MPhil Econometrics – Limited Dependent Variables and Selection

Francis J. DiTraglia

University of Oxford

Complied on 2020-02-03 at 15:13:07

Housekeeping

Lecturer:	Francis J. DiTraglia
Email:	francis.ditraglia@ox.ac.uk
Course Materials:	`http://ditraglia.com/teaching`
Office:	2132 Manor Road Building
Meetings:	after lecture on Wednesdays & Fridays or by appointment

References

- **Wooldridge (2010) – *Econometric Analysis of Cross Section & Panel Data***
- Cameron & Trivedi (2005) – *Microeconometrics: Methods and Applications*
- Train (2009) – *Discrete Choice Methods with Simulation*

Lecture #3 – Models for Binary Outcomes

Properties of Binary Outcome Models

Linear Probability Model

Index Models (e.g. Logit & Probit)

Partial Effects

Conditional MLE for Index Models

Pseudo R-squared

Models for Binary Outcomes

Example

- Outcome: $y = 1$ if employed, 0 otherwise
- Predictors/Regressors: $\mathbf{x} = \{\text{age, sex, education, experience, ...}\}$

Question

How does x_j affect our prediction of y holding the other regressors constant?

We'll consider three models:

1. Linear Probability Model (LPM)
2. Logistic Regression (Logit)
3. Probit Regression (Probit)

Properties of Binary Outcome Models: $y \in \{0, 1\}$

Notation

$p(\mathbf{x}) \equiv \mathbb{P}(y = 1|\mathbf{x})$

Conditional Mean

$\mathbb{E}(y|\mathbf{x}) = p(\mathbf{x})$

Conditional Variance

$\text{Var}(y|\mathbf{x}) = p(\mathbf{x})\,[1 - p(\mathbf{x})]$

$$
\begin{aligned}
\mathbb{E}(y|\mathbf{x}) &= 0 \times \mathbb{P}(y = 0|\mathbf{x}) + 1 \times \mathbb{P}(y = 1|\mathbf{x}) \\
&= \mathbb{P}(y = 1|\mathbf{x}) \equiv p(\mathbf{x})
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(y^2|\mathbf{x}) &= \left\{0^2 \times [1 - p(\mathbf{x})] + 1^2 \times p(\mathbf{x})\right\} \\
&= p(\mathbf{x})
\end{aligned}
$$

$$
\begin{aligned}
\text{Var}(y|\mathbf{x}) &= \mathbb{E}(y^2|\mathbf{x}) - \mathbb{E}(y|\mathbf{x})^2 \\
&= \left\{0^2 \times [1 - p(\mathbf{x})] + 1^2 \times p(\mathbf{x})\right\} - p(\mathbf{x})^2 \\
&= p(\mathbf{x})\,[1 - p(\mathbf{x})]
\end{aligned}
$$

The Linear Probability Model: Assume $p(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$

Conditional Mean & Variance

- $\mathbb{E}(y|\mathbf{x}) = p(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$
- $\text{Var}(y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}\,(1 - \mathbf{x}'\boldsymbol{\beta})$

This is just Linear Regression!

$$y = \mathbf{x}'\boldsymbol{\beta} + u, \quad \mathbb{E}(u|\mathbf{x}) = 0$$

But u is Heteroskedastic

$$\text{Var}(u|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}(1 - \mathbf{x}'\boldsymbol{\beta})$$

$$\begin{aligned}
\mathbb{E}(u|\mathbf{x}) &= \mathbb{E}(y - \mathbf{x}'\boldsymbol{\beta}|\mathbf{x}) = \mathbb{E}(y|\mathbf{x}) - \mathbf{x}'\boldsymbol{\beta} \\
&= \mathbf{x}'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta} = 0
\end{aligned}$$

