Zombie Econometrics: The Linear Probability Model

William Greene
University of South Florida
and
Paul Wilson
Clemson University

http://people.stern.nyu.edu/wgreene/zombie-econometrics.pdf

Econometric Idea:

The Linear Probability Model

Binary outcome,
Linear regression: e.g., Labor force participation given “treatment”
(1) $y \in \{0,1\}$; $E[y|x] = \alpha + \beta x$

Linear probability of binary outcome
(2) $\text{Prob}(y = 1|x) = \alpha + \beta x$

?

Additive disturbance with the usual properties
(3) $y = \alpha + \beta x + \varepsilon$; $E[\varepsilon|x] = 0$

Linear least squares regression, estimation and inference
(4) $\mathbf{b} = [\mathbf{X'X}]^{-1}\mathbf{X'y}$ The usual robust covariance matrix.

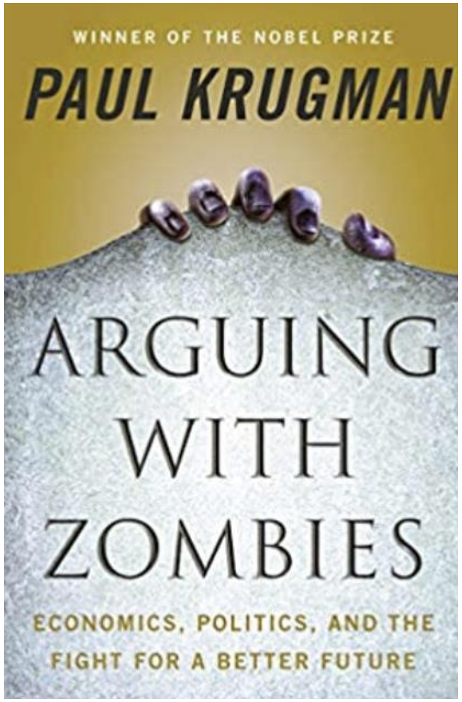
Zombie Ideas

According to Nobel Laureate Paul Krugman, a zombie idea is "a proposition that has been thoroughly refuted by analysis and evidence and should be dead – but won't stay dead …

[*The Skeptic's Dictionary, (2022)*.]

Political Zombie Idea: Tax cuts for wealthy individuals pay for themselves.

Econometric Zombie Idea: The Linear Probability Model


WINNER OF THE NOBEL PRIZE
PAUL KRUGMAN
ARGUING WITH ZOMBIES
ECONOMICS, POLITICS, AND THE FIGHT FOR A BETTER FUTURE

Social Science Econometric Modeling

- **Pre- 1970s** Usually (log) linear regression and linear simultaneous equations models of **measures** of the micro- or macroeconomy

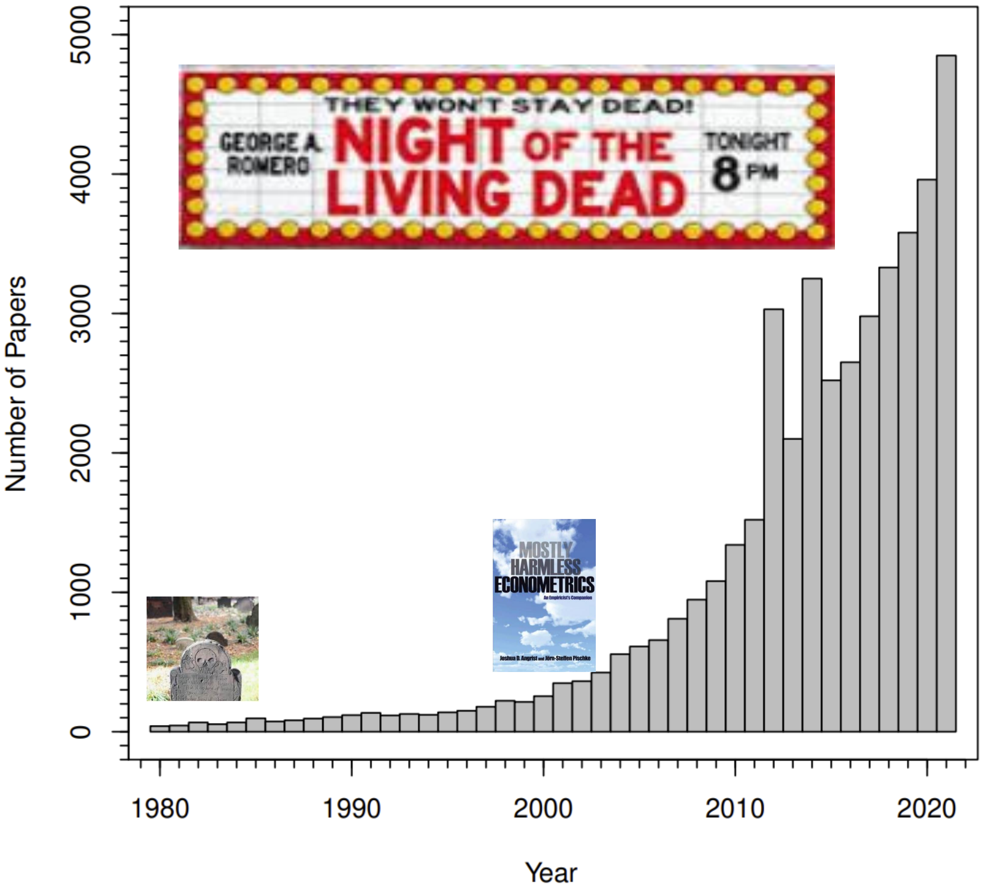
- **Post 1970s -** Microeconomic data and **discrete outcomes** mandate intrinsically nonlinear models.
  - Theil (1971): **binary responses by consumers**
  - McFadden (1972): **multinomial choice of travel mode**
  - Zavoina and McElvey (1975): **ordered choices and preference (rating) scales**

Technological Progress for Binary Choice Models In the 1980s

- Nonlinear binary choice models accepted.
- Probit/Logit MLE standard approach.
- Model builders reconcile probit and logit
  - Coefficients differ
  - Average partial effects don't differ
  - We care about partial effects.
- Experimental research on choice models; including how to save LPM.
- Linear regression generally abandoned.

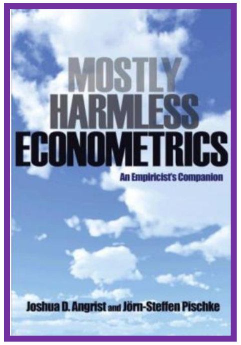
RIP LPM by 1980

The Linear Probability Model Returns with the Credibility Revolution, Undead



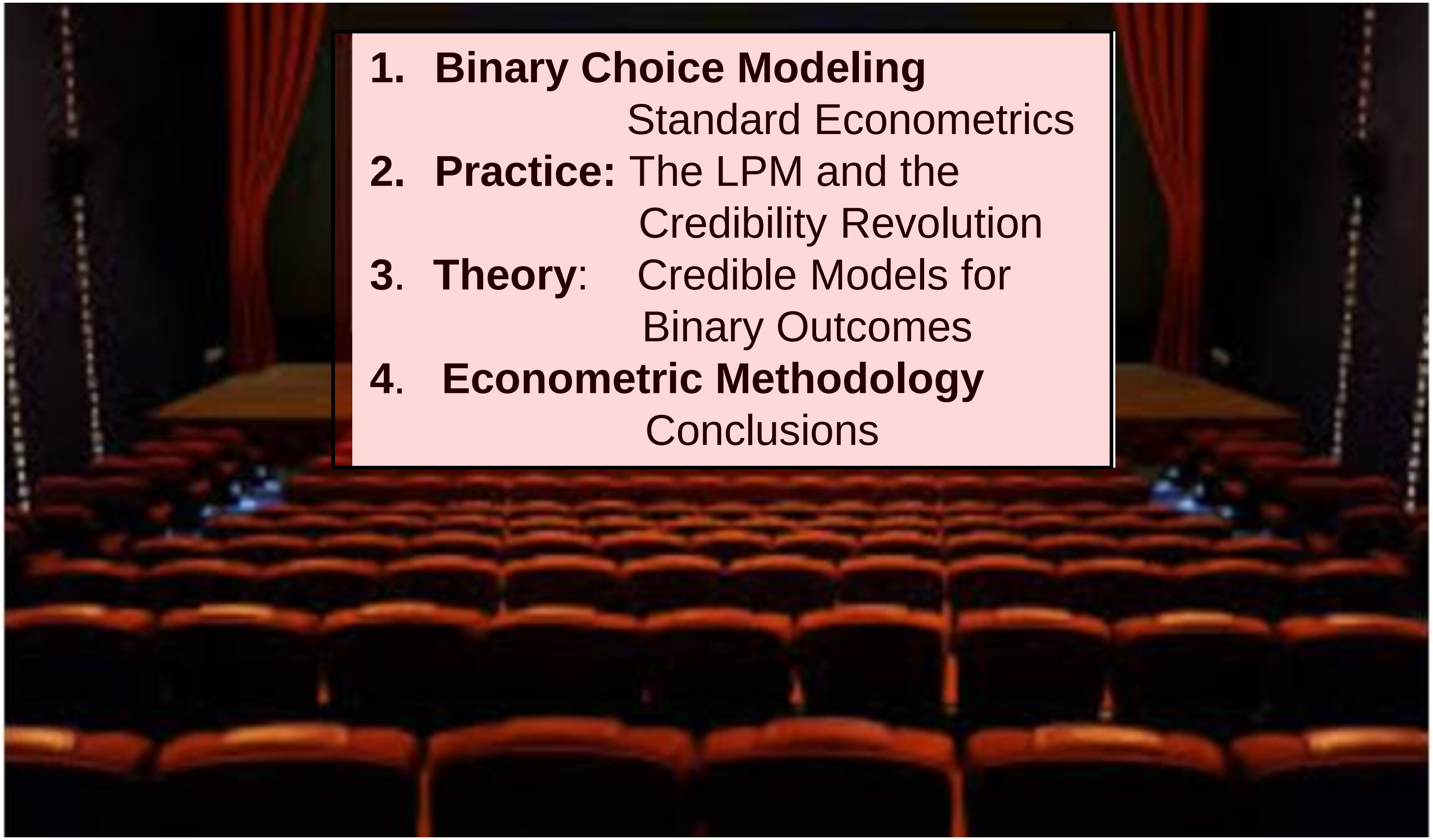
1980-2000 Organic fringe of the growing empirical literature.

