PPOL 503-03, PPOL 503-04, Fall 2016
Course Notes #4: Binary Dependent Variables: the Linear Probability Model

I. Dichotomous Dependent Variables

Examples: labor force participation, union membership, home ownership.
- The response is: yes or no.
- Dependent variable: 0 or 1. This violates the assumption that the dependent variable be continuous (normal).
- Alternative solutions.
- Linear Probability Model (LPM).

- How do we think about *Y* now in the LPM model? Now, with the dependent variable defined as a variable that can only take on values of zero or one, $E(Y|\mathbf{X})$ has a particular interpretation:

$E(Y|\mathbf{X}) = Probability\ (Y=1|\mathbf{X}),\ or\ P\ (Y=1|\mathbf{X}).$

$P\ (Y=1|\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2\ X_2 + \ \beta_3 X_3 .... \ \beta_k X_k$

- In all previous multiple regression models, our interpretation of β_k was the predicted change in *Y* given a one-unit increase in X_k, holding all other variables in the model constant.

- With LPM models, the interpretation of β_k will be slightly different, due to the fact that the dependent variable is now a variable that can only range from 0 to 1 and is now interpreted as a probability:

  For LPM models, β_k is the predicted probability of "success" (i.e., the case when the dependent variable is equal to one) when X_k increases by one unit (examples on the next page)

  Q: What is the interpretation of β_0 in LPM models?

- You can obtain predicted values from LPM models just as you can from regression models with an interval/ratio dependent variable. There will be one problem though: sometimes the prediction can be outside the range of 0 to 1.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_k X_k + \eta$$

Let P_i= Probability that Y_i=1 (the event occurs), and (1-P_i) the probability that Y_i=0 (the event does not occur).

Y_i	Probability
0	$1-P_i$
1	P_i
Total	1

$$E(Y_i) = 0\,(1 - P_i) + 1\,(P_i) = P_i$$

$$E(Y_i / X_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_k X_k = P_i$$

Since the probability P_i must lie between 0 and 1, we impose the restriction $0 <= E(Y_i / X_i) <= 1$. A linear regression model with a dependent variable equal to 0 or 1 is called a linear probability model (LPM).

Problems with the Linear Probability Model

1) Non-normality of the error term (disturbance)

$$U_i = Y_i - \beta_0 - \beta_1 X_1 - \beta_2 X_2 - \beta_3 X_3 \ldots \beta_k X_k$$

Where $Y_i = 1$ then $U_i = 1 - \beta_0 - \beta_1 X_1$

$Y_i = 0$ then $U_i = -\beta_0 - \beta_1 X_1$

Normality is no longer tenable because like Yi, u_i takes 2 values. Still obtain unbiased estimates. As sample size increases, OLS estimators tend to be normally distributed.

2) Heteroskedastic Variances of the Error Term

Even of $E(u_i) = 0$ and $E(u_i\, u_j) = 0$ for $i \neq j$, that is, no serial correlation, it can no longer be maintained that the error terms (disturbances) are homoskedastic.

Y_i	u_i	Probability
0	$-\beta_1 - \beta_2 X_2$	$1 - P_i$
1	$1 - \beta_1 - \beta_2 X_2$	P_i

$$Var\,(u_i) = E(Y_i / X_i)\,[\,1 - E(Y_i / X_i)\,] = P_i\,(1 - P_i)$$

The variance of (u_i) is heteroskedastic because it depends on the conditional expectation of Y, which depends on the value of X.

One way to solve the problem of heteroskedasticity is to transform the data. Divide both side by a factor:

$$\sqrt{E(Y_i / X_i)\left[1 - E(Y_i / X_i)\right]} = \sqrt{P_i(1 - P_i)} = \sqrt{W_i}$$

$$\frac{Y_i}{\sqrt{W_i}} = \frac{\beta_1}{\sqrt{W_i}} + \beta_2 \frac{X_i}{\sqrt{W_i}} + \frac{u_i}{\sqrt{W_i}}$$

The factor is 1 / square root of the variance. The disturbance is now homoskedastic.

Since the weights are unknown use the following approach to estimate the weights.

Step 1: Run OLS on the 0 – 1 dependent variable and obtain its predicted value P_i . If the predicted probabilities lie outside the 0 to 1 range, the calculated weight will be imaginary or infinity. To ensure that all weights are calculated, any predicted value outside the range is set equal to 0.5. This minimizes the effects of such cases. If P_i is greater than or equal to 1, let P_i equal .5. If P_i is less than or equal to 0, let P_i equal .5.

$$W = P_i (1 - P_i)$$

Step 2: Use the estimated W to transform the data and run the regression using weighted least squares.

