POLS/CS&SS 503:
Advanced Quantitative Political Methodology

BINARY DEPENDENT VARIABLES

May 26, 2015

Jeffrey B. Arnold

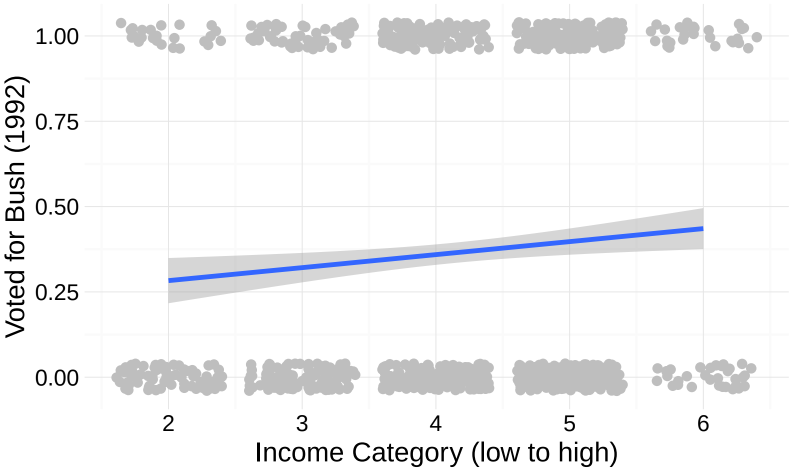
UNIVERSITY of WASHINGTON
DEPARTMENT OF POLITICAL SCIENCE

CENTER for STATISTICS and the SOCIAL SCIENCES

Overview

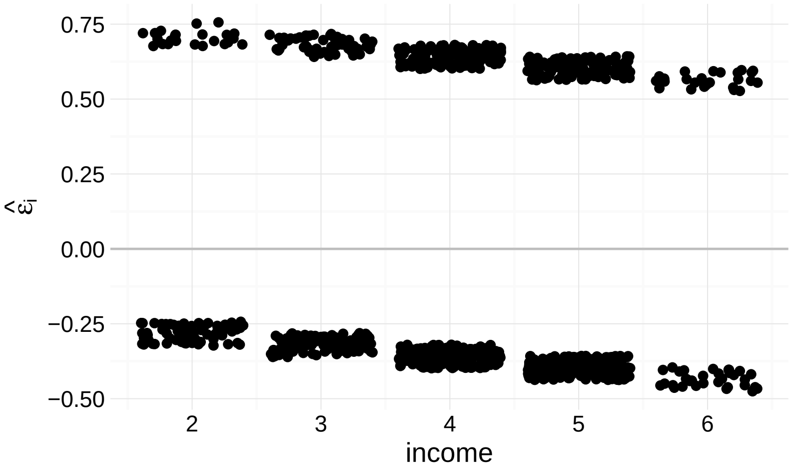
Example of Linear Probability Model

Vote for Bush in U.S. Presidential Election 1992



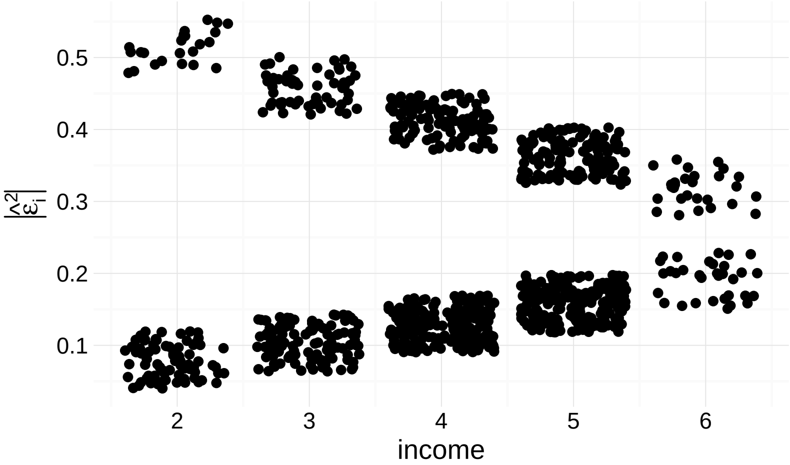
Residuals in LPM

Vote for Bush in U.S. Presidential Election 1992