2000- The living dead awakened by the credibility revolution.


MOSTLY HARMLESS ECONOMETRICS
An Empiricist's Companion
Joshua D. Angrist and Jörn-Steffen Pischke

Mostly Harmless Econometrics (Angrist and Pischke, 2005) **is *the* widely cited source for supporting use of the newly credible linear probability model.**

Zombie Econometrics:
The Linear Probability Model



1. **Binary Choice Modeling**
   Standard Econometrics
2. **Practice:** The LPM and the
   Credibility Revolution
3. **Theory**: Credible Models for
   Binary Outcomes
4. **Econometric Methodology**
   Conclusions

Critics rave…

It's not just nit picking about functional form! … William Greene

Origin: The Probit *Method*
Chester Bliss, 1930s, Bio-Assay*

- **Theory**: What proportion of 1,000 Aphids will ‘respond’ to insecticide dose **X** by dying?
  - o Prob[y_i ($\mathbf{X}_i$) = 1] = Prob[(DIE_i = YES)|$\mathbf{X}_i$]
  - o Depends on dosage, **X**, and random bug specific resistance factor, ε_i
  - o If stimulus, $\mathbf{X}_i$, exceeds resistance, ε_i, response is ‘DIE_i =YES’.

- **Data**: Proportion of Aphids that respond to ‘stimulus,’ insecticide dose **X**, by dying. y_i = **1**[(DIE_i = YES)|$\mathbf{X}_i$].

- **Model:** Prob[DIE_i = YES)|$\mathbf{X}_i$] = Prob(y_i = 1|$\mathbf{X}_i$) = F($\mathbf{X}_i$)

***Bliss, C. (1934) “The Method of Probits.” *Science* 79, pp. 38-39.**

Sample Evidence

Bliss described the nonlinear relationship between dosage and response as *sigmoidal*.

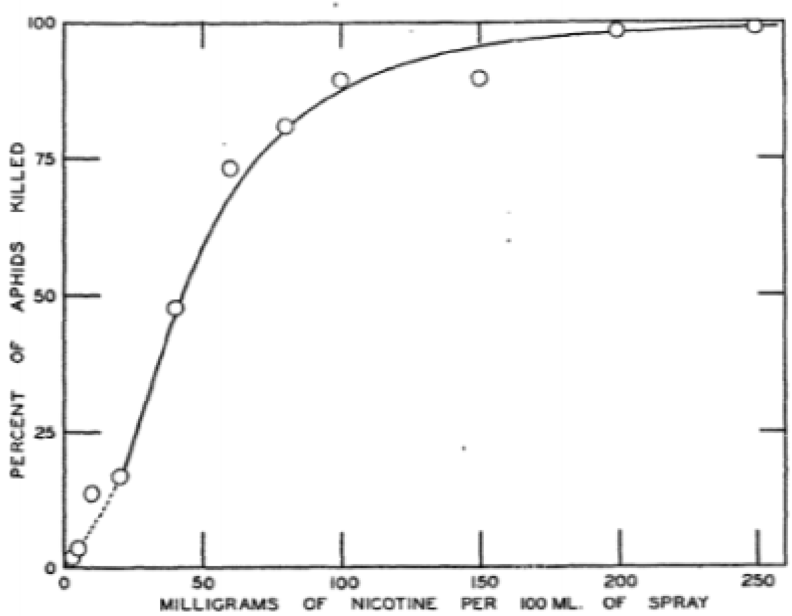


FIG. 1. Net mortality of *Aphis rumicis* L. sprayed in laboratory with different solutions of nicotine; summary of results over 3-year period. Tattersfield and Gimingham.[4] Heavy curve is same as that in Fig. 2 transposed back to original units.

Application: The Normit ***Method***

What dosage is required to achieve 50% death rate?
What is "Lethal Dose 50" = LD50 (or LD90 or LD95)?

Probability = .95

Probability = .90

Probability = .50

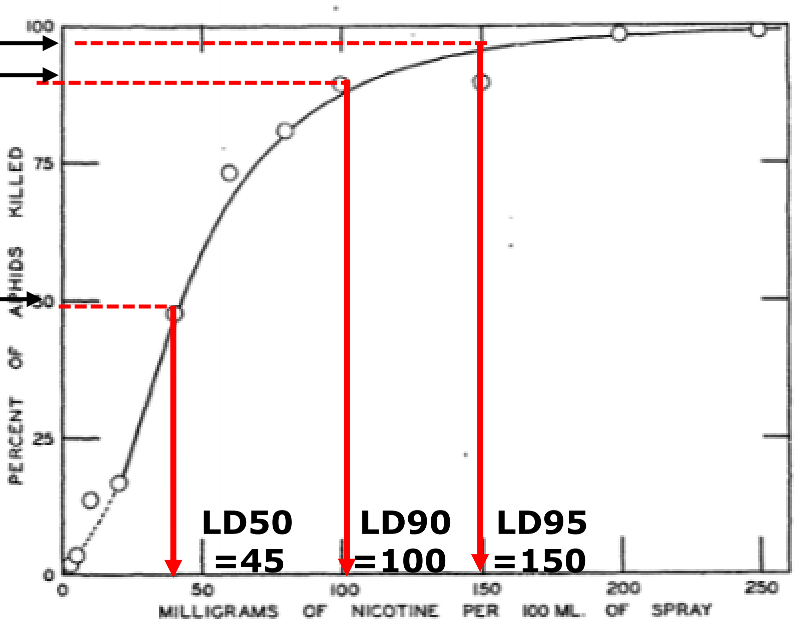


FIG. 1. Net mortality of *Aphis rumicis* L. sprayed in laboratory with different solutions of nicotine; summary of results over 3-year period. Tattersfield and Gimingham.[4] Heavy curve is same as that in Fig. 2 transposed back to original units.

Normit(P)
= LDP
= X
= F^{-1}(P)

LD50 = F^{-1}(.5)

for some function F(X).

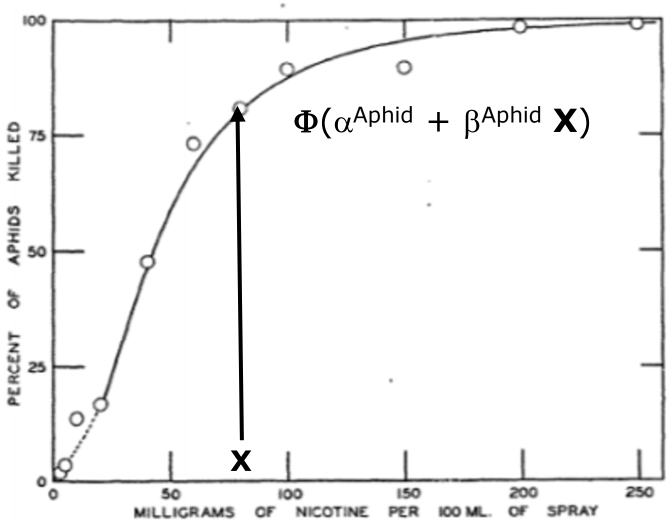
The Probit *Model**

How does the percentage of insects killed respond to changes in the dosage, X?

