Econ 312

Wednesday, April 22

Limited Dependent Variables: Probit and Logit

Readings: Wooldridge, Section 17.1

Class notes: 154 - 159

Today's Far Side offering

"Mr. Osborne, may I be excused? My brain is full."

How we're all feeling at this time of year!

Context and overview

- The final major section of the course deals with dependent variables that have limited ranges, not $-\infty$ to $+\infty$
- This class looks in detail at models of a **dummy dependent variable**
  - **Linear probability model** is simple
  - **Probit and logit** models are more statistically reasonable, but require careful interpretation of the coefficients
- In the next few classes we will examine other situations in which the dependent variable is limited in range or discontinuous

Linear probability model

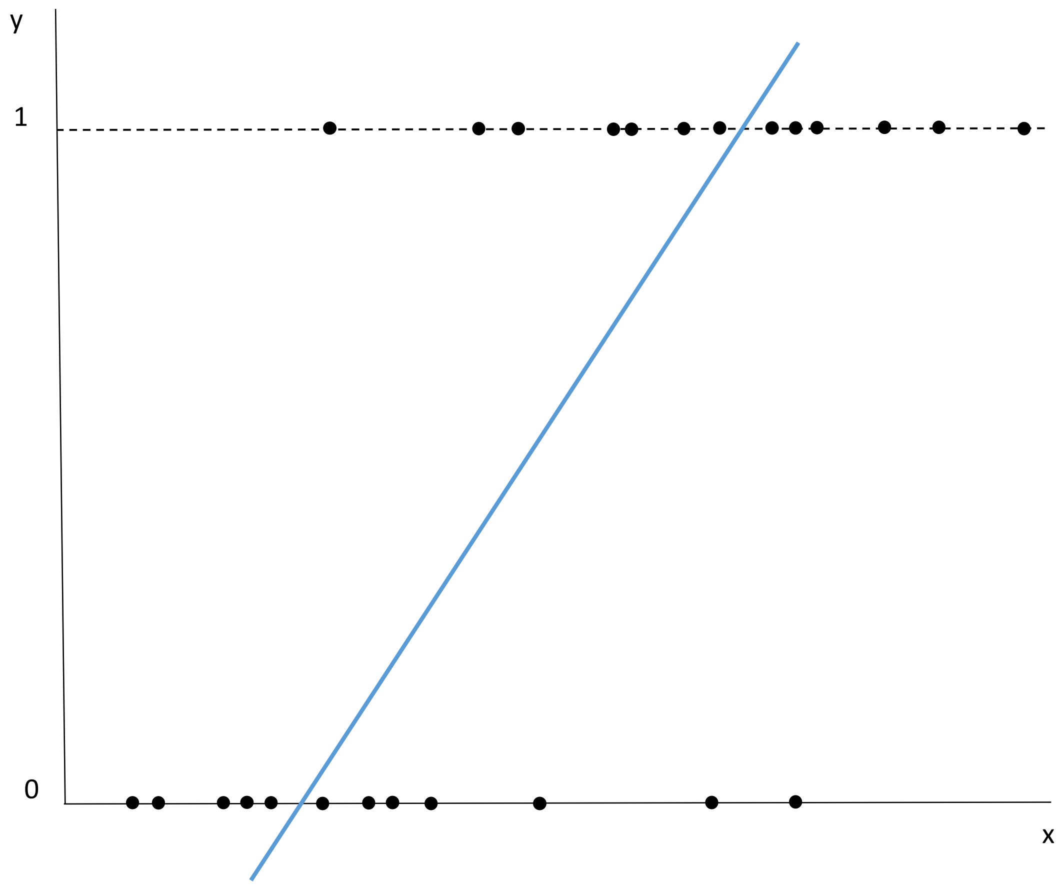
- $y = 0$ or 1, so

$$E(y_i \mid x_i) = \Pr[y_i = 1 \mid x_i]$$

- Linear probability model (LPM) just applies OLS by making this a **linear function** of the x variables:

$$\Pr[y_i = 1 \mid x_i] = E(y_i \mid x_i) = \beta_0 + \beta_1 x_i$$

- Problem #1: **line doesn't fit data well**



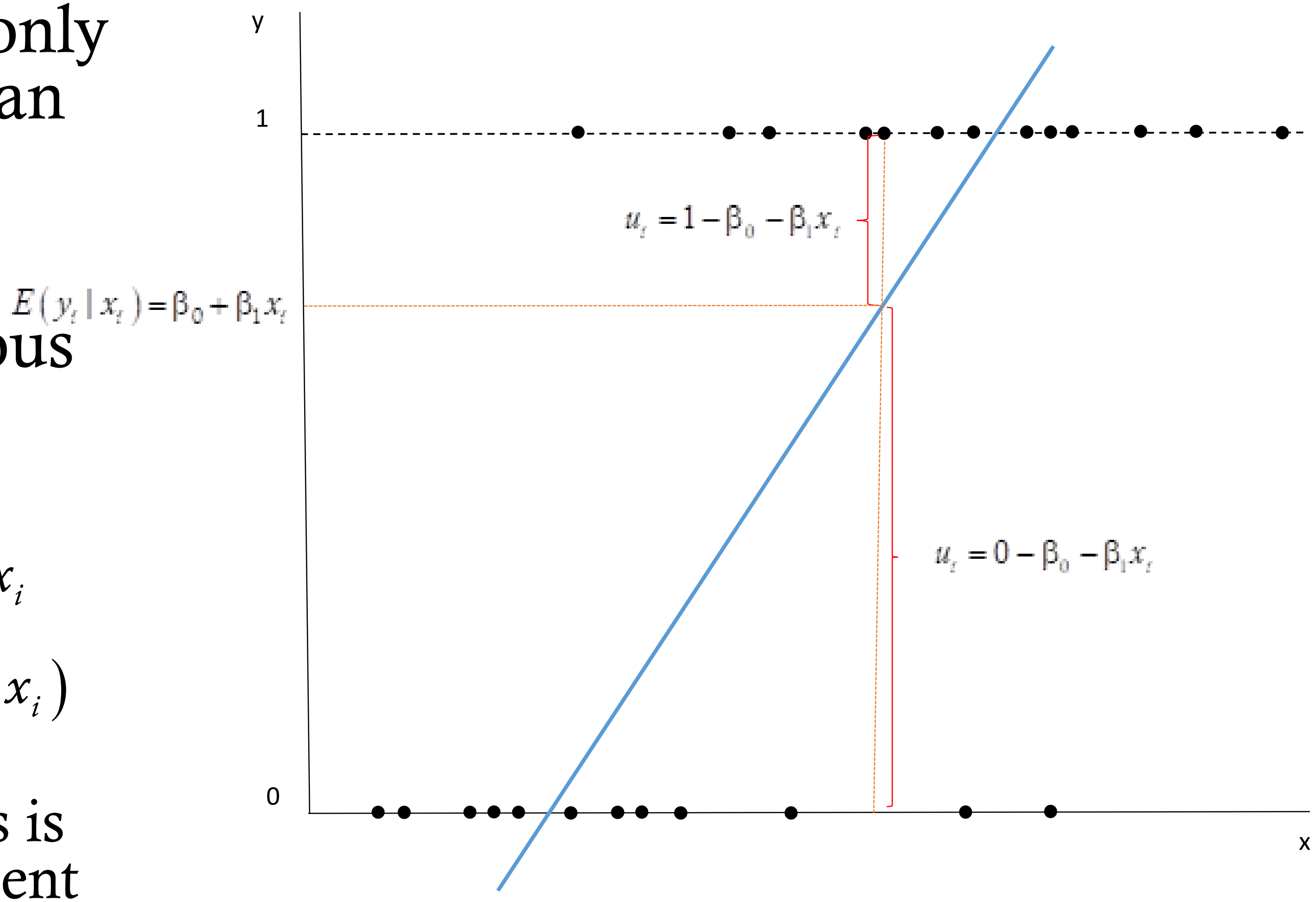
Error term in LPM

- Problem #2: Since y can only be 0 or 1, the error term can only be $1-\beta_0-\beta_1 x_i$ or $0-\beta_0-\beta_1 x_i$
- u is **discrete**, not continuous
  - Bernoulli distribution, not normal

$$\Pr\left[u_i = 1-\left(\beta_0+\beta_1 x_i\right)\right] = \beta_0+\beta_1 x_i$$

$$\Pr\left[u_i = -\left(\beta_0+\beta_1 x_i\right)\right] = 1-\left(\beta_0+\beta_1 x_i\right)$$

  - Sum of Bernoulli variables is normal in limit, so coefficient estimates may still be normal