$$\begin{aligned}
\text{Var}(u|\mathbf{x}) &= \mathbb{E}\left[\{u - \mathbb{E}(u|\mathbf{x})\}^2 |\mathbf{x}\right] = \mathbb{E}\left[u^2|\mathbf{x}\right] \\
&= \mathbb{E}\left[(y - \mathbf{x}'\boldsymbol{\beta})^2 |\mathbf{x}\right] \\
&= \mathbb{E}\left(y^2|\mathbf{x}\right) - 2\mathbb{E}\left(y|\mathbf{x}\right)\mathbf{x}'\boldsymbol{\beta} + \left(\mathbf{x}'\boldsymbol{\beta}\right)^2 \\
&= p(\mathbf{x}) - 2p(\mathbf{x})p(\mathbf{x}) + p(\mathbf{x})^2 \\
&= p(\mathbf{x})\left[1 - p(\mathbf{x})\right]
\end{aligned}$$

The Linear Probability Model: Assume $p(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$

Estimation

Since $\mathbb{E}(u|\mathbf{x}) = 0$ OLS estimation of $y = \mathbf{x}'\boldsymbol{\beta} + u$ is unbiased and consistent.

Inference

Since u is heteroskedastic, tests and CIs should use robust standard errors.

Is the LPM actually reasonable?

- Assumes $p(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta} \implies$ changing x_j by Δ changes $p(\mathbf{x})$ by $\beta_j\Delta$
- If $\mathbf{x}$ contains regressors without upper/lower bounds, $p(\mathbf{x})$ could be > 1 or < 0!
- LPM could be a reasonable approximation but cannot be *literally* true.

Index Models: Constrain $p(\mathbf{x})$ to lie in $[0, 1]$

Index Model: $p(\mathbf{x}) = G(\mathbf{x}'\boldsymbol{\beta})$

Assume $\mathbf{x}$ includes a constant, $0 \leq G(\cdot) \leq 1$, G is differentiable and strictly increasing, $\lim_{z\to\infty} G(z) = 1$, and $\lim_{z\to-\infty} G(z) = 0$.

Terminology

We call $\mathbf{x}'\boldsymbol{\beta}$ the linear index and G the index function.

Partial Effects

Let $g(z) \equiv \frac{d}{dz}G(z)$. Then $\frac{\partial}{\partial x_j}p(\mathbf{x}) = g(\mathbf{x}'\boldsymbol{\beta})\beta_j$. Hence:

- The partial effect of x_j depends on the value of $\mathbf{x}$ at which we evaluate g.
- G strictly increasing $\implies$ $g(\cdot) > 0$ $\implies$ sign of partial effect determined by β_j.

Possible Choices of Index Function

CDFs as Index Functions

G satisfies the index model assumptions (prev. slide) iff it is a continuous CDF.

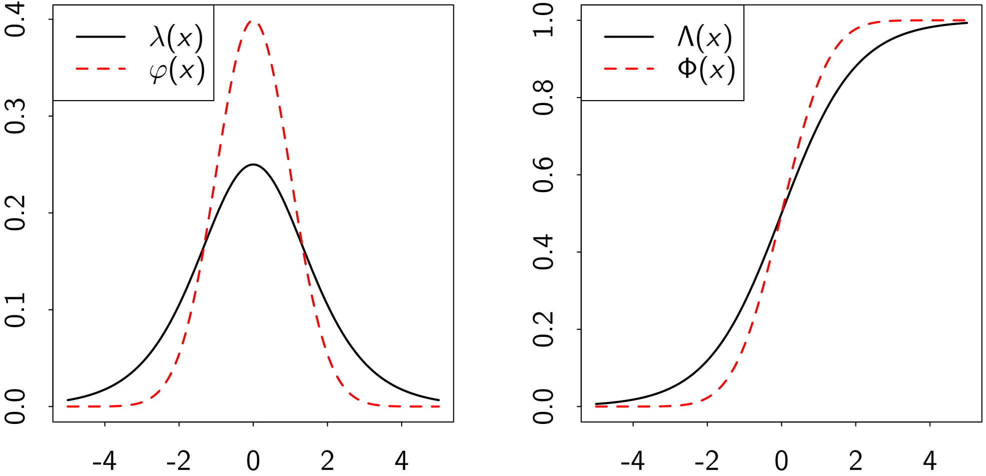
We focus on two examples:

1. Logit: $G(z) = \Lambda(z) \equiv \exp(z) / [1 + \exp(z)]$
2. Probit: $G(z) = \Phi(z) \equiv \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-t^2/2\right)\, dt$

Notation:

- Λ is the CDF of a "standard logistic" RV and Φ of a standard normal RV.
- λ is the density of a "standard logistic" RV and φ of a standard normal
- To treat Logit and Probit simultaneously, we'll write G as a placeholder.

Standard $\boxed{\text{Logistic}}$ and $\boxed{\text{Normal}}$ Densities and CDFs



Partial Effects: $\partial p(\mathbf{x})/\partial x_j$

$$\frac{\partial}{\partial x_j} G(\mathbf{x}'\boldsymbol{\beta}) = g(\mathbf{x}'\boldsymbol{\beta})\beta_j$$

LPM

$$\frac{\partial}{\partial x_j}\mathbf{x}'\boldsymbol{\beta} = \beta_j$$

$$\frac{d}{dz}\Lambda(z) \equiv \lambda(z) = \frac{d}{dz}\left(\frac{e^z}{1+e^z}\right) = \frac{e^z(1+e^z) - e^z e^z}{(1+e^z)^2} = \frac{e^z}{(1+e^z)^2}$$

Logit

$$\frac{\partial}{\partial x_j}\Lambda(\mathbf{x}'\boldsymbol{\beta}) = \frac{\beta_j \exp(\mathbf{x}'\boldsymbol{\beta})}{[1+\exp(\mathbf{x}'\boldsymbol{\beta})]^2}$$

Probit

$$\frac{\partial}{\partial x_j}\Phi(\mathbf{x}'\beta) = \frac{\beta_j \exp\{-(\mathbf{x}'\boldsymbol{\beta})^2/2\}}{\sqrt{2\pi}}$$

$$\frac{d}{dz}\Phi(z) = \varphi(z) = \frac{\exp\left\{-z^2/2\right\}}{\sqrt{2\pi}}$$

Comparing Logit, Probit, and LPM Partial Effects

$$\boxed{\frac{\partial}{\partial x_j}G(\mathbf{x}'\boldsymbol{\beta}) = g(\mathbf{x}'\boldsymbol{\beta})\beta_j, \quad \frac{d}{dz}\Lambda(z) \equiv \lambda(z) = \frac{e^z}{(1+e^z)^2}, \quad \frac{d}{dz}\Phi(z) \equiv \varphi(z) = \frac{\exp\left\{-z^2/2\right\}}{\sqrt{2\pi}}}$$

Maximum Partial Effects

- λ and φ are unimodal with mode at 0

  Logit $\lambda(0) = 0.25$

  Probit $\varphi(0) = (2\pi)^{-1/2} \approx 0.4$

- *Maximum* partial effect when $\mathbf{x}'\boldsymbol{\beta} = 0$

  Logit $\beta_j\lambda(0) = 0.25\beta_j$

  Probit $\beta_j\varphi(0) \approx 0.4\beta_j$

- LPM has constant partial effects β_j

Relative Effects

$$\frac{\frac{\partial}{\partial x_j}p(\mathbf{x})}{\frac{\partial}{\partial x_h}p(\mathbf{x})} = \frac{\beta_j g(\mathbf{x}'\boldsymbol{\beta})}{\beta_h g(\mathbf{x}'\boldsymbol{\beta})} = \frac{\beta_j}{\beta_h}$$

Relative effects do not depend on $\mathbf{x}$.