Probability = Prob($\varepsilon < \alpha^{Aphid} + \beta^{Aphid}\mathbf{X}$) = $\Phi(\alpha^{Aphid} + \beta^{Aphid}\mathbf{X})$

α and β are specific to the particular pest (aphid, fruit fly, etc.)

Probabilities are computed using the normal distribution.



<u>Probability Unit</u>
Probit equals Normit + 5 to avoid negative numbers
Note that α and β depend on the normal distribution.

***Finney, D., (1947) "Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve," Cambridge University Press; and Bliss (1934).**

How to Compute the Parameters

How to find α^{Aphid} and β^{Aphid} ?

Solve two equations:

$\Phi(\alpha + \beta \text{LD50}) = 0.5$, $\text{LD50} = 45$

$\Phi(0.00) = 0.5$ so $\alpha + 45\beta = 0.00$

$\Phi(\alpha + \beta \text{LD90}) = 0.9$, $\text{LD90} = 100$

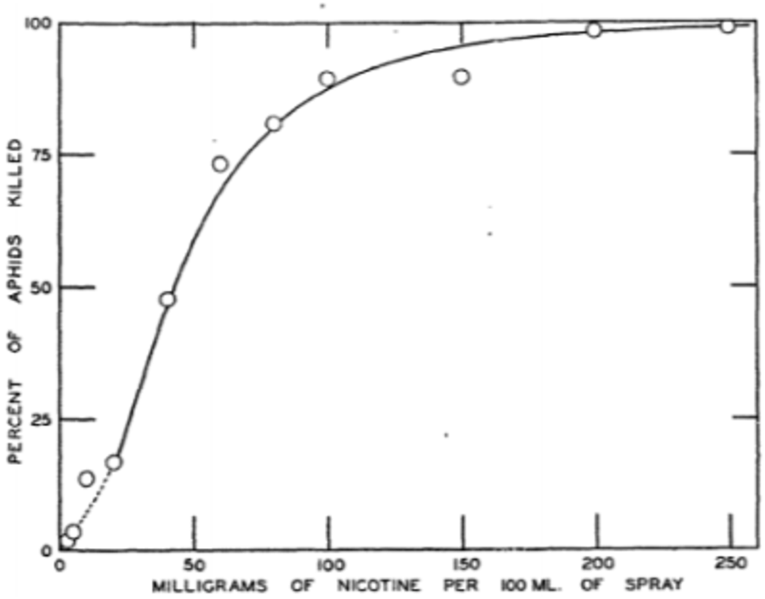
$\Phi(1.28) = 0.9$ so $\alpha + 100\beta = 1.28$

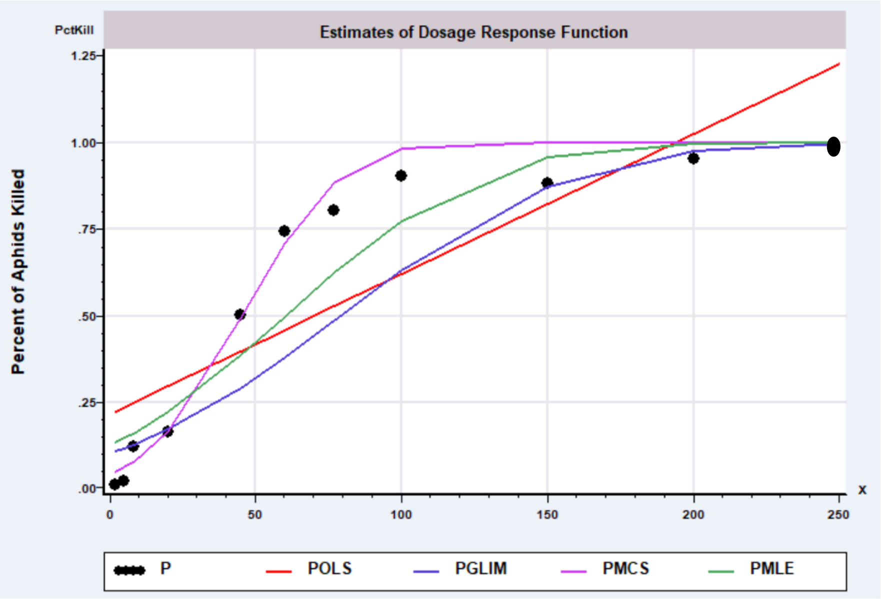
$$\alpha = -1.049, \beta = 0.0233$$

Use α and β to find Pct|**X**. If **X**=105,

$\text{Pct} = \Phi(-1.049 + .0233 * 105) = .92$

Modern Estimators Use All Data





$$\begin{bmatrix} P_i \\ \mathbf{x}_i \end{bmatrix}_{\mathrm{i=1}}^{11} = \begin{bmatrix} .01 & .02 & .12 & .16 & .50 & .74 & .80 & .90 & .88 & .95 & .99 \\ 2 & 5 & 8 & 20 & 45 & 60 & 77 & 100 & 150 & 200 & 250 \end{bmatrix}.$$

Three possible common estimators of (α,β) are:

Generalized Linear Model; $\mathrm{Min}(\alpha,\beta)\ \sum_{\mathrm{i=1}}^{11}\left[\Phi^{-1}(\mathrm{P_i})-(\alpha+\beta\mathbf{x}_i)\right]^2 = (-1.27,+0.016).$

Minimum Chi Squared; $\mathrm{Min}(\alpha,\beta)\ \sum_{\mathrm{i=1}}^{11}\dfrac{\left[\mathrm{P_i}-\Phi(\alpha+\beta\mathbf{x}_i)\right]^2}{\Phi(\alpha+\beta\mathbf{x}_i)} = (-1.733,+0.038).$

Maximum Likelihood; $\mathrm{Max}(\alpha,\beta)\ \sum_{\mathrm{i=1}}^{11}\left[\mathrm{P_i}\ln\Phi(\alpha+\beta\mathbf{x}_i)+(1-\mathrm{P_i})\ln[1-\Phi(\alpha+\beta\mathbf{x}_i)]\right] = (-1.150,0.019).$

Bliss and Finney’s Probit Method

- Nonlinearity of $F(\mathbf{X})$ is crucial to the specification
- Identical to modern, ‘threshold’ binary response model.
  - Structural model is *random utility* (resistance).
  - Outcome is the *binary “choice,”* not a measurement.
  - Specification is *probit* based on the normal distribution.

Binary Choice Models Evolve By The 1980s

- Probits and Logits
  - The latent regression approach with normality; Bliss and Finney, 1934/1947, Theil (1971), many others
  - Theil (1971) likes normality and the probit model.
  - Berkson (*Biometrics*, 7(4), 1951) prefers logits to probits. Easier to compute; he likes odds ratios.
  - Others, e.g., Chen and Tsurumi (2010) are not sure.
- A longer menu of choices – not well motivated. Probit, Logit, Burr, Comploglog, Gompertz, ArcTangent,… Raises the question. Does functional form matter? Functional form is not testable. Can we find the right one?
- Probit and logit survive. Preferences split.
- Linear regression is no longer used for binary choice.

Old School Practicalities Motivate LPM

Horrace and Oaxaca (2006) report practical motivations for using OLS instead of Probit:

→ McGarry (2000): ease of interpretation of estimated marginal effects.
♦ All modern software report partial effects. Slopes are not elasticities. See OutTakes.

→ Reiley (2005): perfect correlation problem with the probit model.

Complete separation: y = 0 1 0 1 0 0 0 0
x = 0 0 0 0 1 1 1 1

Linear (a,b) = .5, -.5 Probit (a,b) 0.0 -7.0
(s.e.) (.2) (.3) (s.e.) (.6) (169,000)

♦ This is not a problem with the probit model. It is a problem with the data/specification.

Greene's Aphorism #1: You can regress anything on anything.
The results aren't always meaningful.

→ Bettis and Fairlie (2001): an extremely large sample size and other simplifications.
♦ Probit requires the same computer resources and trivially more time as OLS.

→ Currie and Gruber (1996): logit, probit and OLS are similar.
♦ Not implied by any theory. Sometimes not true.

Intellectual Appeal of Linear Regression

Concern with nonlinearity. **Maybe functional form matters**. Least squares is motivated by a linear model. Reliable approximation?

$y = \mathbf{x}'\beta + \varepsilon$. $y \in \{0,1\}$

$\text{Prob}(y = 1|\mathbf{x}) = E[y|\mathbf{x}] = \mathbf{x}'\beta$

The linear regression model clings to life in the laboratory.

- Amemiya (1985), Maddala (1983) derive "properties" of OLS and WLS
- Horrace and Oaxaca (2006) looked for an unbiased estimator.
- Wooldridge (2010) treats it as an ordinary regression with inconvenient flaws.