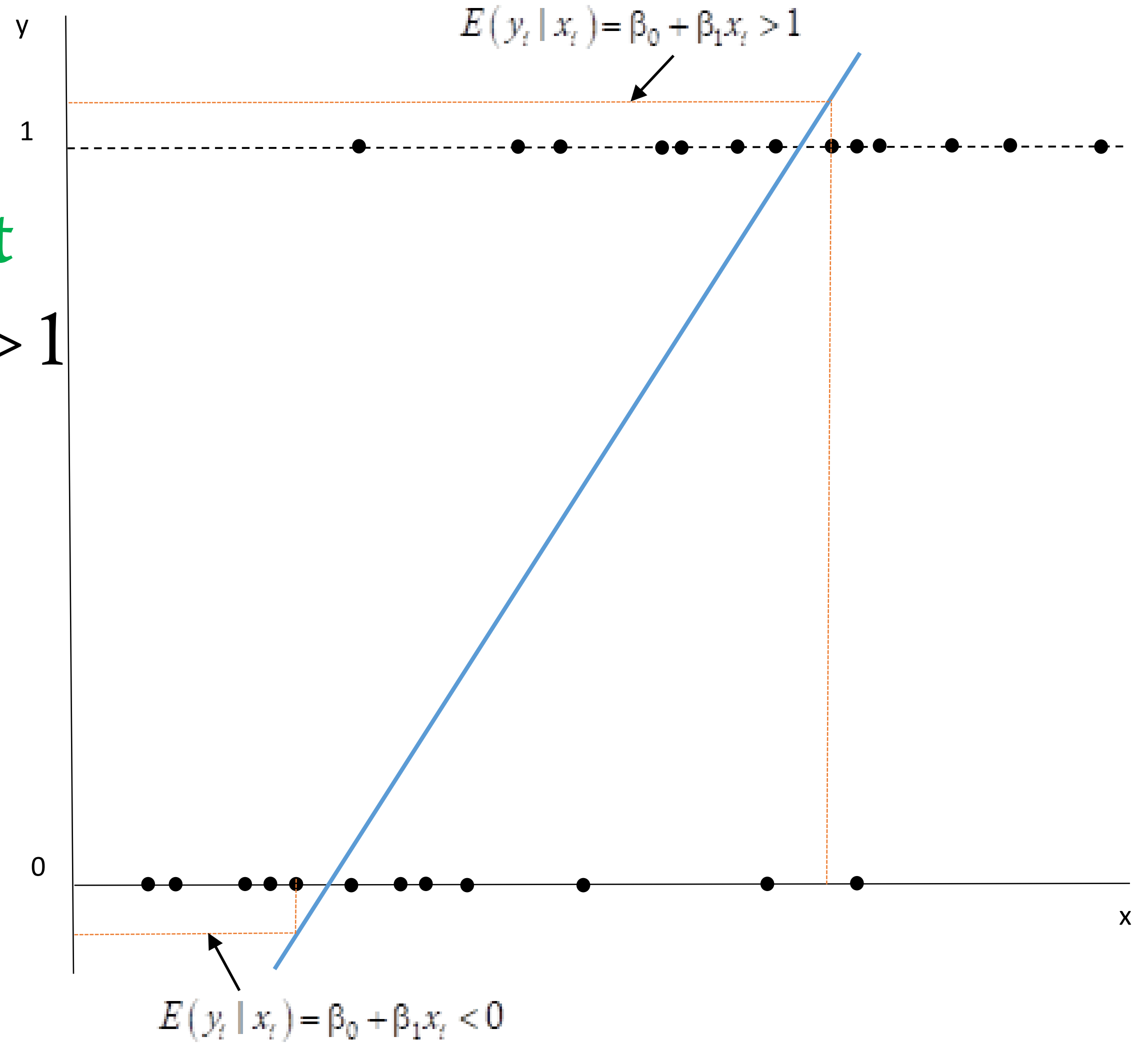


Prediction in LPM

- Problem #3: For extreme values of x we always **predict**

$$\Pr[y=1\,|\,x]<0 \text{ or } \Pr[y=1\,|\,x]>1$$

- Also has heteroskedasticity
- Bottom line:
  - LPM is **simple**
  - Might be **usable** for x close to sample mean
  - Simply **not the best model**!