Average Partial Effects for Index Models

Partial Effect

$$\frac{\partial}{\partial x_j} G(\mathbf{x}'\boldsymbol{\beta}) = g(\mathbf{x}_i'\boldsymbol{\beta})\beta_j$$

Average Partial Effect

$$\mathbb{E}\left[\frac{\partial}{\partial x_j} G(\mathbf{x}'\boldsymbol{\beta})\right] = \mathbb{E}[g(\mathbf{x}_i'\boldsymbol{\beta})]\beta_j$$

Estimated Partial Effect

$$\frac{\partial}{\partial x_j} G(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}) = g(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})\widehat{\beta}_j$$

Estimated Average Partial Effect

$$\left[\frac{1}{N}\sum_{i=1}^{N} g(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})\right]\widehat{\beta}_j$$

Warning:

APE $\neq$ partial effect evaluated at the average value of $\mathbf{x}$ since $\mathbb{E}[f(Z)] \neq f(\mathbb{E}[Z])$.

Conditional MLE for Index Models: iid Observations

Conditional Likelihood

$$f(y_i|\mathbf{x}_i,\boldsymbol{\beta}) = \begin{cases} 1 - G(\mathbf{x}_i'\boldsymbol{\beta}) & \text{if } y_i = 0 \\ G(\mathbf{x}_i'\boldsymbol{\beta}) & \text{if } y_i = 1 \end{cases} \iff f(y_i|\mathbf{x}_i,\boldsymbol{\beta}) = G(\mathbf{x}_i'\boldsymbol{\beta})^{y_i}\left[1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right]^{1-y_i}$$

Conditional Log-Likelihood

$$\ell_i(\boldsymbol{\beta}) \equiv \log f(y_i|\mathbf{x}_i,\boldsymbol{\beta}) = y_i \log\left[G(\mathbf{x}_i'\boldsymbol{\beta})\right] + (1 - y_i)\log\left[1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right]$$

Sample

$$\widehat{\boldsymbol{\beta}} \equiv \underset{\boldsymbol{\beta}\in\boldsymbol{\Theta}}{\arg\max} \frac{1}{N}\sum_{i=1}^{N} \ell_i(\boldsymbol{\beta})$$

Population

$$\boldsymbol{\beta}_o \equiv \underset{\boldsymbol{\beta}\in\boldsymbol{\Theta}}{\arg\max}\, \mathbb{E}\left[\ell(\boldsymbol{\beta})\right]$$

Correct specification: $\mathbb{E}(y|\mathbf{x}) = p(\mathbf{x}) = G(\mathbf{x}'\boldsymbol{\beta}_o)$. Otherwise $\boldsymbol{\beta}_o$ = KL-minimizer.

Asymptotic Variance Calculations for Index Models

Recall from last lecture.

Possibly Mis-specified Model

$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \rightarrow_d \mathcal{N}(\mathbf{0}, \mathbf{J}^{-1}\mathbf{K}\mathbf{J}^{-1})$ where $\mathbf{J} = -\mathbb{E}\left[\mathbf{H}_i(\boldsymbol{\beta}_o)\right]$ and $\mathbf{K} = \mathbb{E}\left[\mathbf{s}_i(\boldsymbol{\beta}_o)\mathbf{s}_i(\boldsymbol{\beta}_o)'\right]$

Correct Specification

$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \rightarrow_d \mathcal{N}(\mathbf{0}, \mathbf{J}^{-1})$ where $\mathbf{J} = -\mathbb{E}\left[\mathbf{H}_i(\boldsymbol{\beta}_o)\right]$

Asymptotic variance calculations for index models are complicated, but there's a clever trick for computing $\mathbf{J}$ under correct specification.