Until ca. 2000, linear binary choice model is viewed with skepticism

- Need for nonlinearity outweighs discomfort with specific choice of functional form.
- It turns out the nonlinear functional form doesn't matter very much.

Ca. 2000 the now "Linear Probability Model" gains acceptance

- Angrist and Pischke (2005) promote it as credible and harmless
- "Properties" are unimportant. Credible experimental design matters more.

Conventional Econometrics: The Linear Probability Model's Well Known, Obvious Shortcomings

A. Probabilities reside outside [0,1]. (Uncomfortable)

- **Use restricted least squares? Impossible**
- Requires theoretical restrictions on β that depend on the data. $0 < \beta'\mathbf{x} < 1$ for all admissable β and all supported $\mathbf{x}$. *Impossible*.
- It doesn't matter; partial effects matter. OLS always estimates partial effects.

B. Heteroskedasticity (Irrelevant side issue)

- Inherently heteroskedastic; $\text{Var}(y|x) = [\beta'\mathbf{x}(1 - \beta'\mathbf{x})]$
- Inefficient**.** Weighted LS? Wt = $[\beta'\mathbf{x}(1 - \beta'\mathbf{x})]^{-1/2}$ *Might be < 0*!
- White estimator? Clustering? *Robust to what? There is no model!*

Conventional Econometrics: The Model

Cameron and Trivedi (2005)

> [T]he LPM usually provides a reasonable *approximation*.
> (**To something**) They do not argue that LPM is a good model.
> Discomfort with peculiar probability function that allows $\mathbf{x}'\beta \notin [0,1]$

Wooldridge (2010) assumes only $E[y|\mathbf{x}] = \mathbf{x}'\beta$.

> "If the main purpose of estimating a binary response model is to approximate the partial effects of the explanatory variables, averaged across the distribution of x, then the LPM often does a very good job… The fact that some predicted probabilities are outside the unit interval need not be a serious concern. *But, there is no guarantee that the LPM provides good estimates of the partial effects…*" (p. 563)

- **How do you know it does a very good job?**
- **Why are nonsense probabilities "not a serious concern?"**
- **No theory for the estimation or inference tools.**
- **Discomfort with LPM as a model. Seems to suggest OLS is a statistic.**

Conventional Econometrics: The Function

Can we establish consistency of OLS within a valid model?

Start from $E[y|x] = F(\mathbf{x}'\beta)$. It follows $y = F(\mathbf{x}'\beta) + \varepsilon$ with $E[\varepsilon|F(\mathbf{x}'\beta)] = 0$ and

Prob($y=1|\mathbf{x}$) = $F(\mathbf{x}'\beta)$. Assume $F(\mathbf{x}'\beta) = \mathbf{x}'\beta$. Implies $E[\varepsilon|\mathbf{x}] = 0$

Amemiya (1985), pp 268-269.

> *The LPM has an obvious defect in that F for this model is not a proper distribution function as it is not constrained to lie between 0 and 1. This defect can be corrected by defining F = 1 if F($\mathbf{x}'\beta$) > 1 and F = 0 if F($\mathbf{x}'\beta$) < 0, but the procedure produces unrealistic kinks at the truncation points.*

Not a solution. Erroneously retains $y = F(\mathbf{x}'\beta) + \varepsilon$ Now ε must be restricted.
Function is nonlinear. Unclear what the conditional mean is. NOT $F(\mathbf{x}'\beta)$
Relies on *Uniform* distribution

Conventional Econometrics: Horrace/Oaxaca

Horrace and Oaxaca (2006): Conclusions:

As long as the sample *always* has **x'β** in (0,1) then OLS is unbiased and consistent. OLS is inconsistent when **x'β** not in (0,1). **x'β** = 0 is problematic. First observation outside, OLS becomes biased.
This mixes sample data and theoretical dgp.
Requires knowledge of population values. Invalid argument.

Solution?

As long as y always = **x'β** + ε, asymptotic properties are conventional.

Necessary and sufficient to observe which observations in a random sample are in the set with 0 < **x'β** < 1**.** Logical inconsistency is if DGP ***can*** produce "bad" observations.

Actual conclusion: Model works if it works. Need theory restriction on DGP: 0 < **x'β** < 1.

Incoherence: Spanos (1986)

Simulating the data generating process

[1]. Choose "true" parameters, β.

[2a]. Choose or simulate generation of the exogenous data, X.

[2b]. Simulate the generation of the random outcomes.

Probit DGP

Simulate normally distributed ε independent of X.

Compute latent $y^* = X\beta + \varepsilon$.

Compute $y = 1(y^* > 0)$

Linear Probability DGP

Enforce binary $y = X\beta + \varepsilon$

Choose $\varepsilon = -X\beta$ if $y = 0$,

$\varepsilon = 1 - X\beta$ if $y = 1$.

Compute $y = X\beta + \varepsilon$

It is not possible to simulate data from the LPM DGP.

The linear probability model is incoherent even if $0 < \mathbf{x}'\beta < 1$ for all theoretical values of $\mathbf{x}'\beta$. The problem is equating discrete outcome, y, to continuous probability of y.

$y = \text{Prob}(y=1|\mathbf{x}) + \varepsilon$ makes no sense. Try $\varepsilon = 0$.

Summary of the Econometric Issue

- The Linear Probability *Model* is incoherent. (There is no “reduced form.”)

- What’s wrong with that?
    - There is no theory to explain calculations based on the “model”
    - There is no theory that justifies OLS. Only casual observation that OLS often resembles partial effects from probit models.

- How do we know that OLS provides a good approximation? Usually resembles probit.

- **What should the researcher do when OLS does not resemble the probit model?**

An Ordinary Example: “Modeling Innovation*”

***Bertschuk, I. and M. Lechner. (1998). Convenient Estimators for the Panel Probit Model,” *Journal of Econometrics*, 87(2).**

Did firm *i* produce an innovation in year *t* ? y_{it} : 1=Yes / 0=No

Observed *N*=1270 German firms for *T*=5 years, 1984-1988

Observed covariates: $\mathbf{x}_{it}$ = Industry, competitive pressures, size, productivity, etc.

Important input variables of interest.

SP = ratio of industry imports to sales + industry imports

FDIUM = ratio of industry foreign direct investment to sales + imports

Probit and Logit coefficients appear uncomfortably different. Logit coefficients are about 60% larger.

Average partial effects look similar.
It's partial effects that matter anyway, not coefficients.
The choice between probit and logit is inconsequential.

	Probit		Logit		Linear
	Coefficient	APE	Coefficient	APE	Coefficient
Constant	-1.96031	-	-3.22324	-	-0.10424
Log Sales	0.17711	0.06573	0.29639	0.06766	0.05198
SP	1.07274	0.39812	1.92656	0.43993	0.09492
IMUM	1.13384	0.42080	1.79889	0.41101	0.45284
FDIUM	2.85318	1.05890	4.76252	1.08753	1.07787
PROD	-2.34116	-0.86887	-4.42565	-1.01060	-0.55012
RawMtl	-0.27858	-.010569	-0.41120	-0.09635	-0.09861
InvGood	0.18796	0.07045	0.29327	0.06758	0.07879

Which to Report, Probit or LPM?

Partial Effects from the German Innovation Model

	Partial Effects	
	Probit	Linear
LogSales	0.0657	0.0520
SP	0.3981	0.0949
PROD	-0.8689	-0.5501

Econometric Conclusions

We need to distinguish between the *statistic*, **b** = OLS(**X**,**y**) and the *estimator* of the slopes in a linear probability *model*.

