POLS/CS&SS 503:
Advanced Quantitative Political Methodology

BINARY DEPENDENT VARIABLES

May 19, 2015

Jeffrey B. Arnold

UNIVERSITY of WASHINGTON
DEPARTMENT OF POLITICAL SCIENCE

CENTER for STATISTICS and the SOCIAL SCIENCES

Overview

Linear Probability Model

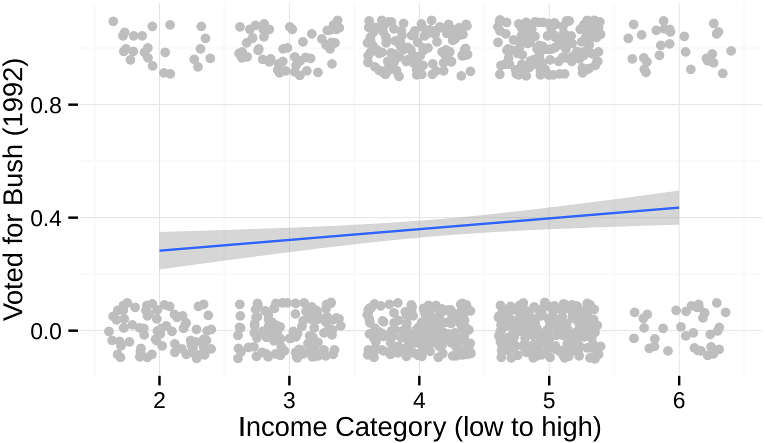
Logit Models

LPM vs. Logit

References

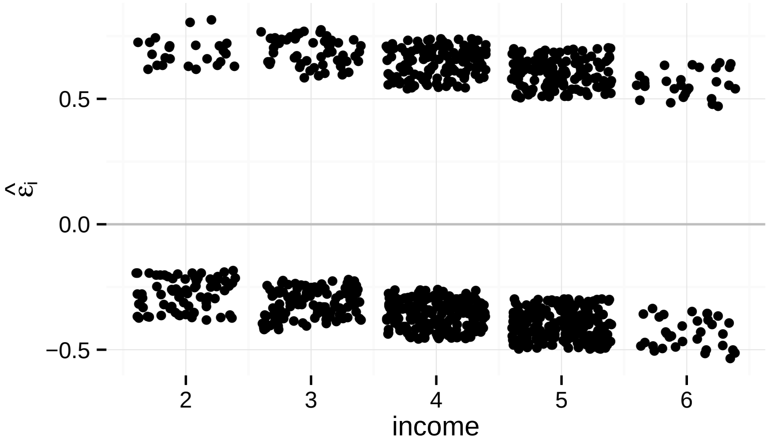
Example of Linear Probability Model

Vote for Bush in U.S. Presidential Election 1992



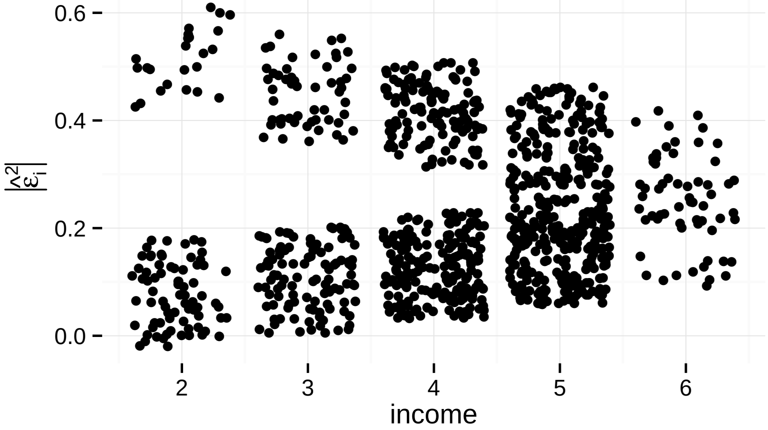
Residuals in LPM

Vote for Bush in U.S. Presidential Election 1992



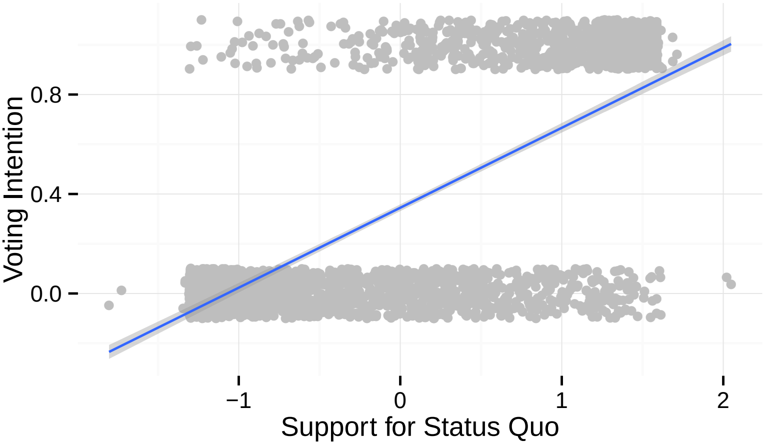
Residuals in LPM

Vote for Bush in U.S. Presidential Election 1992



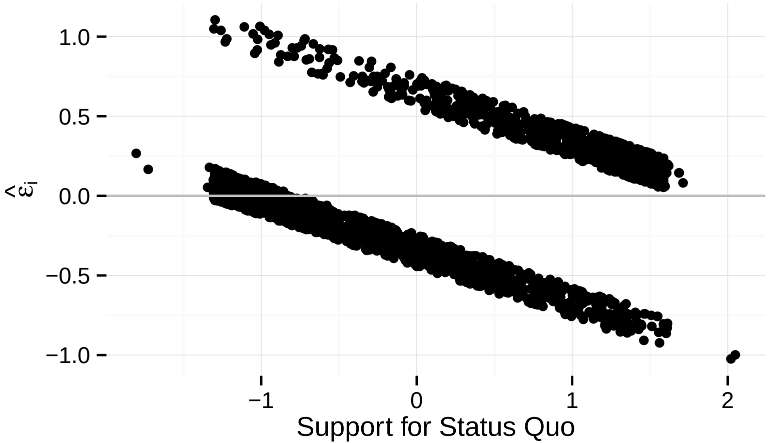
Example of Linear Probability Model

Vote Intention in Chilean Plebiscite in 1973