Alternative models: Probit and logit

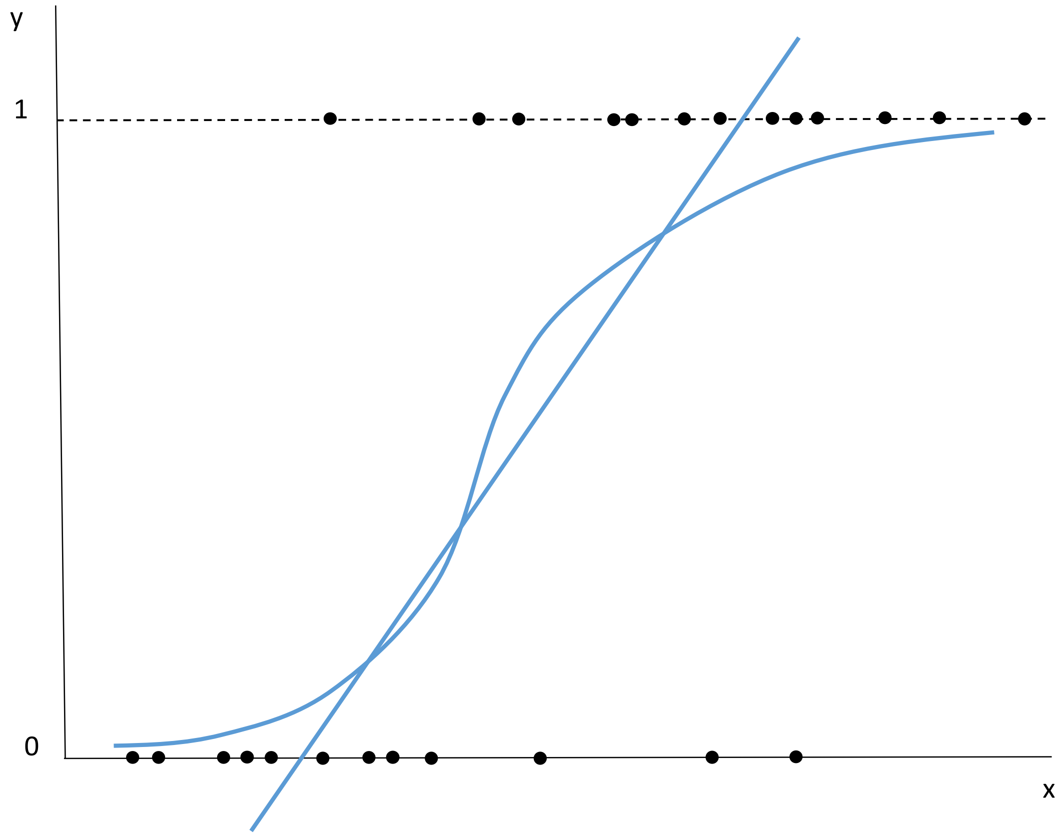
- $E(y_i \mid x) = \Pr[y_i = 1 \mid x] = G[\beta_0 + \beta_1 x_i] = G(z_i)$
- **Probit** fits cumulative normal distribution function

$$G(z_i) = \Phi(z_i) = \int_{-\infty}^{z_i} \phi(\zeta) d\zeta = \int_{-\infty}^{z_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\zeta^2} d\zeta$$

- **Logit** fits cumulative logistic distribution function

$$G(z_i) = \Lambda(z_i) = \frac{1}{1+e^{-z_i}} = \frac{e^{z_i}}{1+e^{z_i}}$$

- Similar shapes that fit the 0/1 data better than linear



Estimation of probit and logit

- Nonlinear **maximum likelihood**
  - Discrete density function:

    $$\Pr[y_i = 1 \mid x_i, \beta] = G(x_i\beta),\ \Pr[y_i = 0 \mid x_i, \beta] = 1 - G(x_i\beta)$$

  - Or $f(y_i \mid x_i, \beta) = [G(x_i\beta)]^{y_i}[1 - G(x_i\beta)]^{(1-y_i)}$, $y_i = 0, 1$
- **Log likelihood function** with IID sample:

  $$\ln L(\beta; y, x) = \sum_{i=1}^{n}\left[y_i \ln[G(x_i\beta)] + (1 - y_i)\ln[1 - G(x_i\beta)]\right].$$

- Choose β, evaluate ln*L*, then **search** over β to find maximum
- Estimator is **consistent, asymptotically normal/efficient**