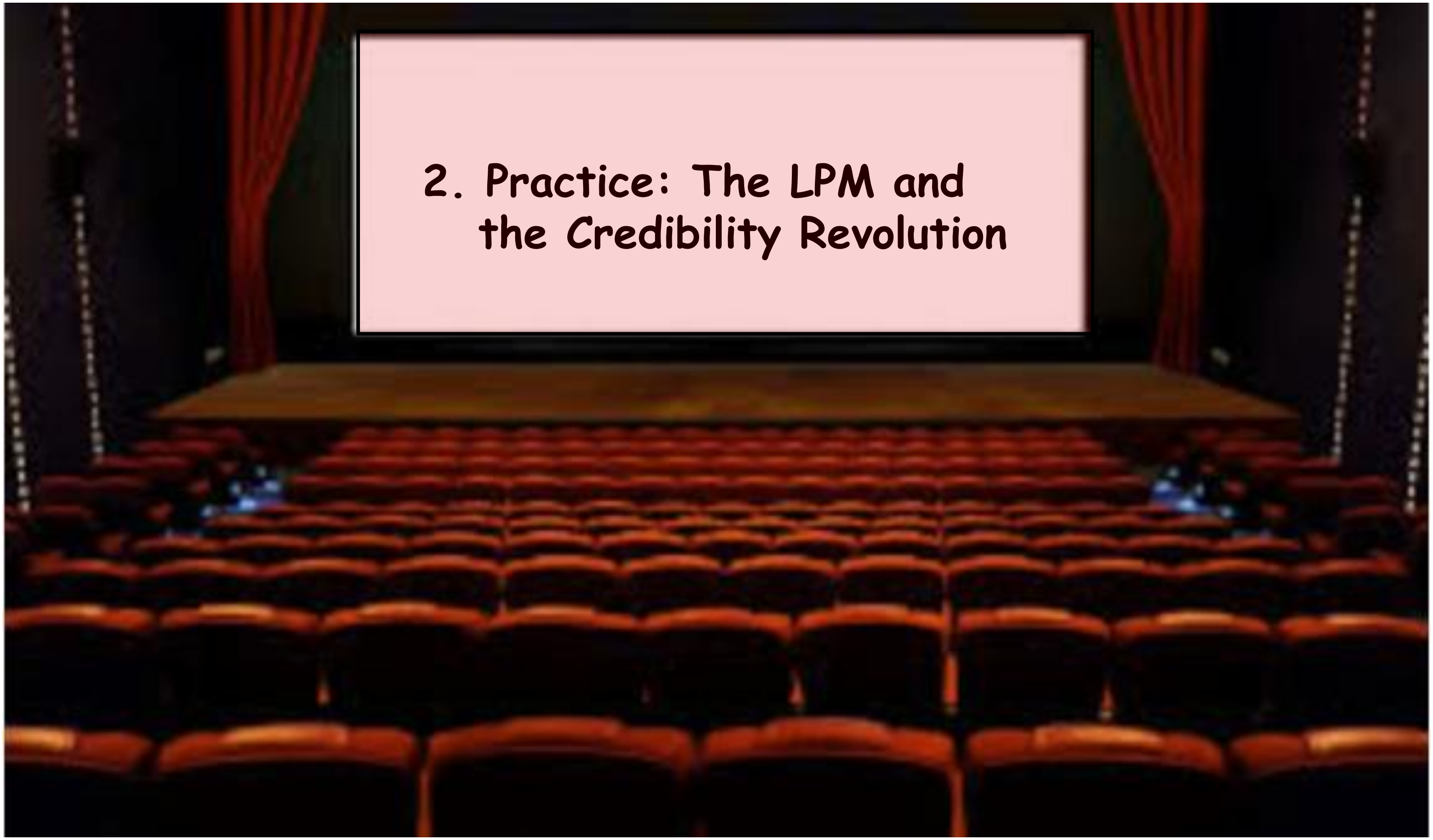
The properties of the statistic, **b**, have not been shown.
The LPM is incoherent, so OLS is not estimating the slopes of a conditional mean.
It is unclear what OLS is estimating. It seems to resemble an average partial effect.

Angrist and Pischke (2005, p. xii)

"… the estimators in common use have a simple interpretation that is not heavily model dependent." (Using OLS and 2SLS for binary outcomes.)

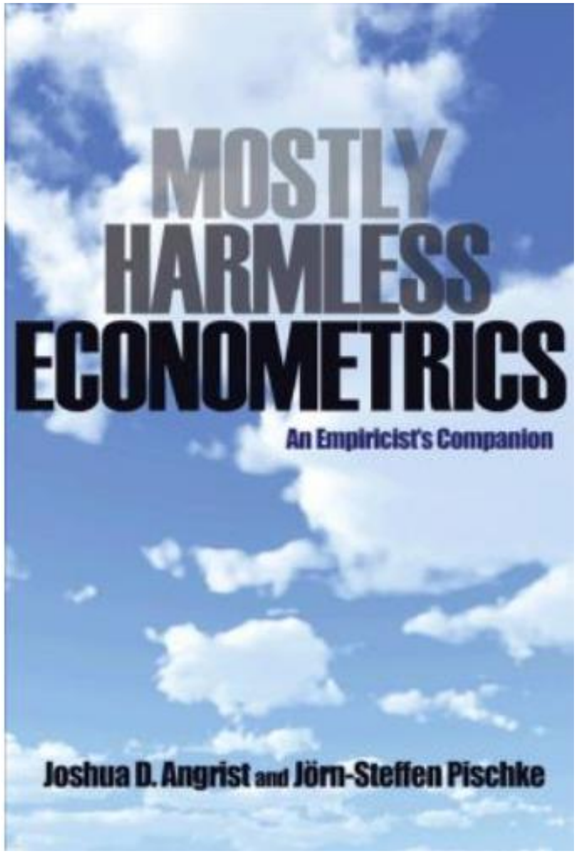
" … linear regression … provides useful information about the conditional mean function regardless of the shape of this function."

Ultimately amounts to saying the thing we are interested in is what is estimated by OLS (or 2SLS).



2. Practice: The LPM and the Credibility Revolution

The Credibility Revolution Arrives at the Fin de Siècle

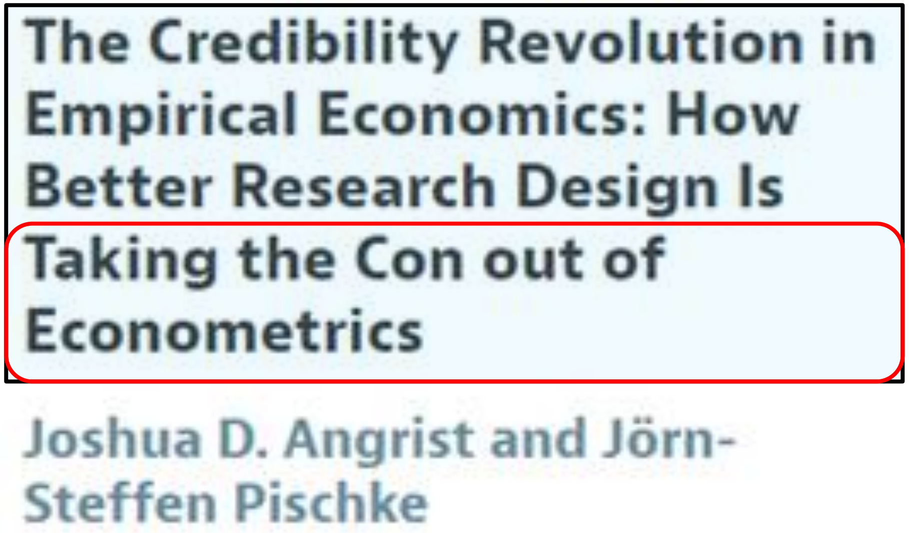


MOSTLY HARMLESS ECONOMETRICS

An Empiricist's Companion

Joshua D. Angrist and Jörn-Steffen Pischke

The Manifesto, ca. 2005



The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics

Joshua D. Angrist and Jörn-Steffen Pischke

The American Economic Review

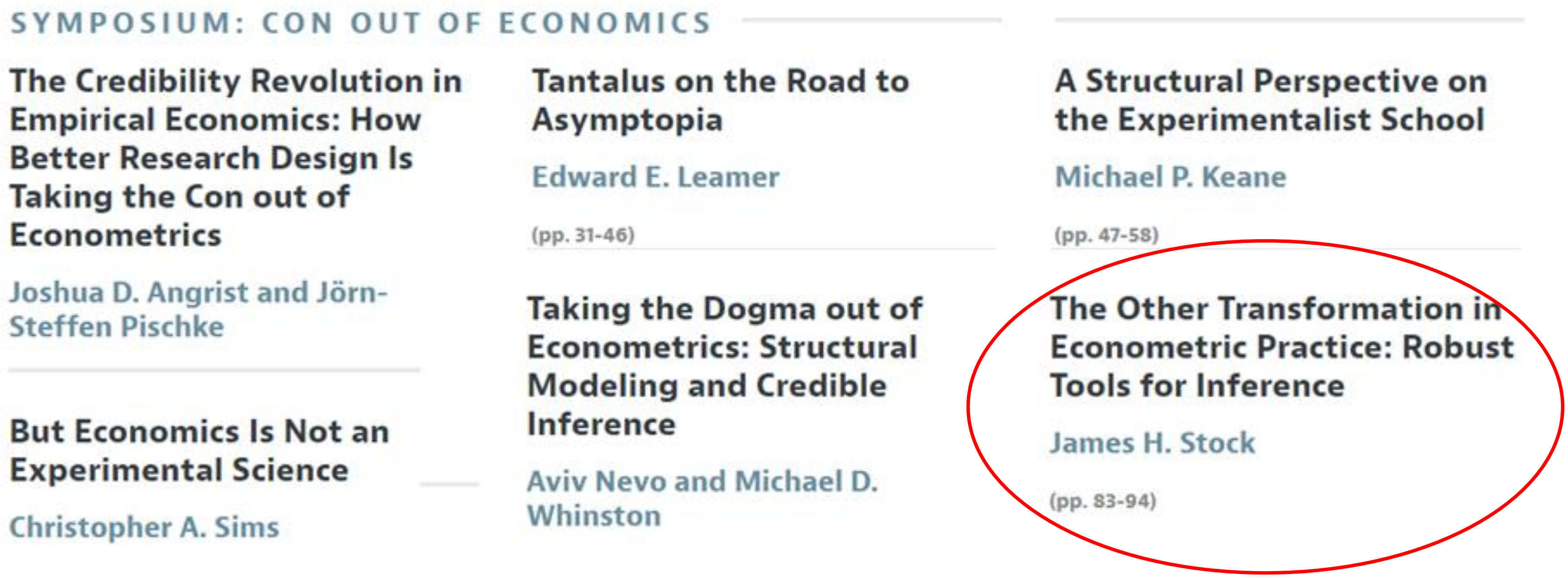
1983 Article
Let's Take the Con Out of Econometrics

Edward Leamer
AER, 73,1, May 1983.