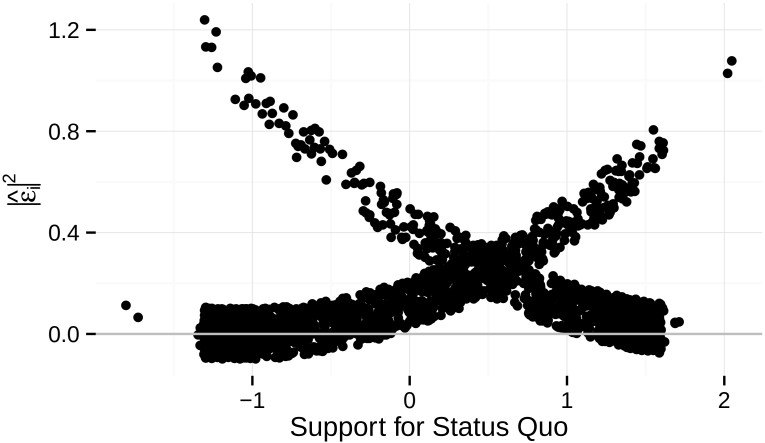
Residuals in LPM

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

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

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

$$\begin{tikzcd}[column sep=huge] (-\infty, \infty) \arrow[r, bend right, "\mathsf{logit}^{-1}(x) = \frac{e^x}{e^x+1} = \frac{1}{1+e^{-x}}"'] & (0, 1) \arrow[l, bend right, "\mathsf{logit}(x) = \mathsf{log}\left(\frac{x}{1-x}\right)"'] \end{tikzcd}$$

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

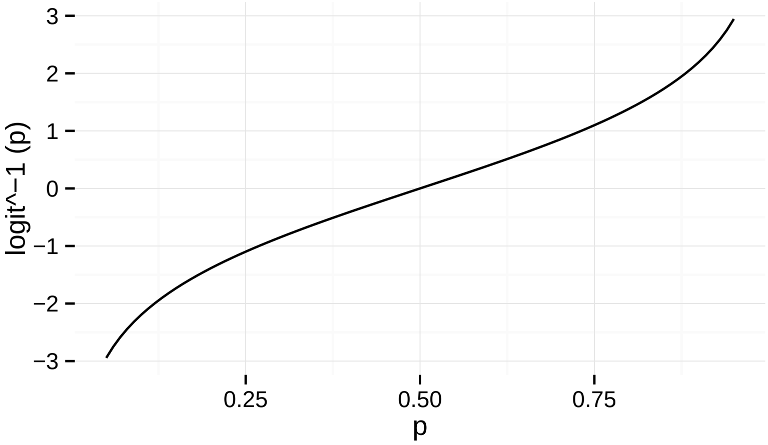
Function $(\infty, -\infty) \to (0, 1)$

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

Logistic and logit functions are inverses of each other

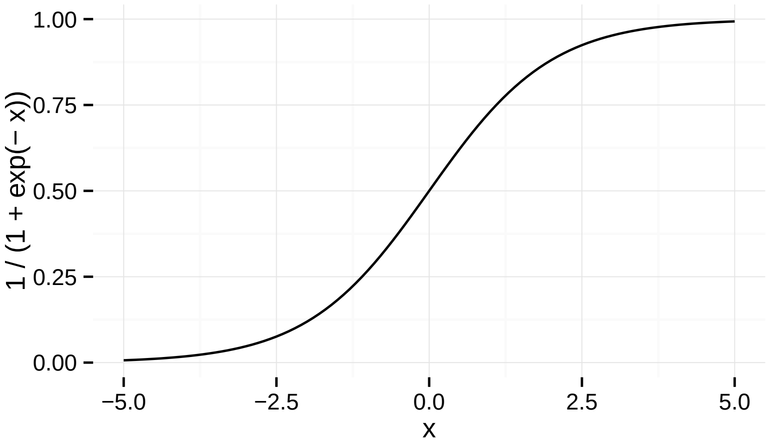
$$\text{logit}^{-1}(\text{logit}(x)) = x$$