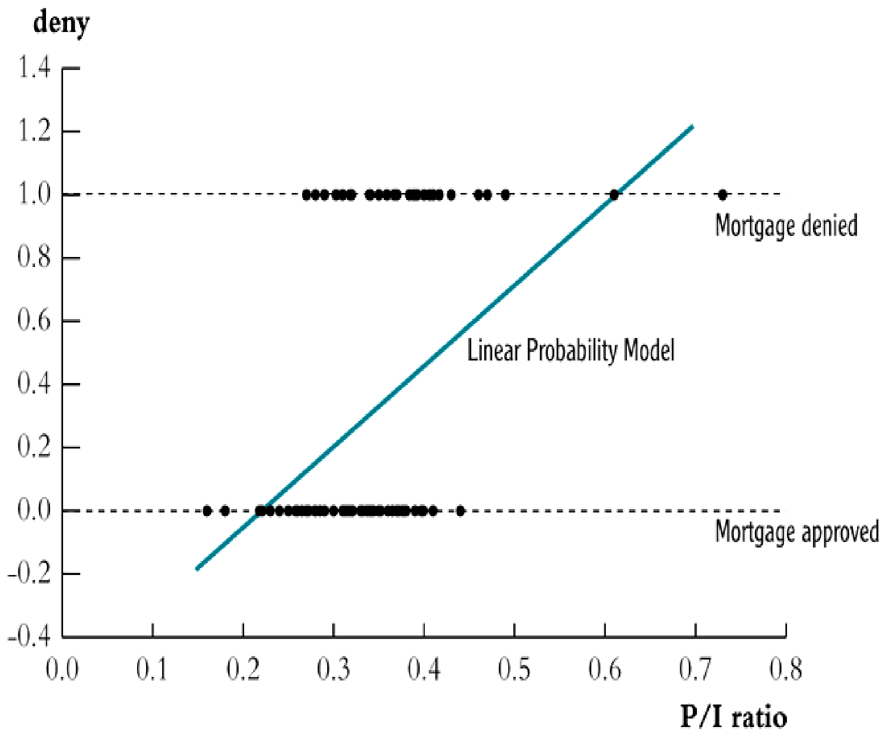
$$Weight = \frac{1}{\sqrt{P_i (1 - P_i}}$$

LPM Problems and Solutions:

1. non-normality of the error (disturbance); solution: increase sample size to minimize the effects of non-normality.
2. Heteroskedasticity of the error term; solution: use weighted least squares.
3. estimated probabilities that lie outside the range 0 to 1; use restricted least squares to force the probabilities to lie inside the range 0 to 1.

FIGURE 9.1 **Scatterplot of Mortgage Application Denial and the Payment-to-Income Ratio**

Mortgage applicants with a high ratio of debt payments to income (*P/I ratio*) are more likely to have their application denied (*deny* = 1 if denied, *deny* = 0 if approved). The linear probability model uses a straight line to model the probability of denial, conditional on the *P/I ratio*.



EXAMPLE of LPM: $\hat{Deny} = -0.080 + 0.604(P / I\ Ratio)$

EXAMPLE: Probability of being in a clerical occupation

$$CLEROCC = \beta_0 + \beta_1 EDUC + \beta_2\ FEMALE + \beta_3 NONWHITE + \ \eta$$

clerocc =1 if in clerical occupation
educ years of education
nonwhite =1 if nonwhite
female =1 if female

```
Variable      N            Mean         Std Dev         Minimum         Maximum
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
clerocc     526       0.1673004       0.3735991               0       1.0000000
educ        526      12.5627376       2.7690224               0      18.0000000
female      526       0.4790875       0.5000380               0       1.0000000
nonwhite    526       0.1026616       0.3038053               0       1.0000000
     ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
```

```
Dependent Variable: clerocc
                                Analysis of Variance
                                        Sum of           Mean
Source                      DF         Squares         Square    F Value    Pr > F
Model                        3         9.34173        3.11391      25.42    <.0001
Error                      522        63.93584        0.12248
Corrected Total            525        73.27757

Root MSE                 0.34997    R-Square     0.1275
Dependent Mean           0.16730    Adj R-Sq     0.1225
Coeff Var              209.18956

                        Parameter        Standard
Variable     DF          Estimate           Error    t Value    Pr > |t|
Intercept     1           0.01009         0.07477       0.13      0.8927
educ          1           0.00264         0.00556       0.48      0.6348
female        1           0.26642         0.03066       8.69      <.0001
nonwhite      1          -0.03516         0.05047      -0.70      0.4864
```

- predicted probability of being a clerical worker for (e.g.):

  (a) a person with no education, male, who is white = **0.01**

  (b) a female with no education, who is white = 0.01+0.266 = **0.276**

(c) a female with 10 years of education who is nonwhite = 0.01+ (0.00264*10)+0.266-0.035 = **0.3374**

In other words, our best guess of the percent of the (c) group who work in a clerical occupation is 33.74%.