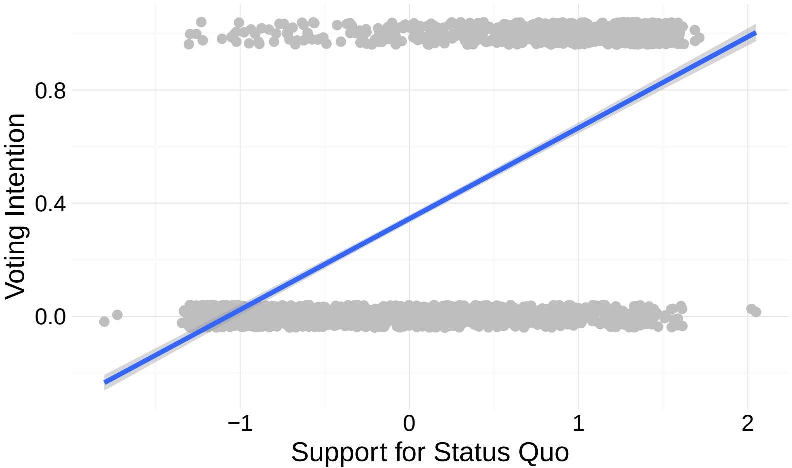
Residuals Squared in LPM

Vote for Bush in U.S. Presidential Election 1992



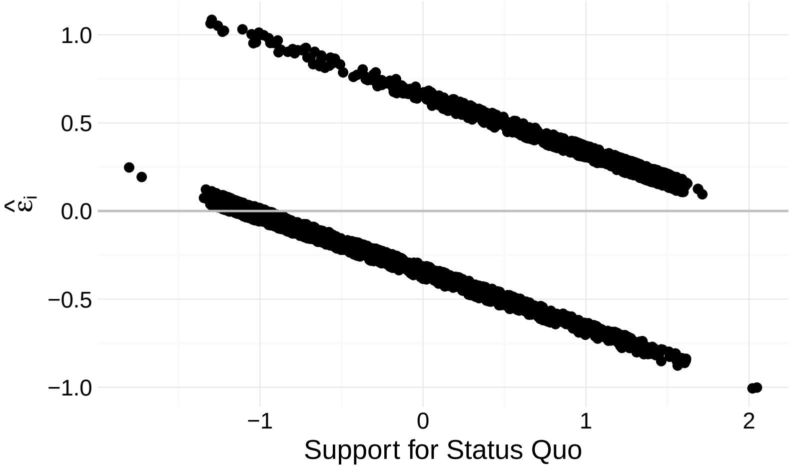
Example of Linear Probability Model

Vote Intention in Chilean Plebiscite in 1973



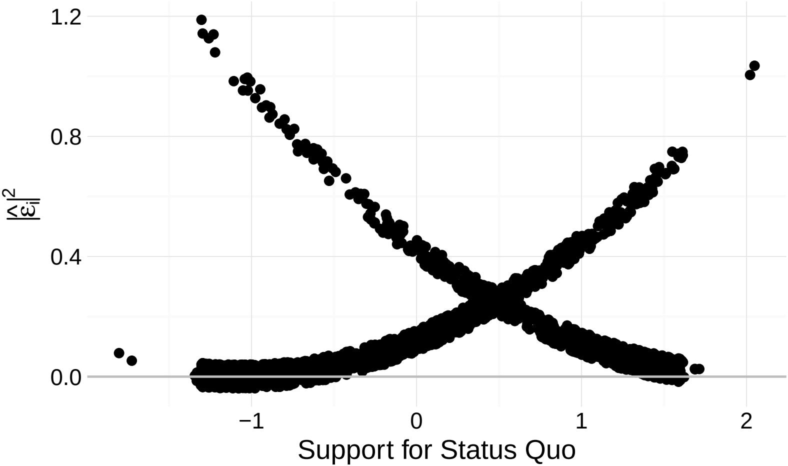
Residuals in LPM

Vote Intention in Chilean Plebiscite in 1973



Residuals Squared in LPM

Vote Intention in Chilean Plebiscite in 1973



Linear Probability Model

OLS with a binary dependent variable. When $Y_i \in \{0, 1\}$:

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

The expected value is a probability

$$E(Y_i|X_i) = \text{Pr}(Y_i = 1|X_i) = \alpha + \beta X_i$$

Problems with the LPM

- Errors are not normally distributed

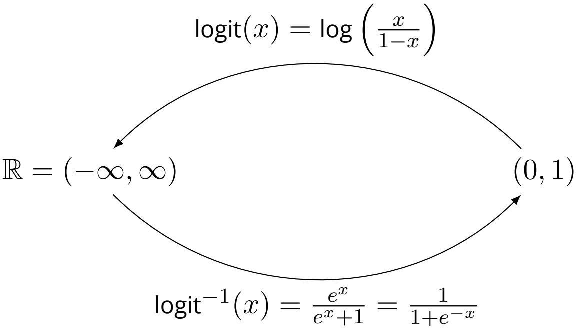
$$\epsilon_i = 1 - E(Y_i|X_i) = 1 - (\alpha - \beta X_i) = 1 - \pi_i$$
$$\epsilon_i = 0 - E(Y_i|X_i) = 0 - (\alpha - \beta X_i) = -\pi_i$$

- Errors have non-constant variance (heteroskedasticity)

$$V(\epsilon_i) = \pi(1 - \pi_i)$$

- $E(Y_i|X_i) = \alpha + \beta X_i$ can extend beyond (0, 1)
- Improper specification leads to bias; heteroskedasticity and errors leads to incorrect standard errors.

Logit and Logistic Function



Logit and Logistic Function

Logit Function

Function $(0, 1) \to (\infty, -\infty)$

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p)$$

Interpreted as the log of the odds ratio ($p/(1-p)$).

Logistic or Inverse Logit Function

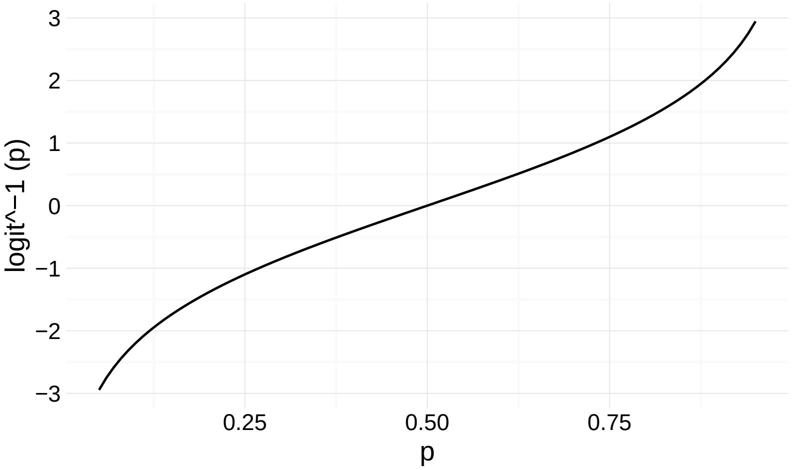
Function $(\infty, -\infty) \to (0, 1)$

$$\text{logit}^{-1}(x) = \frac{1}{1+\exp(-x)} = \frac{\exp(x)}{\exp(x)+1}$$

Logistic and logit functions are inverses of each other

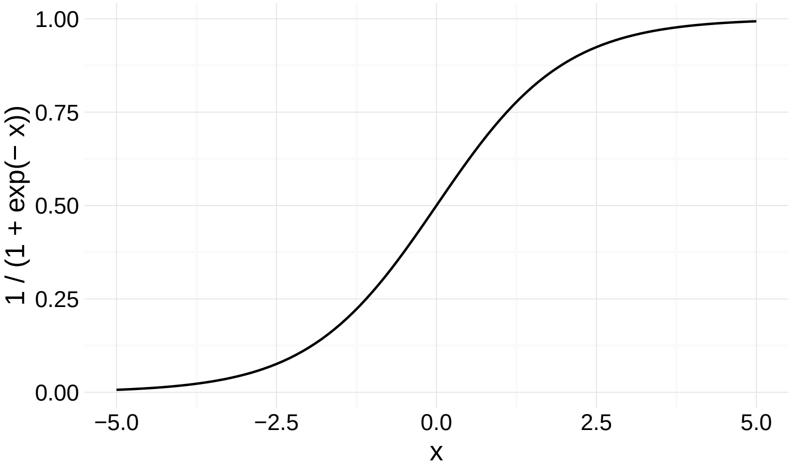
$$\text{logit}^{-1}(\text{logit}(x)) = x$$