Logit Function



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

Inverse Logit (Logistic) Function



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

Logit Model

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

- Models $\text{Pr}(Y_i = 1)$ instead of Y_i directly
- Although $X_i\beta$ is linear, $\text{logit}(x)$ is non-linear.

Logit Model as a Linear Function of Log Odds

Alternative specification:

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

Log-odds of the probability of Y is a linear function

Logit Objective Function

OLS minimizes squared errors

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

Logit minimizes a **different** function

$$\hat{\beta} = \arg\min_b \sum_i (y_i \log P_i + (1 - y_i) \log(1 - P_i))$$

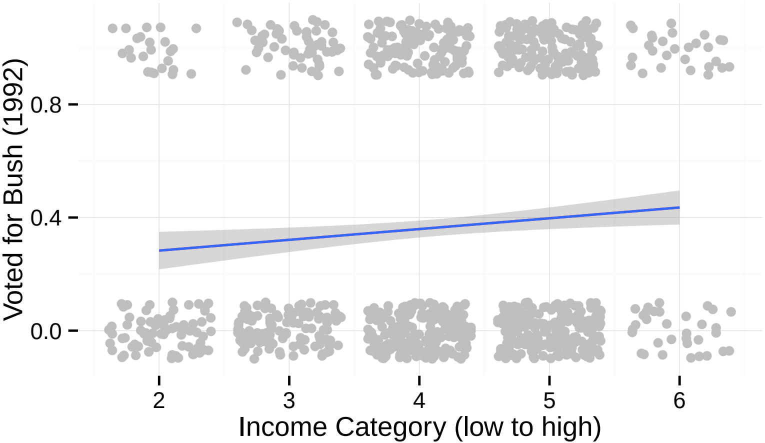
$$P_i = \text{logit}^{-1}(X_i b) = \frac{1}{1 + \exp(-X_i b)}$$

```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

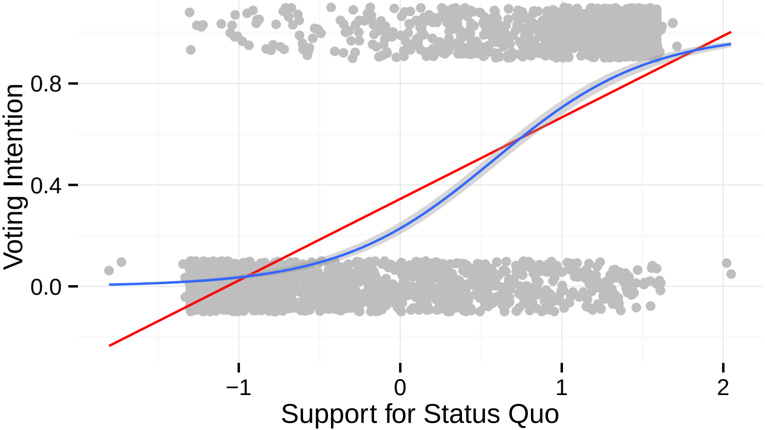


```r
summary(glm(vote_yes ~ statusquo, data = Chile,
            family = binomial(link = "logit")))
```

```
##
## 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

In linear regression, $\partial Y / \partial X_j = \beta_j$

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

In logistic regression, $\partial Y / \partial X_j = \beta_j$

$$\frac{\partial \,\mathsf{Pr}(Y_i = 1)}{\partial X_j} = \frac{\partial}{\partial X_j} \frac{1}{1 + \mathsf{exp}(\alpha + \beta X_i)} = \mathsf{Pr}(Y = 1 | X_i)\, \mathsf{Pr}(Y = 0 | X_i) \beta_j$$

or

$$\frac{\partial logit(Y_i = 1)}{\partial X_j} = \frac{\partial}{\partial X_j} X_i \beta_j = \beta_j$$

- Unlike OLS, the partial derivative depends on value of X_i

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
- It is biased if functional form is wrong; but so it logit / probit. And the functional form is **always** wrong
- If you care about **average marginal effects** OLS does well

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

- Angrist and Pischke recommend LPM with heteroskedasticity consistent errors

Comparing Average Marginal Effects of Logit and LPM

1992 U.S. Election Example

```r
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.