earlier estimates, the equation becomes:

$$L = -2.3868 + 0.0014 * \text{GQES} + 0.4777 * \text{GPA}$$

The Logit …

- Interpretation of the log odds for the individual variables:
- For GQES: The results show that for a unit increase in GQES, the log odds in favour of gaining admission into graduate school increase by 0.0014.

The Logit ...

- For GPA: The results show that the log odds in favour of gaining admission into the graduate school increases by 0.48 with each additional unit increase in GPA.

The Logit ...

- Note the following about the Odds Ratio and the Log Odds Ratio:

➢ Log odds = log (odds ratio)= $X\hat{\beta}$

➢ Odds ratio = exp (log (odds ratio))= $\exp\left(X\hat{\beta}\right)$

- Check our illustration carefully, the estimates reported for all the coefficients are in favour of log odds hence, the need to exponentiate the coefficients.

The Logit ...

- Also, note that the interpretation of the log odds was drawn directly from the reported estimates without any further transformation.
- On the other hand, the odds ratios were obtained from the log odds by exponentiating the latter to compute the odds ratio.
- Kindly check to confirm.

The Logit ...

- Nonetheless, you may decide to start with the Odds ratio but you have to be mindful of the required transformation.
  - Again, not to worry, there are commands in Stata that can be used to generate all the relevant statistics.
  - However, understanding the underlying theory is as important as knowing how to operate the statistical softwares.

The Logit ...

- **Marginal Effect:** It measures the change in the probability of Y=1 as a result of a unit change in a particular explanatory variable.
- In the case of logit models however, obtaining the marginal effects is more complicated.
- Recall that:

$$\Lambda(X\beta) = \frac{\exp(X\beta)}{1+\exp(X\beta)}$$

The Logit ...

- The marginal effect is computed as:

$$\frac{\partial\Lambda(X\beta)}{\partial X} = \beta^* \Lambda'(X\beta) = \beta^* \Lambda(X\beta)\left(1-\Lambda(X\beta)\right)$$

- where $\Lambda(X\beta) = \dfrac{\exp(X\beta)}{1+\exp(X\beta)}$

- Let us continue with our illustration.

The Logit ...

- Marginal Effect for GQES

$$= \beta_1 * \Lambda\left(X\beta\right)\left(1 - \Lambda\left(X\beta\right)\right)$$

$$\Lambda\left(X\beta\right) = \frac{\exp\left(\beta_0 + \beta_1 \mathrm{GQES} + \beta_2 \mathrm{GPA}\right)}{1 + \exp\left(\beta_0 + \beta_1 \mathrm{GQES} + \beta_2 \mathrm{GPA}\right)}$$

- The $\Lambda(X\beta)$ gives the probability value of being admitted into graduate school evaluated at the mean values of the regressors.

The Logit ...

- Let us assume that $\overline{\text{GPA}} = 3.3899$ & $\overline{\text{GQES}} = 587.7$.

$$\exp\left(X\hat{\beta}\right) = \exp\left(-2.3868 + 0.0014 * 587.7 + 0.4777 * 3.3899\right)$$

$$\exp\left(X\hat{\beta}\right) = \exp\left(0.0553\right) = 1.0569$$

- Thus,

$$\Lambda\left(X\beta\right) = \frac{1.0569}{1 + 1.0569} = 0.514$$

The Logit ...

- The results indicate that students with GQES & GPA of approximately 600 and 4.0 respectively have about 51% probability (fairly above average chance) of being admitted into the graduate school.

- The ME for GQES = 0.0014*0.514[1-0.514]

  = 0.0003

The Logit ...

- Interpretation:
- A unit increase in GQES will increase the probability of gaining admission into the graduate school by 0.0003 or 0.03%.

- Class Exercise 1:
- Compute the Marginal effect for GPA and interpret your results.

The Logit ...

- Dealing with Binary regressors in logit models:
- We can extend our earlier example to include a dummy regressor such as the prestige of the undergraduate institutions attended by the applicants.
- INST =1 if a student attended one of the top-ranked universities and zero other wise.

The Logit ...

- The extended equation is given as:

$$\Pr\left(\text{GSA}=1 \middle| \text{GQES}, \text{GPA}, \text{INST}\right)$$

$$= \Lambda\left(X\beta\right) = \Lambda\left(\beta_0 + \beta_1\text{GQES} + \beta_2\text{GPA} + \beta_3\text{INST}\right)$$

$$= \frac{\exp\left(\beta_0 + \beta_1\text{GQES} + \beta_2\text{GPA} + \beta_3\text{INST}\right)}{1+\exp\left(\beta_0 + \beta_1\text{GQES} + \beta_2\text{GPA} + \beta_3\text{INST}\right)}$$

- Recall that GQES and GPA are continuous variables while INST is a discrete (binary) variable.