Goodness of fit and hypothesis tests

- Goodness of fit always looks bad:
  - **Fraction predicted correctly**, predicting 1 for $G\left(x_i\hat{\beta}\right) > 0.5$
  - **Pseudo-R^2**:

$$1 - \frac{\ln L\left(\hat{\boldsymbol{\beta}}; x, y\right)}{\ln L\left(\boldsymbol{\beta}_Z; x, y\right)}, \boldsymbol{\beta}_Z{}' \equiv \left(\bar{y}, 0, 0, \ldots, 0\right)$$

- **Likelihood-ratio test**: $2\left[\ln L_u - \ln L_r\right] \sim \chi^2_q$.
- Can also do the **standard *t* test**:

$$t = \frac{\hat{\beta}_j - c}{se\left(\hat{\beta}_j\right)} \sim t_{(n-k-1)}$$

Interpretation of coefficients

- In OLS (or LPM): $\beta_j = \dfrac{\partial E[y_i \mid x_i]}{\partial x_j}$
- In probit or logit: $\beta_j = \dfrac{\partial z}{\partial x_j}$ with no logical interpretation
  - Remember that we define $z = x\beta$
- For continuous regressor, we want

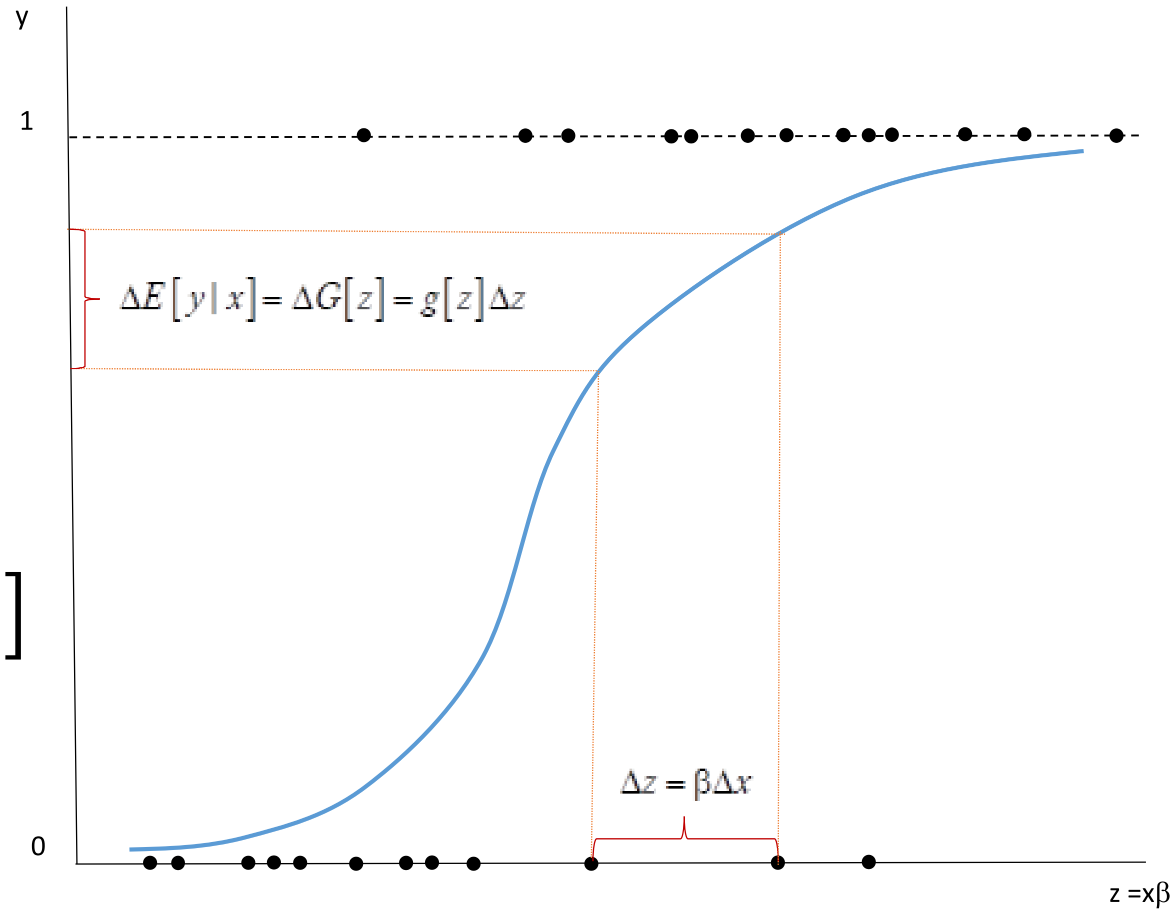
$$\frac{\partial \Pr[y=1]}{\partial x_j} = \frac{d \Pr[y=1]}{dz} \frac{\partial z}{\partial x_j} = G'(z)\beta_j = g(z)\beta_j,$$