Journal of Economic Perspectives

Vol. 24 No. 2 Spring 2010

The profession was not amused.



SYMPOSIUM: CON OUT OF ECONOMICS

The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics

Joshua D. Angrist and Jörn-Steffen Pischke

But Economics Is Not an Experimental Science

Christopher A. Sims

Tantalus on the Road to Asymptopia

Edward E. Leamer

(pp. 31-46)

Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference

Aviv Nevo and Michael D. Whinston

A Structural Perspective on the Experimentalist School

Michael P. Keane

(pp. 47-58)

The Other Transformation in Econometric Practice: Robust Tools for Inference

James H. Stock

(pp. 83-94)

A Sea Change in Econometrics

The Other Transformation in Econometric Practice: Robust Tools for Inference

James H. Stock

(pp. 83-94)

Pre revolution: consistency is essential; inconsistency is unacceptable
efficiency is paramount; robustness is a side issue.

Post revolution: consistency is useful; credible approximation is sufficient;
efficiency is inessential; robustness is paramount.

Angrist and Pischke (2005), page xiii.

> "… *we are not much concerned with asymptotic efficiency. Rather, our discussion is devoted mostly to the finite-sample bugaboos that should bother practitioners.*"

On Models

Angrist and Pischke (2005), page xii.

> *"Most econometrics texts appear to take econometric models very seriously…. We take a more forgiving and less literal-minded approach… The estimators in common use almost always have a simple interpretation that is not heavily model dependent."*

This entire line of reasonable appears to be applied to the univariate and bivariate probit models. The literature at large appears still to be comfortable with intricate settings such as ordered probit and multinomial choice modeling, latent class models for count data, survival modeling and so on.

The econometrics are carried out without reliance on specific models when possible.

On Politics

Why do credible revolutionaries *prefer* the discredited linear probability ‘model’ to the coherent probit model?

Ultimately, I think the preference for one or the other is largely generational, with people who went to graduate school prior to the Credibility Revolution preferring the probit or logit [model] …

Marc F Bellemare: A Rant on Estimation with Binary Dependent Variables (Technical)
http://marcfbellemare.com/wordpress/8951
(visited 08/09/22)

On Partial Effects

> *[W]hile a nonlinear model may fit the CEF for LDVs more closely than a linear model, when it comes to marginal effects, this probably matters little. This optimistic conclusion is **not a theorem**, but, as in the empirical example, **it seems to be fairly robustly true**.* (*Angrist and Pischke (2005), p. 107*)

The LPM is not a model, so there are no theorems. Just hope for the best. The econometric platform *seems* to work.

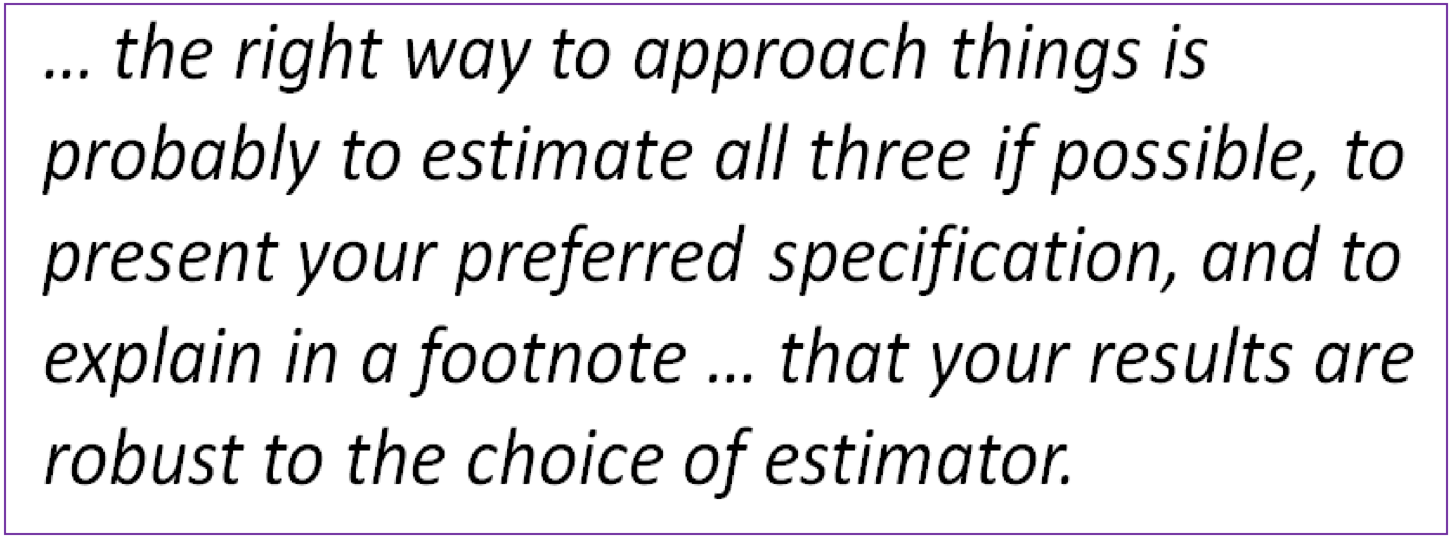
Two questions:

(1) How do you know it is "fairly robustly true?"
Because it returns results similar to probit.

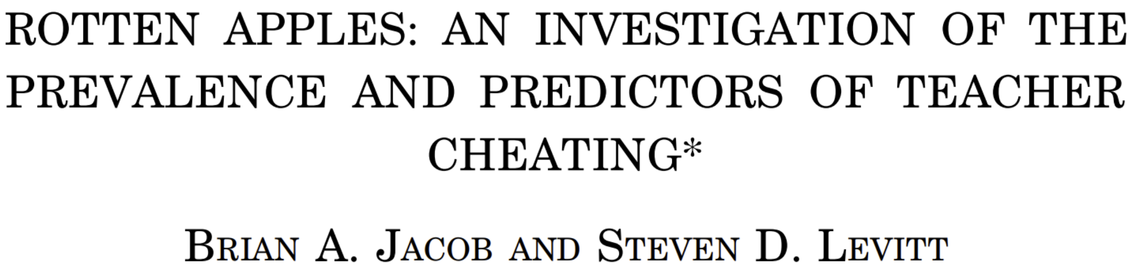
(2) What should you do if it is not true?
What to do if results are not similar to probit?

On Robustness

Marc F Bellemare: A Rant on Estimation with Binary Dependent Variables (Technical)
http://marcfbellemare.com/wordpress/8951



… the right way to approach things is probably to estimate all three if possible, to present your preferred specification, and to explain in a footnote … that your results are robust to the choice of estimator.



ROTTEN APPLES: AN INVESTIGATION OF THE PREVALENCE AND PREDICTORS OF TEACHER CHEATING*

BRIAN A. JACOB AND STEVEN D. LEVITT

Quarterly Journal of Economics, 2003, 118(3,Aug.), 843-878

*We report estimates from linear probability models (**probits** yield similar marginal effects), with standard errors clustered at the school level.* Page 863, online.