The Logit ...

- Our focus now is on INST.
- Log Odds:
- Let us assume that the coefficient for the INST is 0.3523 after estimating the log odds function.
- The log odds in favour of gaining admission into the graduate school is 0.3523 higher for students from top-ranked universities relative to those from other universities.

The Logit ...

- Odds Ratio:
- The odds ratio is exp(0.3523) = 1.4223. This suggests that students from top-ranked universities are 1.4223 times (about 42.23%) more likely to gain admission into the graduate school than students from other universities.

The Logit ...

- Marginal Effect for a Binary regressor (INST):

$$\Lambda\left(\hat{\beta}_0 + \hat{\beta}_1 \overline{\text{GQES}} + \hat{\beta}_2 \overline{\text{GPA}} + \hat{\beta}_3 (1)\right)$$

$$-\Lambda\left(\hat{\beta}_0 + \hat{\beta}_1 \overline{\text{GQES}} + \hat{\beta}_2 \overline{\text{GPA}} + \hat{\beta}_3 (0)\right)$$

- An Empirical Application of Logistic Models using Stata

The Logit ....

- The Problem:
- We want to examine whether a new method of teaching economics significantly influenced performance in later economics courses.

The Logit …

- Variables are Grade, PSI, GPA & TUCE.

➢ Grade (Dependent variable) indicates whether a student improved his/her grades after the new teaching method PSI had been introduced (0 = no, 1 = yes).

- PSI indicates if a student attended courses that used the new method (0 = no, 1 = yes).

➢ GPA (Average grade of the student)

➢ TUCE is the Score of an intermediate test which shows previous knowledge of a topic.

The Logit …

- Data:

S/N	grade	psi	tuce	gpa	S/N	grade	psi	tuce	gpa
1	0	0	20	2.66	17	0	0	25	2.75
2	0	0	22	2.89	18	0	0	19	2.83
3	0	0	24	3.28	19	0	1	23	3.12
4	0	0	12	2.92	20	1	1	25	3.16
5	1	0	21	4	21	0	1	22	2.06
6	0	0	17	2.86	22	1	1	28	3.62
7	0	0	17	2.76	23	0	1	14	2.89
8	0	0	21	2.87	24	0	1	26	3.51
9	0	0	25	3.03	25	1	1	24	3.54
10	1	0	29	3.92	26	1	1	27	2.83
11	0	0	20	2.63	27	1	1	17	3.39
12	0	0	23	3.32	28	0	1	24	2.67
13	0	0	23	3.57	29	1	1	21	3.65
14	1	0	25	3.26	30	1	1	23	4
15	0	0	26	3.53	31	0	1	21	3.1
16	0	0	19	2.74	32	1	1	19	2.39

The Logit …

- Estimation Procedure:

❑ Descriptive Statistics

❑ Estimation

❑ Scenario Analyses

The Logit …

- Descriptive Statistics
- Step 1: Load your data into Stata
- Step 2: Use relevant commands in Stata to generate the summary statistics.
- Since we already have our data in .dta format, you may use stata command or menu approach to load the required data.

The Logit …

- Compute summary statistics for continuous variables

```
. su tuce gpa
```

Variable	Obs	Mean	Std. Dev.	Min	Max
tuce	32	21.9375	3.901509	12	29
gpa	32	3.117188	.4667128	2.06	4

The Logit…

- Highlights of the descriptive statistics:
  - The average score of the students in an intermediate test (TUCE) is about 22 their GPAs average 3.12.
  - We find that the variations in TUCE and GPA across the respondents seem minimal although the former is higher than the latter.

The Logit …

You are to tabulate discrete variables and not to summarize. Why?

```
. tab grade
```

grade	Freq.	Percent	Cum.
0	21	65.63	65.63
1	11	34.38	100.00
Total	32	100.00	

The Logit…

- Some highlights of the statistics:
  - About 34.4% (equivalent to 11) and 65.6% (equivalent to 21) of the respondents report high and low academic performance respectively.
  - In other words, more than half of the respondents report low academic performance.

The Logit …

```
. tab psi
```

psi	Freq.	Percent	Cum.
0	18	56.25	56.25
1	14	43.75	100.00
Total	32	100.00	

The Logit…

- Some highlights of the statistics:
  - About 43.7% (equivalent to 14) of the respondents were exposed to the new teaching method (PSI) while 56% (equivalent to 18) were not.
  - In other words, more than half of the respondents were not exposed to the new teaching method.