- β_{EDUC}: This effect is *not* statistically significant. However, if it were, we would interpret this coefficient as:

    > Holding gender and race constant, for each additional year of schooling, the predicted probability of working in a clerical occupation increases by 0.003. (Note: remember that the possible outcome scale is 0 to 1).

    In other words, there is a 0.3 *percentage point* effect of an additional year of education (i.e., less than ½ of 1 percentage point).

    NOTE: the "percentage point" interpretation refers to a possible outcome scale of 0 to 100 (instead of 0 to 1 as the variable was originally defined). Each 1-point increase is a percentage point increase. To convert from probability to percentage points, just move the decimal two places to the right; i.e., multiply by 100).

    Holding gender and race constant, what's the change in predicted probability of being a clerical worker for going to school an additional ten years?

- β_{FEMALE}: Holding education and race constant, the predicted probability of working in a clerical occupation for females (compared to males) increases by 0.266. In other words, there is a 26.6 *percentage point* effect on working in a clerical occupation of being female (holding the other vars in the model constant). Is this effect statistically and/or substantively significant?

- $\beta_{NONWHITE}$: Holding gender and education constant, the predicted probability of working in a clerical occupation for nonwhites (compared to whites) decreases by 0.04. In other words, the effect is a reduction of 4 percentage points.

EXAMPLE: Probability of Arrest

Let *arr86* be a binary variable =1 if a man was arrested in 1986 and zero otherwise. The population is a group of young men in California born in 1960 or 1961 who have at least one arrest prior to 1986. The following linear probability model is estimated:

$$Arr86 \;=\; \beta_0 + \beta_1\, pcnv + \beta_2\, avgsen + \beta_3\, tottime + \beta_4\, ptime86 + \beta_5\, qemp86 + u$$

pcnv = the proportion of prior arrests that led to a conviction;

avgsen = the average sentence served from prior convictions (in months)

tottime = months spent in prison since age 18 prior to 1986

ptime86 = months spent in prison in 1986

qemp86 = the number of quarters (0 to 4) that the man was legally employed in 1986

Only 7.2% of men had more than one arrest, while about 27.7% were arrested at least once during 1986. The estimated equation is:

$$\begin{array}{llllll} \mathit{arr86} \;=\; & .441 & -\ .162\, \mathit{pcnv} & +\ .0061\, \mathit{avgsen} & -\ .0023\, \mathit{tottime} & -\ .022\, \mathit{ptime86} \\ & (.017) & (.021) & (.0065) & (.0050) & (.005) \end{array}$$

$$\begin{array}{l} -\ .043\, \mathit{qemp86} \\ \quad (.005) \end{array}$$

1. The intercept of .441 is the predicted probability of arrest for a man who has not been convicted (so *pcnv* and *avgsen* are both zero), has spent no time in prison since age 18, spent no time in prison in 1986 and was unemployed during the entire year.
2. *avgsen* and *tottime* are not statistically significant.
3. Increasing the probability of conviction does lower the probability of arrest, but one needs to be careful in interpreting the magnitude of the coefficient. The variable *pcnv* is a proportion between zero and one. Changing *pcnv* from zero to 1 means a change from no chance of being convicted to being

convicted with certainty; the corresponding change in the probability of arrest is .162 or 16.2 percentage points.

4. The variable *pttime86* is measured in months, six more months in prison reduces the probability of arrest by .022 (6) = .132 or 13.2 percentage points.
5. The variable *qemp86* affects the probability of arrest; a man employed all four quarters is (.044) (4) = .172 less likely to be arrested than a man who was not employed at all.

Re-estimate the model but include controls for race/ethnicity

$$arr86 = .380 - .152\,pcnv + .0046\,avgsen - .0026\,tottime - .024\,ptime86$$

$$(.019)\quad(.021)\qquad(.0064)\qquad(.0049)\qquad(.005)$$

$$- .038\,qemp86 + .170\,black + .096\,hispanic$$

$$(.005)\qquad(.024)\qquad(.021)$$

The coefficient on *black* means that other thing equal, a black man has a .17 or 17 percentage point higher chance of being arrested than a white man. The coefficient on *Hispanic* means that Hispanic men have a .096 higher chance of being arrested than white men. Alternatively, the probability of arrest is 9.6 percentage points higher for Hispanic men than for white men.

Major problem with the LPM is that it assumes that P_i increases linearly with X; in other words the marginal effect of X is constant over the range of X values. To address this shortcoming of the LPM model we need a probability model with two features:

1. at higher values of X, P_i increases but it never goes outside the range 0 to 1.
2. The relationship between P_i and X_i is non-linear. That is the probability changes more slowly as the two ends of the X distribution are approached.