Logit Function



$\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$

Inverse Logit (Logistic) Function



$\text{logit}^{-1}(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$

Logit Objective Function

OLS minimizes squared errors

$$\hat{\beta} = \underset{b}{\text{arg min}} \sum_i (y_i - X_i b)^2$$

Logit minimizes a **different** function

$$\hat{\beta} = \underset{b}{\text{arg min}} \sum_i (y_i \log P_i + (1 - y_i) \log(1 - P_i))$$

$$P_i = \text{logit}^{-1}(X_i b) = \frac{1}{1 + \text{exp}(-X_i b)}$$

Logit needs to be estimated by an interative maximization method

Logit Model

In logit, Pr$(Y_i = 1)$ not Y_i is directly a function of $X_i\beta$

- Probabilty of $Y_i = 1$:

$$\begin{aligned} \text{Pr}(Y_i = 1) &= f(X_i\beta) \\ &= \frac{1}{1 + \text{exp}(-(X_i\beta))} \\ &= \text{logit}^{-1}(X_i\beta) \end{aligned}$$

- Alternative interpretation, log odds ratio (log$(p/(1-p))$):

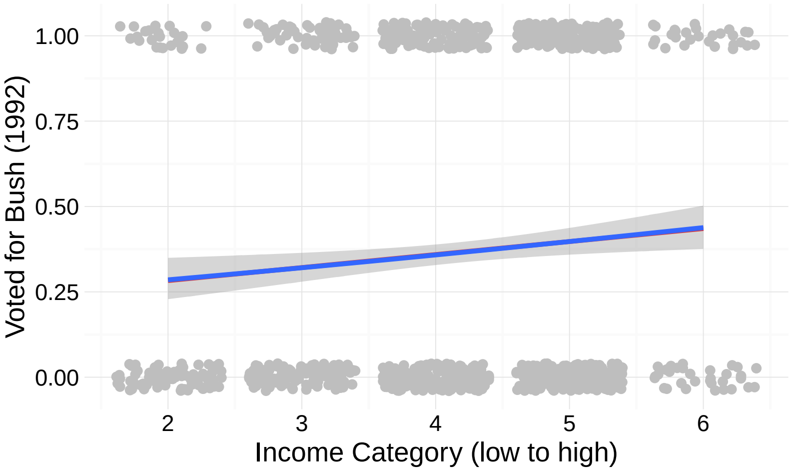
$$\begin{aligned} \text{Pr}(Y_i = 1) &= \pi_i \\ \text{logit}(\pi_i) &= \alpha + X_i\beta \end{aligned}$$

```r
summary(glm(voterep ~ income, data = nes_sample,
            family = binomial(link = "logit")))
```

```
##
## Call:
## glm(formula = voterep ~ income, family = binomial(link = "logit"),
##     data = nes_sample)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0738  -1.0066  -0.8793   1.3584   1.5838
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.25311    0.27045  -4.633  3.6e-06 ***
## income       0.16741    0.06276   2.668  0.00764 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1311.4  on 999  degrees of freedom
## Residual deviance: 1304.1  on 998  degrees of freedom
## AIC: 1308.1
##
## Number of Fisher Scoring iterations: 4
```

Example of Linear Probability Model

Vote for Bush in U.S. Presidential Election 1992

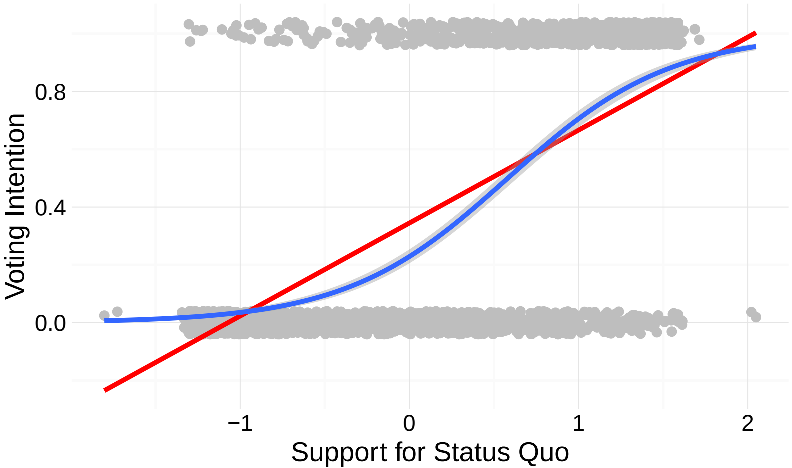


```r
summary(glm(vote_yes ~ statusquo, data = Chile,
            family = binomial(link = "logit")))
```

```
##
## Call:
## glm(formula = vote_yes ~ statusquo, family = binomial(link = "logit"),
##     data = Chile)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4942  -0.4747  -0.2290   0.5747   2.8140
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.21597    0.06955  -17.48   <2e-16 ***
## statusquo    2.08971    0.07805   26.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3242.0  on 2518  degrees of freedom
## Residual deviance: 1874.9  on 2517  degrees of freedom
##   (181 observations deleted due to missingness)
## AIC: 1878.9
##
## Number of Fisher Scoring iterations: 5
```