The logit …

```
. tab grade psi
```

grade	psi 0	1	Total
0	15	6	21
1	3	8	11
Total	18	14	32

The Logit…

- You can as well determine the distribution of tuce and gpa between high and low performance.

```
. su tuce gpa if grade==1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
tuce	11	23.54545	3.777926	17	29
gpa	11	3.432727	.503132	2.39	4

```
. su tuce gpa if grade==0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
tuce	21	21.09524	3.780275	12	26
gpa	21	2.951905	.3572201	2.06	3.57

The Logit …

- Estimation:
  - Odds Ratio
  - Log Odds
  - Marginal Effects
  - Some plausible scenarios

The Logit …

- Odds ratio:

```
. logistic grade psi tuce gpa
```

Logistic regression — Number of obs = 32; LR chi2(3) = 15.40; Prob > chi2 = 0.0015; Pseudo R2 = 0.3740

Log likelihood = -12.889633

grade	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
psi	10.79073	11.48743	2.23	0.025	1.339344	86.93802
tuce	1.099832	.1556859	0.67	0.501	.8333651	1.451502
gpa	16.87972	21.31809	2.24	0.025	1.420194	200.6239
_cons	2.21e-06	.0000109	-2.64	0.008	1.40e-10	.03487

The Logit…

- Highlights of the results:
- TUCE: The odds ratio in favour of high academic performance increases by 10% [i.e. (1.10-1)*100] with each additional unit increase in tuce.
- GPA: A unit increase in gpa will lead to 16.88 units (1588%) increase in the odds ratio in favour of high academic performance.
- PSI: Students with exposure to the new method are 10.79 times more likely to have high performance than students without exposure.

The Logit …

- Log Odds

```
. logit grade psi tuce gpa

Iteration 0:   log likelihood =  -20.59173
Iteration 1:   log likelihood = -13.259768
Iteration 2:   log likelihood = -12.894606
Iteration 3:   log likelihood = -12.889639
Iteration 4:   log likelihood = -12.889633
Iteration 5:   log likelihood = -12.889633

Logistic regression                               Number of obs   =         32
                                                  LR chi2(3)      =      15.40
                                                  Prob > chi2     =     0.0015
Log likelihood = -12.889633                       Pseudo R2       =     0.3740
```

grade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
psi	2.378688	1.064564	2.23	0.025	.29218	4.465195
tuce	.0951577	.1415542	0.67	0.501	-.1822835	.3725988
gpa	2.826113	1.262941	2.24	0.025	.3507938	5.301432
_cons	-13.02135	4.931325	-2.64	0.008	-22.68657	-3.35613

The Logit …

- Confirm the computational relationship between Odds Ratio and Log Odds from the results.
- What is your observation?

The Logit …

- Marginal Effects

```
. mfx compute

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .25282025
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.4564984	.18105	2.52	0.012	.10164	.811357	.4375
tuce	.0179755	.02624	0.69	0.493	-.033448	.069399	21.9375
gpa	.5338589	.23704	2.25	0.024	.069273	.998445	3.11719

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Logit …

- Some plausible scenarios for marginal effects:
- Case 1A: PSI=1, TUCE=20, GPA=4
- Case 1B: PSI=1, TUCE=20, GPA=2
- Case 2A: PSI=0, TUCE=20, GPA=4
- Case 2B: PSI=0, TUCE=20, GPA=2
- Case 3A: PSI=1, TUCE=20, GPA=4
- Case 3B: PSI=0, TUCE=20, GPA=4
- Case 4A: PSI=1, TUCE=20, GPA=2
- Case 4B: PSI=0, TUCE=20, GPA=2

The Logit …

- Case 1A: PSI=1, TUCE=20, GPA=4

```
. mfx, at(psi=1, tuce=20, gpa=4)

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .92857112
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
psi*	.382141	.25122	1.52	0.128	-.110233	.874515	1
tuce	.0063115	.01418	0.45	0.656	-.021483	.034106	20
gpa	.187447	.15883	1.18	0.238	-.123848	.498742	4

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Logit …

- Case 1B: PSI=1, TUCE=20, GPA=2

```
. mfx, at(psi=1, tuce=20, gpa=2)

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .04363495
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
psi*	.0394245	.05736	0.69	0.492	-.073006	.151855	1
tuce	.003971	.00802	0.50	0.620	-.01174	.019682	20
gpa	.1179364	.12099	0.97	0.330	-.1192	.355072	2

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Logit …

- Case 2A: PSI=0, TUCE=20, GPA=4