- β measures effect of x on z, $g(z)$ measures effect of z on $\Pr[y = 1]$

Geometric interpretation of coefficients

- Put $z = x\beta$ on horizontal axis
- Increase of Δx units in x → change of $\Delta z = \beta\Delta x$ units in z
- Change of Δz units in z → change of $\Delta G[z] = g[z]\Delta z$ units in $E[y|x]$
- So partial effect of x on $\Pr[y=1]$ is

$$\frac{\partial \Pr[y=1]}{\partial x_j} = g(z)\beta_j$$



Odds ratio in logit

$$\Pr[y_i = 1 | x_i] = \Lambda(x_i\beta) = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}}$$

$$(1 + e^{x_i\beta})\Lambda(x_i\beta) = e^{x_i\beta}$$

$$\Lambda(x_i\beta) = e^{x_i\beta} - \Lambda(x_i\beta)e^{x_i\beta} = (1 - \Lambda(x_i\beta))e^{x_i\beta}$$

$$e^{x_i\beta} = \frac{\Lambda(x_i\beta)}{1 - \Lambda(x_i\beta)} = \frac{\Pr[y_i = 1 | x_i]}{\Pr[y_i = 0 | x_i]} = \text{"odds ratio"}$$

- β is the effect of *x* on the "log odds ratio"
- Stata table reports e^{β} as proportional effect of *x* on odds
  - Always > 0, $e^{\beta} > 1$ → *x* increases odds, $e^{\beta} < 1$ → *x* decreases odds
  - Depends on *x*: set to mean

Partial effects in probit

- Probit uses normal rather than logistic distribution

$$\frac{\partial \Pr\left[y=1\right]}{\partial x_j} = \Phi'\left(z\right)\beta_j = \phi\left(z\right)\beta_j = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(x_i\beta\right)^2} \beta_j$$

- Again, this depends on x, so we typically evaluate the effect at the means of the regressors
- Partial effects in both probit and logit have same sign as coefficient and can be tested by $\beta_j = 0$
- If x is a dummy, we want $\Pr\left[y=1 \mid x_j = 1\right] - \Pr\left[y=1 \mid x_j = 0\right]$

Probit and logit in Stata

Probit

- probit reports coefficients
- dprobit reports partial effects (or effects of 0 → 1 for dummies) evaluated at means of all x
- Use margins command to evaluate at other x
- Same t statistics and tests from both

Logit

- logit reports coefficients
- logistic reports effect of x on log odds ratio at means
- Remember that zero effect means $e^{\beta} = 1$
- Tricky to interpret; see notes

Issues in probit and logit estimation

- **Nonlinearity** makes finding best estimator less reliable
  - Algorithm can break down in cases of high multicollinearity
  - Complex models can take a long time to converge
- **Omitted-variable bias** affects all coefficients, even if x_j not correlated
- **Heteroskedasticity** makes estimator inconsistent
  - White's robust standard error fixes the standard error, but not the coefficient
  - Try to rescale the model to reduce probability of heteroskedasticity

Review and summary

- When the dependent variable is a 0/1 dummy we have several choices of estimators
  - **Linear probability model** is simple, but unrealistic
  - **Probit** and **logit** are better suited to the situation
- Both probit and logit use **cumulative probability distribution functions** to approximate relationship instead of straight line
- Must be estimated by **nonlinear least squares**
- Coefficients no longer have the usual interpretations
- Stata can transform coefficients into meaningful "**partial effects**"

Another bad economist joke …

"Let us remember the unfortunate econometrician who, in one of the major functions of his system, had to use a proxy for risk and a dummy for sex."

-- Fritz Machlup

--Taken from Jeff Thredgold, *On the One Hand: The Economist's Joke Book*

What's next?

In the next class, we discuss estimators appropriate to other unusual dependent variables:

- Generalizing probit/logit to **more than two choices** (0/1/2, for example)
- Models for **ordered dependent variables** (A > B > C)
- Models for **count dependent variables** (0, 1, 2, 3, …)