Asymptotic Variance Calculation for Correctly Specified Index Models

$$\boxed{\ell_i(\boldsymbol{\beta}) = y_i \log\left\{G(\mathbf{x}_i'\boldsymbol{\beta})\right\} + (1-y_i)\log\left\{1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right\}}$$

Step 1: Calculate The Score Vector

$$\mathbf{s}_i \equiv \frac{\partial}{\partial \boldsymbol{\beta}}\ell_i(\boldsymbol{\beta}) = \frac{y_i g(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i}{G(\mathbf{x}_i'\boldsymbol{\beta})} - \frac{(1-y_i)g(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i}{1 - G(\mathbf{x}_i'\boldsymbol{\beta})}$$

$$= \frac{g(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i}{G(\mathbf{x}_i'\boldsymbol{\beta})\left[1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right]}\left\{\left[1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right] - G(\mathbf{x}_i'\boldsymbol{\beta})(1-y_i)\right\}$$

$$= \frac{g(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i\left[y_i - G(\mathbf{x}_i'\boldsymbol{\beta})\right]}{G(\mathbf{x}_i'\boldsymbol{\beta})\left[1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right]}$$

Asymptotic Variance Calculation for Correctly Specified Index Models

$$\boxed{\mathbf{s}_i = \frac{g(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i \left\{y_i - G(\mathbf{x}_i'\boldsymbol{\beta})\right\}}{G(\mathbf{x}_i'\boldsymbol{\beta})\left\{1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right\}}}$$

Step 2: Start Calculating the Hessian but give up because it's a nightmare.

$$\mathbf{H}_i(\boldsymbol{\beta}) \equiv \frac{\partial \mathbf{s}_i}{\partial \boldsymbol{\beta}'} = \frac{\partial}{\partial \boldsymbol{\beta}}\left([y_i - G(\mathbf{x}_i'\boldsymbol{\beta})]\left[\frac{g(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i}{G(\mathbf{x}_i'\boldsymbol{\beta})\left\{1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right\}}\right]\right)$$

$$= \frac{-g(\mathbf{x}_i'\boldsymbol{\beta})^2\mathbf{x}_i\mathbf{x}_i'}{G(\mathbf{x}_i'\boldsymbol{\beta})\left\{1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right\}} + [y_i - G(\mathbf{x}_i'\boldsymbol{\beta})]\underbrace{\frac{\partial}{\partial \boldsymbol{\beta}'}\left\{\frac{g(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i}{G(\mathbf{x}_i'\boldsymbol{\beta})\left[1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right]}\right\}}_{\text{a nasty awful mess: call it } \mathbf{M}(\mathbf{x}_i,\boldsymbol{\beta})}$$

Asymptotic Variance Calculation for Correctly Specified Index Models

$$\boxed{\mathbf{H}_i(\boldsymbol{\beta}) = \frac{-g(\mathbf{x}_i'\boldsymbol{\beta})^2\mathbf{x}_i\mathbf{x}_i'}{G(\mathbf{x}_i'\boldsymbol{\beta})\left\{1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right\}} + \left[y_i - G(\mathbf{x}_i'\boldsymbol{\beta})\right]\mathbf{M}(\mathbf{x}_i, \boldsymbol{\beta})}$$

Step 3: Calculate the *Conditional Expectation* instead...

$$\begin{aligned}
-\mathbb{E}\left[\mathbf{H}_i(\boldsymbol{\beta})|\mathbf{x}_i\right] &= \frac{g(\mathbf{x}_i'\boldsymbol{\beta})^2\mathbf{x}_i\mathbf{x}_i'}{G(\mathbf{x}_i'\boldsymbol{\beta})\left\{1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right\}} + \underbrace{\mathbb{E}\left[y_i - G(\mathbf{x}_i'\boldsymbol{\beta})|\mathbf{x}_i\right]}_{\text{equals zero under correct spec.}} \mathbf{M}(\mathbf{x}_i, \boldsymbol{\beta}) \\
&= \frac{g(\mathbf{x}_i'\boldsymbol{\beta})^2\mathbf{x}_i\mathbf{x}_i'}{G(\mathbf{x}_i'\boldsymbol{\beta})\left\{1 - G(\mathbf{x}_i'\boldsymbol{\beta})\right\}}
\end{aligned}$$