```
. mfx, at(psi=0, tuce=20, gpa=4)

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .54643009
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
psi*	.382141	.25122	1.52	0.128	-.110233	.874515	0
tuce	.0235843	.0366	0.64	0.519	-.048141	.09531	20
gpa	.7004358	.26442	2.65	0.008	.182177	1.21869	4

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Logit …

- Case 2B: PSI=0, TUCE=20, GPA=2

```
. mfx, at(psi=0, tuce=20, gpa=2)

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .00421044
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
psi*	.0394245	.05736	0.69	0.492	-.073006	.151855	0
tuce	.000399	.00092	0.43	0.666	-.001411	.002209	20
gpa	.0118491	.01782	0.66	0.506	-.023081	.046779	2

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Logit …

- Case 3A: PSI=1, TUCE=20, GPA=4

```
. mfx, at(psi=1, tuce=20, gpa=4)

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .92857112
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.382141	.25122	1.52	0.128	-.110233	.874515	1
tuce	.0063115	.01418	0.45	0.656	-.021483	.034106	20
gpa	.187447	.15883	1.18	0.238	-.123848	.498742	4

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Logit …

- Case 3B: PSI=0, TUCE=20, GPA=4

```
. mfx, at(psi=0, tuce=20, gpa=4)

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .54643009
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.382141	.25122	1.52	0.128	-.110233	.874515	0
tuce	.0235843	.0366	0.64	0.519	-.048141	.09531	20
gpa	.7004358	.26442	2.65	0.008	.182177	1.21869	4

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Logit …

- Case 4A: PSI=1, TUCE=20, GPA=2

```
. mfx, at(psi=1, tuce=20, gpa=2)

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .04363495
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.0394245	.05736	0.69	0.492	-.073006	.151855	1
tuce	.003971	.00802	0.50	0.620	-.01174	.019682	20
gpa	.1179364	.12099	0.97	0.330	-.1192	.355072	2