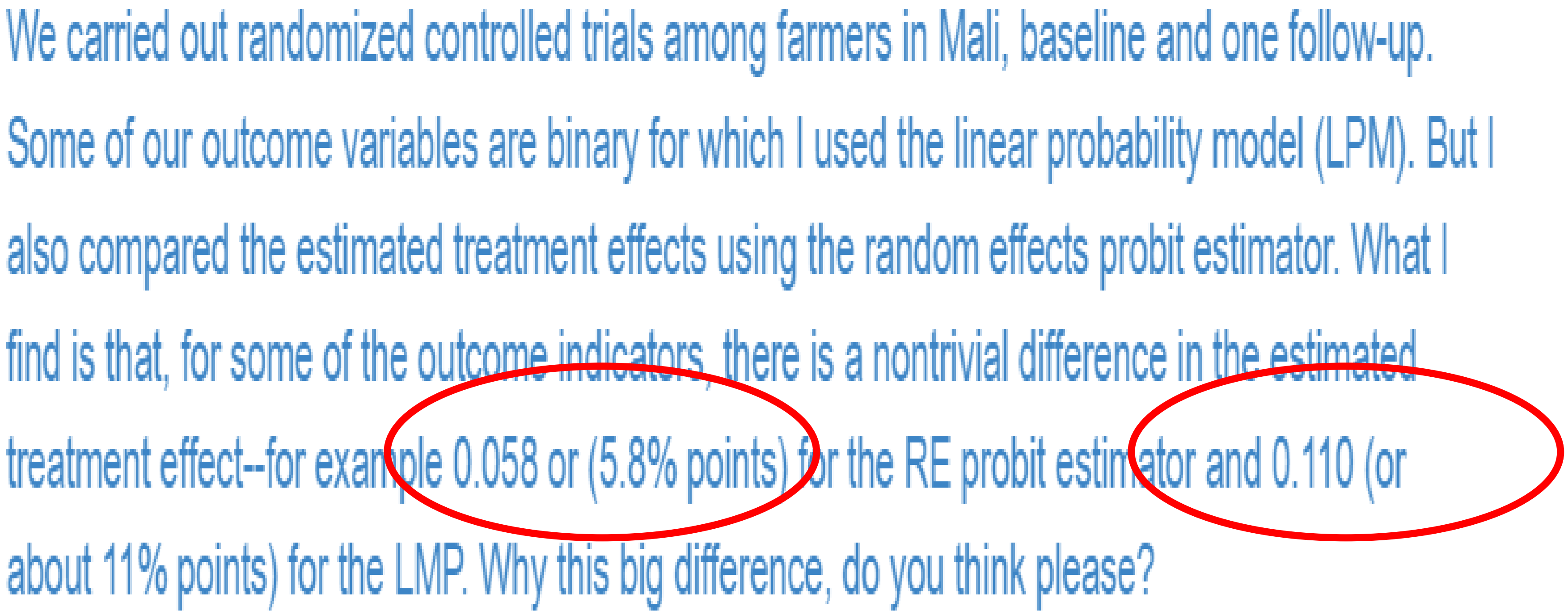
NBER Working Paper 9413, 2002. p. 37 of 69, fn 31.

***Logit** models evaluated at the mean yield comparable results, so the estimates from a linear probability model are presented for ease of interpretation.*

(The working paper also uses the probit model on p. 32 and no counterpoint for 2SLS on p. 41)

An Econometrics 911
It’s “robustly true.” Except when it isn’t.

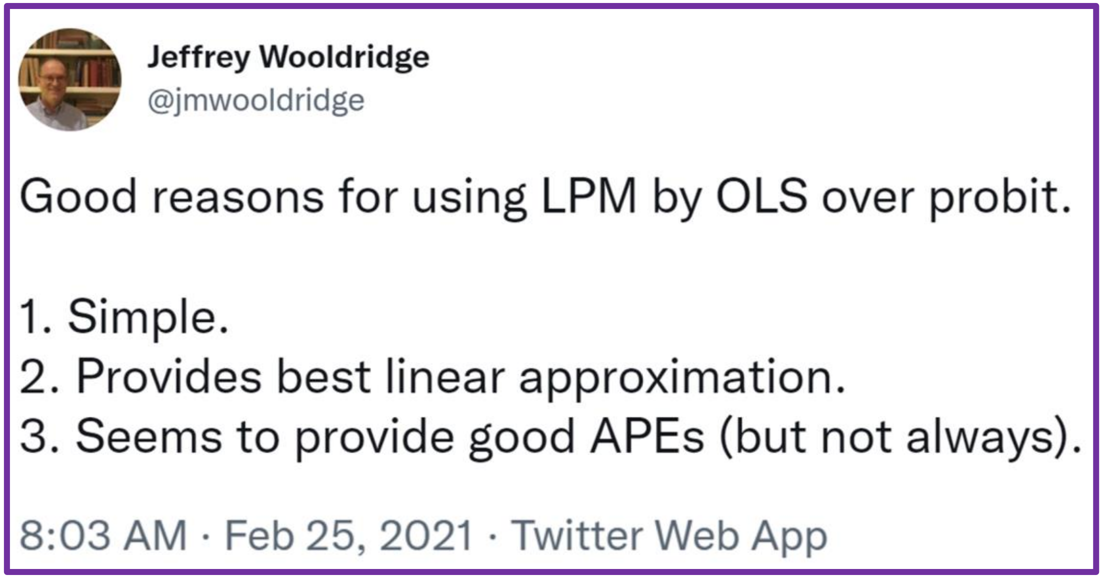
Dear Professor Greene:



We carried out randomized controlled trials among farmers in Mali, baseline and one follow-up. Some of our outcome variables are binary for which I used the linear probability model (LPM). But I also compared the estimated treatment effects using the random effects probit estimator. What I find is that, for some of the outcome indicators, there is a nontrivial difference in the estimated treatment effect--for example 0.058 or (5.8% points) for the RE probit estimator and 0.110 (or about 11% points) for the LMP. Why this big difference, do you think please?

Regards, Fred

Compelling Virtues of the LPM

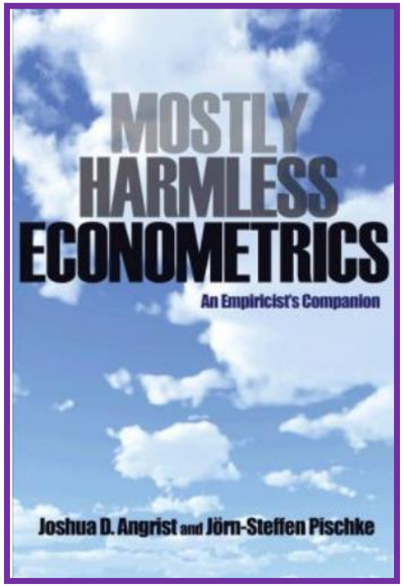


Jeffrey Wooldridge
@jmwooldridge

Good reasons for using LPM by OLS over probit.

1. Simple.
2. Provides best linear approximation.
3. Seems to provide good APEs (but not always).

8:03 AM · Feb 25, 2021 · Twitter Web App


MOSTLY HARMLESS ECONOMETRICS
An Empiricist's Companion
Joshua D. Angrist and Jörn-Steffen Pischke

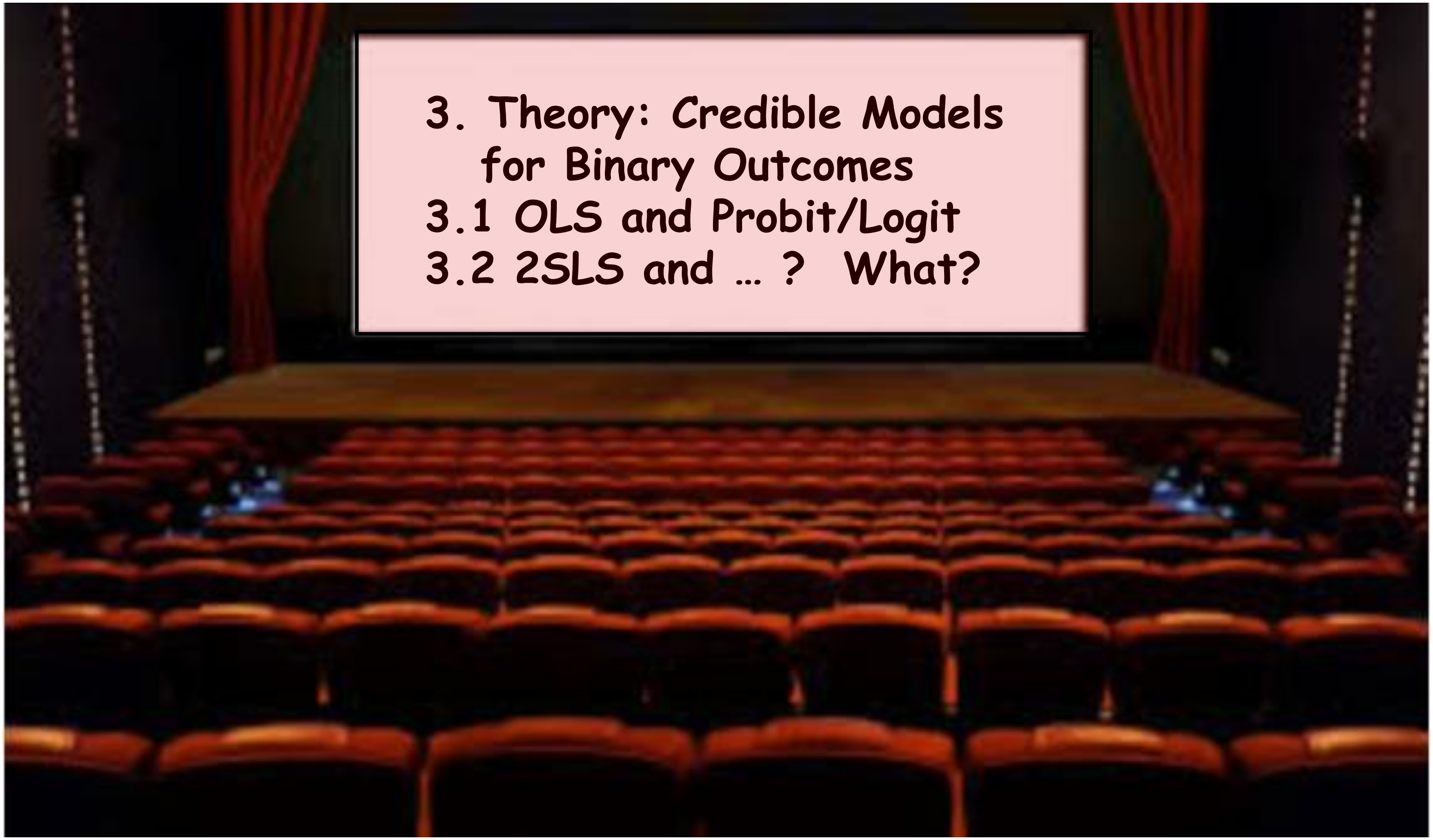
“It seems to be fairly robustly true.” (p. 107)