Example of Linear Probability Model

Vote Intention in Chilean Plebiscite in 1973



Logit Coefficients are Less Transparent

Linear Regression Coeficients

$$\frac{\partial Y}{\partial X_j} = \frac{\partial}{\partial X_j}(\alpha + \beta_1 X_1 + \dots \beta_k X_k) = \beta_j$$

Coefficient equals the marginal effect of x

Logistic Regression Coeficients

$$\frac{\partial \, \text{logit}(Y)}{\partial X_j} = \frac{\partial}{\partial X_j}\left(\frac{1}{1 + \text{exp}(\alpha + \beta X_i)}\right) = \text{Pr}(Y = 1|X_i)\,\text{Pr}(Y = 0|X_i)\beta_j$$

or

$$\frac{\partial \, \text{logit}(Y)}{\partial X_j} = \frac{\partial}{\partial X_j} X_i \beta_j = \beta_j$$

Coefficient does not equal the marginal effect of x_j

Linear Probability Model

Logit Models

LPM vs. Logit

References

The LPM Strikes Back

- LPM has renewed popular among econometricians, causal inference folks -
- See the debate here
- OLS is still Min MSE linear approx of Conditional Expectation Function
- If the functional form is wrong ; but so it logit / probit. And the functional form is **always** wrong;
- OLS coefficients are a good estimate of the average marginal effects even if not good for the marginal effects at a given x.
- OLS coefficients are directly interpretable
- Angrist and Pischke recommend LPM with heteroskedasticity consistent errors

Average Marginal Effects

- The **average marginal effect** summarizes the marginal effect $\frac{\partial y}{\partial x_j}$ averaging over the sample of x.

$$\text{Avg. Marginal Effect of } x_j = \frac{1}{n} \sum_i \left. \frac{\partial Y}{\partial x_j} \right|_{X_i}$$

- In OLS, the marginal effect of x_j (assuming no interactions, polynomials, etc.) is simply the coefficient

$$\left. \frac{\partial y}{\partial x_j} \right|_{x_i} = \frac{1}{n} \sum_i \hat{\beta}_j = \hat{\beta}_j$$

- In Logit, the average

$$\left. \frac{\partial y}{\partial x_j} \right|_{x_i} = \frac{1}{n} \sum_i \text{Pr}(y_i = 1 | \hat{\beta}, x_i) \, \text{Pr}(y_i = 0 | \hat{\beta}, x_i) \hat{\beta}_j$$

Comparing Average Marginal Effects of Logit and LPM

1992 U.S. Election Example

```
mod <- glm(voterep ~ income, data = nes_sample,
           family = binomial(link = "logit"))
mod_aug <- augment(mod, type.predict = "response")
mean(mod_aug$.fitted * (1 - mod_aug$.fitted) * coef(mod)[2])

## [1] 0.03847867

lm(voterep ~ income, data = nes_sample)

##
## Call:
## lm(formula = voterep ~ income, data = nes_sample)
##
## Coefficients:
## (Intercept)       income
##     0.20679      0.03811
```

Comparing Average Marginal Effects of Logit and LPM

Chile Plebiscite Example

```
mod <- glm(vote_yes ~ statusquo, data = Chile,
           family = binomial(link = "logit"))
mod_aug <- augment(mod, type.predict = "response")
mean(mod_aug$.fitted * (1 - mod_aug$.fitted) * coef(mod)[2])

## [1] 0.2436621

lm(vote_yes ~ statusquo, data = Chile)

##
## Call:
## lm(formula = vote_yes ~ statusquo, data = Chile)
##
## Coefficients:
## (Intercept)    statusquo
##      0.3447       0.3215
```

Linear Probability Model

Logit Models

LPM vs. Logit

References

References

- Fox, Ch. 14
- Gelman and Hill, Ch 5. This should have most material you need.
- Chile Plebicite example: Fox, Ch. 14. Data from **arm** package dataset `Chile`.
- Bush vote in 1992 example: Gelman and Hill, Ch 5. Data from `http://www.stat.columbia.edu/~gelman/arm/examples/ARM_Data.zip` as `ARM_Data/nes/nes5200_processed_voters_realideo.dta`.