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Logit …

- Case 4A: PSI=0, TUCE=20, GPA=2

```
. mfx, at(psi=0, tuce=20, gpa=2)

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .00421044
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.0394245	.05736	0.69	0.492	-.073006	.151855	0
tuce	.000399	.00092	0.43	0.666	-.001411	.002209	20
gpa	.0118491	.01782	0.66	0.506	-.023081	.046779	2

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Logit …

- In sum, the various scenarios reveal that the probability that a student's grade will increase after exposure to PSI is far greater for students with high GPAs than for those with low GPAs.

The Probit Regression

The Probit ...

- Consider a class of BRM of the form:

$$\Pr\left(Y=1\middle|X\right)=\Phi\left(\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k\right)$$

- The equation above defines the conditional probabilities of Y=1 (i.e. Y occurring) given X.
- For a more compact representation:

$$\Pr\left(y=1\middle|X\right)=\Phi\left(X\beta\right)$$

The Probit ...

- In the Probit model, $\Phi(X\beta)$ can be expressed as:

$$\Phi(X\beta) = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right) dZ$$

where $\mathrm{Z} = X\beta$

- The equation above is the cumulative standard normal distribution function and it ranges between zero and one for all values of $X\beta$.

The Probit ...

- Note the following about the equation:
  - Φ is a non-linear function of $X\beta$ and hence, we cannot use OLS.
  - The errors follow standard normal distribution.
  - The estimator is the Maximum Likelihood estimator.

The Probit ...

- When dealing with Probit regressions, the following parameters are usually estimated:

➢ Z-scores

➢ Marginal effects

➢ Conditional Probability

- Let us now take each in turn.

The Probit ...

- An Illustration (using our previous example):
- The Problem:
- A researcher is interested in how variables, such as GPA (grade point average) and GQES (graduate qualifying exam scores) affect admission into graduate school (GSA).
- The response variable (GSA), admit/don't admit, is a binary variable.

The Probit ...

- The model can be expressed as:

$$\Pr\left(\text{GSA}=1 \middle| \text{GQES},\text{GPA}\right) = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right) dZ$$

$$Z = \beta_0 + \beta_1 \text{GQES} + \beta_2 \text{GPA}$$

- If after estimation we obtain the following for the relevant parameters:

$$\hat{\beta}_0 = -1.4968 \quad \hat{\beta}_1 = 0.0009 \quad \hat{\beta}_2 = 0.3827$$

- Not to worry, the estimation process will be demonstrated later using Stata software.

The Probit …

- **Z-score**: It denotes the Z-value of a normal distribution. This statistic is similar to the log odds in terms of the RHS of the underlying specification.
- Recall:
- For log odds;

$$L = X\beta = \beta_0 + \beta_1 \text{GQES} + \beta_2 \text{GPA} \qquad (1)$$

- Similarly, for Z-scores;

$$Z = X\beta = \beta_0 + \beta_1 \text{GQES} + \beta_2 \text{GPA} \qquad (2)$$

The Probit …

- Note the difference between the log of odds & Z-scores:

➢ While the RHS of the equation for (1) is determined by taking the log of the odds, that of the Z-scores is obtained by taking the reciprocal of the standard normal distribution function.

- $$L = \text{In}\left(\frac{\Lambda\left(X\beta\right)}{1-\Lambda\left(X\beta\right)}\right) = X\beta$$

$$Z - score = \Phi^{-1}\left(X\beta\right) = X\beta$$

The Probit …

- While (1) follows the standard logistic distribution, (2) is consistent with the standard normal distribution.
- If the estimation of the coefficients reported follows equation (2) using the MLE (we will demonstrate this using stata), then, the coefficients are already expressed in z-cores.

The Probit ...

- Interpretation of the Z scores:
- In relation to our example, the Z-scores can be expressed as:

$$z = X\beta = \beta_0 + \beta_1 \text{GQES} + \beta_2 \text{GPA}$$

- Given our earlier estimates, the equation becomes:

$$z = -1.4968 + 0.0009 * \text{GQES} + 0.3827 * \text{GPA}$$

The Probit ...

- For GQES: The results show that the z-score in favour of gaining admission into graduate school increases by 0.0009 for each additional unit increase in GQES.

- For GPA: The results show that a unit increase in GPA will lead to an increase in the z-score in favour of gaining admission into graduate school by 0.3827.

The Probit ...

- **Marginal Effect**: As previously defined, it measures the change in the probability of Y=1 as a result of a unit change in a particular explanatory variable.
- That is; $\dfrac{\partial \Phi(Z)}{\partial X}$
- Recall that:

$$\Phi(Z) = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right) dZ$$

The Probit …

- We can rewrite as:

$$\Phi(Z) = \int_{-\infty}^{Z} \phi(Z)\, dZ$$

- where $\phi(Z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right)$ & $Z = X\beta$

- Therefore,

$$\frac{\partial \Phi(Z)}{\partial X} = \beta * \phi(Z)$$

The Probit ...

- Marginal Effect for GQES

$$= \beta_1 * \phi(Z)$$

$$\phi(Z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right)$$

- Where $Z = \beta_0 + \beta_1 \text{GQES} + \beta_2 \text{GPA}$
- From our results, β_1 is known but we need to compute Z and $\phi(Z)$.

The Probit…

- Like the log odds, we need to determine the level of the regressors at which the Z can be computed.
- Recall the equation for the Z:

$$Z = -1.4968 + 0.0009 * \mathrm{GQES} + 0.3827 * \mathrm{GPA}$$

- You may consider the mean values of the regressors.
- Therefore, the equation becomes:

$$Z = -1.4968 + 0.0009 * \overline{\mathrm{GQES}} + 0.3827 * \overline{\mathrm{GPA}}$$

The Probit ...

- As previously noted, the mean values are

$$\overline{\mathrm{GPA}} = 3.3899; \overline{\mathrm{GQES}} = 587.7$$

- By further simplification, we have:

$$\mathrm{Z} = -1.4968 + 0.0009 * 587.7 + 0.3827 * 3.3899$$

$$Z = 0.3295$$

- Thus;

$$\phi(Z) = \frac{1}{\sqrt{\frac{44}{7}}} \exp\left(-\frac{0.3295^2}{2}\right) = 0.3778$$

The Probit ...

- The ME for GQES = 0.0009*0.3778
  = 0.0003
- Interpretation:
- A unit increase in GQES will increase the probability of gaining admission into the graduate school by 0.0003 or 0.03%.

The Probit ...

- Class Exercise 2:
- Compute the Marginal effect for GPA and interpret your results.

The Probit ...

- Dealing with Binary regressors in Probit models:
- Like the Logit case, we can extend our earlier example to include a dummy regressor such as the prestige of the undergraduate institutions attended by the applicants.
- INST =1 if a student attended one of the top-ranked universities and zero other wise.

The Probit ...

- The extended equation is given as:

$$\Pr\left(\text{GSA}=1 \middle| \text{GQES}, \text{GPA}, \text{INST}\right)$$

$$= \Phi\left(X\beta\right) = \Phi\left(\beta_0 + \beta_1 \text{GQES} + \beta_2 \text{GPA} + \beta_3 \text{INST}\right)$$

- Recall that GQES and GPA are continuous variables while INST is a discrete (binary) variable.

The Probit ...

- Our focus now is on INST.
- **Z-Scores**:
- Let us assume that the coefficient for the INST is 0.1739 after estimating the Z-score function.
- The Z-score in favour of gaining admission into the graduate school is 0.1739 higher for students from top-ranked than those from other universities.

The Probit ...

- Marginal Effect for a Binary regressor (INST):

$$\Phi\left(\hat{\beta}_0 + \hat{\beta}_1\overline{\text{GQES}} + \hat{\beta}_2\overline{\text{GPA}} + \hat{\beta}_3(1)\right)$$

$$-\Phi\left(\hat{\beta}_0 + \hat{\beta}_1\overline{\text{GQES}} + \hat{\beta}_2\overline{\text{GPA}} + \hat{\beta}_3(0)\right)$$

- An Empirical Application of Probit Models using Stata

The Probit …

- Estimation:
- Z-scores
- Marginal Effects
- Some plausible scenarios

The Probit …

- Z-scores:

```
. probit grade psi tuce gpa

Iteration 0:   log likelihood =  -20.59173
Iteration 1:   log likelihood = -12.908126
Iteration 2:   log likelihood = -12.818963
Iteration 3:   log likelihood = -12.818803
Iteration 4:   log likelihood = -12.818803

Probit regression                                 Number of obs   =         32
                                                  LR chi2(3)      =      15.55
                                                  Prob > chi2     =     0.0014
Log likelihood = -12.818803                       Pseudo R2       =     0.3775
```

grade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
psi	1.426332	.5950379	2.40	0.017	.2600795	2.592585
tuce	.0517289	.0838903	0.62	0.537	-.1126929	.2161508
gpa	1.62581	.6938825	2.34	0.019	.2658255	2.985795
_cons	-7.45232	2.542472	-2.93	0.003	-12.43547	-2.469166

The Probit …

- Marginal Effects

```
. mfx compute

Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .26580809
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
psi*	.464426	.17028	2.73	0.006	.130682	.79817	.4375
tuce	.0169697	.02712	0.63	0.531	-.036184	.070123	21.9375
gpa	.5333471	.23246	2.29	0.022	.077726	.988968	3.11719

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Probit …

- Some plausible scenarios for marginal effects:
- Case 1A: PSI=1, TUCE=20, GPA=4
- Case 1B: PSI=1, TUCE=20, GPA=2
- Case 2A: PSI=0, TUCE=20, GPA=4
- Case 2B: PSI=0, TUCE=20, GPA=2
- Case 3A: PSI=1, TUCE=20, GPA=4
- Case 3B: PSI=0, TUCE=20, GPA=4
- Case 4A: PSI=1, TUCE=20, GPA=2
- Case 4B: PSI=0, TUCE=20, GPA=2

The Probit …

- Case 1A: PSI=1, TUCE=20, GPA=4

```
. mfx, at(psi=1, tuce=20, gpa=4)

Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .93471169
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.4006438	.23361	1.72	0.086	-.057224	.858512	1
tuce	.0065815	.01626	0.40	0.686	-.02528	.038443	20
gpa	.2068523	.18384	1.13	0.261	-.153468	.567173	4

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Probit …

- Case 1B: PSI=1, TUCE=20, GPA=2

```
. mfx, at(psi=1, tuce=20, gpa=2)

Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .04094808
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.0401756	.06883	0.58	0.559	-.094733	.175085	1
tuce	.0045433	.01004	0.45	0.651	-.01514	.024227	20
gpa	.1427923	.15165	0.94	0.346	-.154438	.440022	2

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Probit …

- Case 2A: PSI=0, TUCE=20, GPA=4

```
. mfx, at(psi=0, tuce=20, gpa=4)

Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .53406788
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.4006438	.23361	1.72	0.086	-.057224	.858512	0
tuce	.0205616	.03412	0.60	0.547	-.046309	.087432	20
gpa	.6462381	.24728	2.61	0.009	.161577	1.1309	4

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Probit …

- Case 2B: PSI=0, TUCE=20, GPA=2

```
. mfx, at(psi=0, tuce=20, gpa=2)

Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .00077243
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
psi*	.0401756	.06883	0.58	0.559	-.094733	.175085	0
tuce	.0001374	.00049	0.28	0.778	-.000815	.00109	20
gpa	.0043174	.01189	0.36	0.717	-.018987	.027621	2

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Probit …

- Case 3A: PSI=1, TUCE=20, GPA=4

```
. mfx, at(psi=1, tuce=20, gpa=4)

Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .93471169
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.4006438	.23361	1.72	0.086	-.057224	.858512	1
tuce	.0065815	.01626	0.40	0.686	-.02528	.038443	20
gpa	.2068523	.18384	1.13	0.261	-.153468	.567173	4

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Probit …

- Case 3B: PSI=0, TUCE=20, GPA=4

```
. mfx, at(psi=0, tuce=20, gpa=4)

Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .53406788
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
psi*	.4006438	.23361	1.72	0.086	-.057224	.858512	0
tuce	.0205616	.03412	0.60	0.547	-.046309	.087432	20
gpa	.6462381	.24728	2.61	0.009	.161577	1.1309	4

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Probit …

- Case 4A: PSI=1, TUCE=20, GPA=2

```
. mfx, at(psi=1, tuce=20, gpa=2)

Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .04094808
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.0401756	.06883	0.58	0.559	-.094733	.175085	1
tuce	.0045433	.01004	0.45	0.651	-.01514	.024227	20
gpa	.1427923	.15165	0.94	0.346	-.154438	.440022	2

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Probit …

- Case 4A: PSI=0, TUCE=20, GPA=2

```
. mfx, at(psi=0, tuce=20, gpa=2)

Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .00077243
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
psi*	.0401756	.06883	0.58	0.559	-.094733	.175085	0
tuce	.0001374	.00049	0.28	0.778	-.000815	.00109	20
gpa	.0043174	.01189	0.36	0.717	-.018987	.027621	2

(*) dy/dx is for discrete change of dummy variable from 0 to 1

The Probit …

- The results are similar to those obtained under Logit models. The various scenarios suggest that the probability that a student's grade will increase after exposure to PSI is far greater for students with high GPAs than for those with low GPAs.

Logit vs. Pobit

- Results tend to be very similar
- Preference for one over the other tends to vary by discipline.
- Amemiya suggests multiplying a Logit estimate by 0.625 to get the corresponding Probit estimate.
- Conversely, multiplying a Probit estimate by 1.6 (i.e. 1/0.625) gives the corresponding logit estimate. Try it!!!

Diagnostics for Binary Response Models

- In order for our analysis to be valid, our model has to satisfy the assumptions of the BRM.
- When the assumptions of the BRM are not met, we may have problems, such as biased coefficient estimates or very large standard errors for the regression coefficients, and these problems may lead to invalid statistical inferences.

Diagnostics …

- Therefore, before we can use our model to make any statistical inference, we need to check that our model fits sufficiently well.

Diagnostics …

- Is the model correctly specified?
    - Specification test
- Is the overall model statistically significant?
    - Goodness-of-fit test
- Are the regressors orthogonal (uncorrelated)?
    - Multicollinearity test

Diagnostics …

- **Specification test:** This is conducted to confirm that the probability function is correctly specified.
- Procedure:
- The test involves two steps
- Step 1: Estimate the probability function (either logit or probit)
- Step 2: Use the information from step 1 to build the model for the test and estimate appropriately.

Diagnostics …

- Specification test for the Logit Model: Step 1.

```
. logit grade psi tuce gpa

Iteration 0:   log likelihood =  -20.59173
Iteration 1:   log likelihood = -13.259768
Iteration 2:   log likelihood = -12.894606
Iteration 3:   log likelihood = -12.889639
Iteration 4:   log likelihood = -12.889633
Iteration 5:   log likelihood = -12.889633

Logistic regression                               Number of obs   =         32
                                                  LR chi2(3)      =      15.40
                                                  Prob > chi2     =     0.0015
Log likelihood = -12.889633                       Pseudo R2       =     0.3740
```

grade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
psi	2.378688	1.064564	2.23	0.025	.29218	4.465195
tuce	.0951577	.1415542	0.67	0.501	-.1822835	.3725988
gpa	2.826113	1.262941	2.24	0.025	.3507938	5.301432
_cons	-13.02135	4.931325	-2.64	0.008	-22.68657	-3.35613