Step 4: Iterated Expectations

$$\mathbf{J} = -\mathbb{E}\left[\mathbf{H}_i(\boldsymbol{\beta}_o)\right] = \mathbb{E}\left\{\mathbb{E}\left[\mathbf{H}_i(\boldsymbol{\beta}_o)|\mathbf{x}_i\right]\right\} = \mathbb{E}\left\{\frac{g(\mathbf{x}_i'\boldsymbol{\beta}_o)^2\mathbf{x}_i\mathbf{x}_i'}{G(\mathbf{x}_i'\boldsymbol{\beta}_o)\left\{1 - G(\mathbf{x}_i'\boldsymbol{\beta}_o)\right\}}\right\}$$

Asymptotic Variance Calculation for Correctly Specified Index Models

Asymptotic Distribution

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \to_d \mathcal{N}\left(\mathbf{0}, \mathbf{J}^{-1}\right), \quad \mathbf{J}^{-1} = \mathbb{E}\left\{\frac{g(\mathbf{x}_i'\boldsymbol{\beta}_o)^2 \mathbf{x}_i \mathbf{x}_i'}{G(\mathbf{x}_i'\boldsymbol{\beta}_o)\{1 - G(\mathbf{x}_i'\boldsymbol{\beta}_o)\}}\right\}$$

Consistent Estimator

$$\widehat{\mathbf{J}}^{-1} \equiv \left\{\frac{1}{N}\sum_{i=1}^{N} \frac{g(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})^2 \mathbf{x}_i \mathbf{x}_i'}{G(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})\left[1 - G(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})\right]}\right\}^{-1}$$

Notes

- Assumes correct specification, i.e. $p(\mathbf{x}) = \mathbb{E}(y|\mathbf{x}) = G(\mathbf{x}'\boldsymbol{\beta}_o)$
- In contrast, *robust* variance matrix $\mathbf{J}^{-1}\mathbf{K}\mathbf{J}^{-1}$ is complicated, but R can do it.

McFadden (1974) – "Pseudo R-squared"

Model with Intercept Only

$L(\bar{y}) \equiv$ maximized sample Likelihood

$\ell(\bar{y}) \equiv$ maximized sample log-likelihood

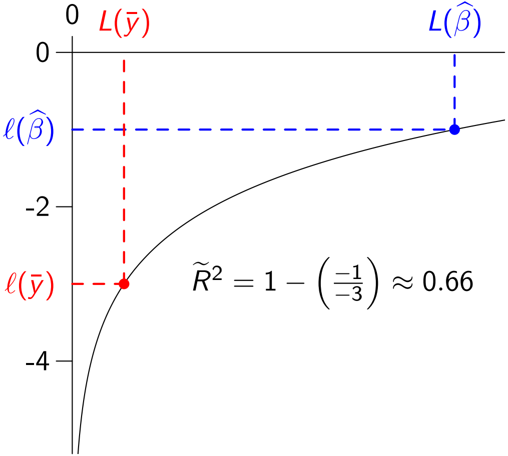
Full Model

$L(\widehat{\beta}) \equiv$ maximized sample Likelihood

$\ell(\widehat{\beta}) \equiv$ maximized sample log-likelihood

Pseudo R-squared

$\widetilde{R}^2 \equiv 1 - \ell(\widehat{\beta})/\ell(\bar{y})$



McFadden (1974) – "Pseudo R-squared"

Pseudo R-squared

$\widetilde{R}^2 \equiv 1 - \ell(\widehat{\beta})/\ell(\bar{y})$

Always between 0 and 1

Show this on the problem set!

Health Warning

I don't recommend using pseudo-R^2: it's arbitrary and can be misleading. Other people use it so I'm telling you what it is.