Diagnostics …

- Specification test: Step 2

```
. linktest

Iteration 0:   log likelihood =  -20.59173
Iteration 1:   log likelihood = -13.220543
Iteration 2:   log likelihood = -12.887479
Iteration 3:   log likelihood = -12.860508
Iteration 4:   log likelihood = -12.860434
Iteration 5:   log likelihood = -12.860434

Logistic regression                               Number of obs   =         32
                                                  LR chi2(2)      =      15.46
                                                  Prob > chi2     =     0.0004
Log likelihood = -12.860434                       Pseudo R2       =     0.3755
```

grade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	.9551764	.3834559	2.49	0.013	.2036166	1.706736
_hatsq	-.0453861	.1881828	-0.24	0.809	-.4142176	.3234455
_cons	.0817277	.6074585	0.13	0.893	-1.108869	1.272324

Diagnostics …

- Interpretation:
- The model is correctly specified if ‘ _hat’ is statistically significant and ‘_hatsq’ is not.
- In our example, the model is correctly specified.

Diagnostics …

- How can we resolve specification bias in probability models?
- Exclude redundant variables
- Include relevant variables such as squared terms & interaction terms
- Note that after modifying the model to correct for specification bias, you have to conduct the specification test again to confirm that the modification is valid.

Diagnostics …

- Goodness-of-fit test:
  - (1) Likelihood ratio (LR) test
  - (2) Hosmer and Lemeshow's goodness-of-fit test

- Note that the LR statistic (1) is reported by default when you estimate logit/probit models in stata.

Diagnostics …

- The estimated model fits the data well if the LR test statistic is statistically significant.

Diagnostics …

- Hosmer and Lemeshow's (HL) goodness-of-fit test: This test examines whether the predicted frequency and observed frequency match closely. The more closely they match, the better the fit.

- Procedure: Two steps
- Step 1: Estimate the model (logit/probit) under consideration.
- Step 2: Perform the test on the regression output obtained in step 1.

Diagnostics …

- HL test for the Logit model

```
. lfit, group(10) table
```

Logistic model for grade, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.0982	1	0.3	3	3.7	4
2	0.1069	0	0.3	3	2.7	3
3	0.1362	0	0.4	3	2.6	3
4	0.1770	0	0.5	3	2.5	3
5	0.2876	1	0.7	2	2.3	3
6	0.3784	2	1.4	2	2.6	4
7	0.4397	1	1.3	2	1.7	3
8	0.6000	2	1.8	1	1.2	3
9	0.7167	1	2.0	2	1.0	3
10	0.8551	3	2.5	0	0.5	3

```
number of observations =        32
      number of groups =        10
Hosmer-Lemeshow chi2(8) =         6.34
           Prob > chi2 =         0.6094
```

Diagnostics …

- The estimated model fits the data well if the HLtest statistic is not statistically significant.

Diagnostics …

- Mutlicollinearity test: This test is conducted to verify if there is presence of severe correlations among regressors in the model under consideration.
- A user written program 'collin' is used to detect the multicollinearity.
- You have to install the program before you conduct the test.

Diagnostics …

- For installation, use the command below (you need internet connection for the installation):

  findit collin

Procedure: Two Steps

Step 1: Install the program if you have not done so.

Step 2: Perform the test (note the command for the test)

Diagnostics …

- If all the variables are orthogonal to each other, in other words, completely uncorrelated with each other, both the tolerance and variance inflation factor (VIF) are 1. If a variable is very closely related to another variable, the tolerance goes to 0, and the VIF gets very large.
- Tolerance = 1 – R^2
- VIF = 1/Tolerance

Diagnostics …

- In other words, the closer the tolerance and VIF values to 1, the less severe the multicollinearity problem in the model.
- As a rule of thumb, a tolerance of 0.1 or less (equivalently VIF of 10 or greater) is a cause for concern.

Diagnostics …

- Multocollinearity test

```
. collin tuce gpa
(obs=32)

  Collinearity Diagnostics

                        SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
      tuce      1.18    1.08     0.8502      0.1498
       gpa      1.18    1.08     0.8502      0.1498
----------------------------------------------------
```

Diagnostics …

- Based on the results obtained for the tolerance and VIF statistics, we can conclude that there is no presence of severe multicollinearity problem in the model.

Diagnostics …

- How can we deal with severe correlations among regressors?

(1) Identify the source(s) of multicollinearity through the multicollinearity test.

(2) The variable with high VIF and Tolerance values is a potential source of multicollinearity in the model.

(3) In connection with (2), you may consider any of the following:

Diagnostics …

(a) transform the affected variable appropriately in such a way as to make it orthogonal to other variables in the model (e.g. by demeaning the affected variable(s))

(b) consider a more appropriate proxy (i.e. a different variable) that will be orthogonal to other variable(s)