Skepticism

- Bypassing the complexity of nonlinear models is a small objective. The world is full of great programmers and great software.

- ‘Robustness’ here means OLS looks like the APEs from the probit model. Apparently not always.

- Not a theorem implies that OLS “jumps” to partial effects. It only “seems” to happen. We don’t know why. Sometimes it doesn’t happen



3. Theory: Credible Models for Binary Outcomes

3.1 OLS and Probit/Logit

3.2 2SLS and … ? What?

A Coherent Linear Probability/Regression Model

A&P's Discussion begins with the "saturated" model, no covariates.

Claim: The LPM is the *right* model when it is saturated.

Random assignment of treatment, T_i**. Prob(**T_i=1**|All Info) =** δ.

(a) **(LP)** Prob(Y_i=1|T_i) = $\alpha + \beta T_i$ Prob(Y_i=0|T_i) = 1 – Prob(Y_i=1|T_i)

(LR) $Y_i = \alpha + \beta T_i + \boldsymbol{\varepsilon_i}$

(b) $E[Y_i \mid T_i] = \alpha + \beta T_i$

(c) $E[\varepsilon_i \mid T_i] = 0$

Proposition:
In the "saturated" (exactly identified) case, **(a), (b), (c)** are not actually part of a model. LPM is just algebra, not a model.

The Sample Data in the Saturated Case

(T_i,Y_i), i = 1,N Random sampling and random assignment assumed.

N00 = #(T=0,Y=0); **P00 = N00/N** = EstProb(T=0,Y=0)
N10 = #(T=1,Y=0); **P10 = N10/N** = EstProb(T=1,Y=0)
N01 = #(T=0,Y=1); **P01 = N01/N** = EstProb(T=0,Y=1)
N11 = #(T=1,Y=1); **P11 = N11/N** = EstProb(T=1,Y=1)
N1• = #(T=1) = N10 + N11; **P1• = N1•/N** = EstProb(T=1); **P0• = 1 – P1•** = EstProb(T=0)
N•1 = #(Y=1) = N01 + N11 ; **P•1 = N •1/N** = EstProb(Y=1); **P•0 = 1 – P•1** = EstProb(Y=0)

All of the sample information is contained in the crosstab

.

		Y		
		0	1	
T	0	P00	P01	P0•
	1	P10	P11	P1•
		P•0	P•1	1

Estimation Based on the Law of Large Numbers (Moments)

Estimate the treatment effect, TE

TE = Prob(Y=1|T=1) - Prob(Y=1|T=0)

Use the definition of conditional probability

$$\boxed{TE = \frac{P11}{P1\bullet} - \frac{P01}{P0\bullet}}$$

LPM Using Least Squares

Estimate TE = β using OLS

$$b = \frac{\sum_{i=1}^{N}(T_i - \bar{T})Y_i}{\sum_{i=1}^{N}(T_i - \bar{T})T_i} = \boxed{\frac{P11 - P1\bullet \times P\bullet 1}{P1\bullet(1 - P1\bullet)}}$$

$$b = TE = \frac{P11}{P1\bullet} - \frac{P01}{P0\bullet}$$

Identical to method of moments estimator

(Hints: 1-P1• = P0• and P•1 = P01+P11)

Direct Maximum Likelihood

$$
\begin{aligned}
\ln L(\alpha,\beta,\delta) &= \sum_{i=1}^{N} \ln[\text{Prob}(Y = Y_i \mid T = T_i)\text{Prob}(T = T_i)] \\
&= \sum_{i=1}^{N} \ln\text{Prob}(Y = Y_i \mid T = T_i) + \sum_{i=1}^{N} \ln\text{Prob}(T = T_i) \\
&= \quad N00 \ln(1-\alpha) \qquad + N01 \ln(\alpha) \\
&\quad + N10 \ln(1-\alpha-\beta) + N11 \ln(\alpha+\beta) \\
&\quad + N0\bullet \ln(1-\delta) \qquad + N1\bullet \ln(\delta)
\end{aligned}
$$

First order conditions for $\hat{\alpha},\hat{\beta}$ imply, again, the axiomatic solution.

$$\hat{\beta} = TE = \frac{P11}{P1\bullet} - \frac{P01}{P0\bullet}$$

Probit Model by MLE

$$\ln L = \sum_{i=1}^{N} \ln \text{Prob}(Y = Y_i \mid T = T_i)\text{Prob}(T = T_i)$$

$$= \text{N00}\ln[(1-\Phi(\alpha))\times(1-\delta)] + \text{N01}\ln[\Phi(\alpha)\times(1-\delta)]$$

$$+ \text{N10}\ln[(1-\Phi(\alpha+\beta))\times\delta] + \text{N11}\ln[\Phi(\alpha+\beta)\times\delta]$$

(Same outcome for logit or any other valid parametric form.)

$$\text{TE} = \Phi(\hat{\alpha}+\hat{\beta}) - \Phi(\hat{\alpha}) = \frac{\text{P11}}{\text{P1}\bullet} - \frac{\text{P01}}{\text{P0}\bullet}$$

Or, use invariance

$$\text{TE} = \Phi(\alpha+\beta) - \Phi(\alpha)$$

All Roads Lead to Rome

Proof:

Four Treatment Effect Estimators All Identical to Definition

1. **Method of moments based on definitions of probability**
2. **Least squares**
3. **ML based on the axioms**
4. **ML Probit with linear index function, nonlinear**

Conclusion: QED

The LPM is not the "right model" in the saturated case.

Other approaches are equally "right."

A Coherent Saturated Linear Probability Model

The “saturated” model with a linear probability interpretation can be obtained as

(1) Linear Probability $\quad$ $\text{Prob}(y_i = \text{Outcome}|T_i) = a + bT_i$
Coherency $\quad$ $0 < a < 1$ and $0 < a+b < 1$

(2) A Coherent DGP for the randomness of the outcome:

→ (a) $\varepsilon_i \sim U[0,1]$ Uniform[0,1]
(b) $\text{Prob}(y_i=1|T_i) = \text{Prob}(\varepsilon \leq a+bT_i)$.
(c) $y_i = \mathbf{1}[\varepsilon_i \leq a+bT_i]$. (This is the DGP)

Linear probability, nonlinear model.

Coherence needs an external source of variation, ε_i

There is no coherent model that has an additive random component

For the saturated case, it doesn’t matter.

The Linear Probability / Linear Regression Model

- There is no internally consistent process within which the linear regression with additive disturbance applies and the probability is the same linear function. Implicitly, the probability model, if it exists, is

$$\text{Prob}(y=1|\mathbf{x},T) = \alpha + \beta T + \gamma'\mathbf{x};\ y \in \{0,1\}$$
$$\varepsilon \sim U[0,1],\ y = \mathbf{1}[\varepsilon \leq \alpha + \beta T + \boldsymbol{\gamma}'\mathbf{x}]$$
$$\boxed{0 < \alpha + \beta T + \boldsymbol{\gamma}'\mathbf{x} < 1 \text{ for all } (\alpha, \beta, \gamma, T, \mathbf{x})}$$

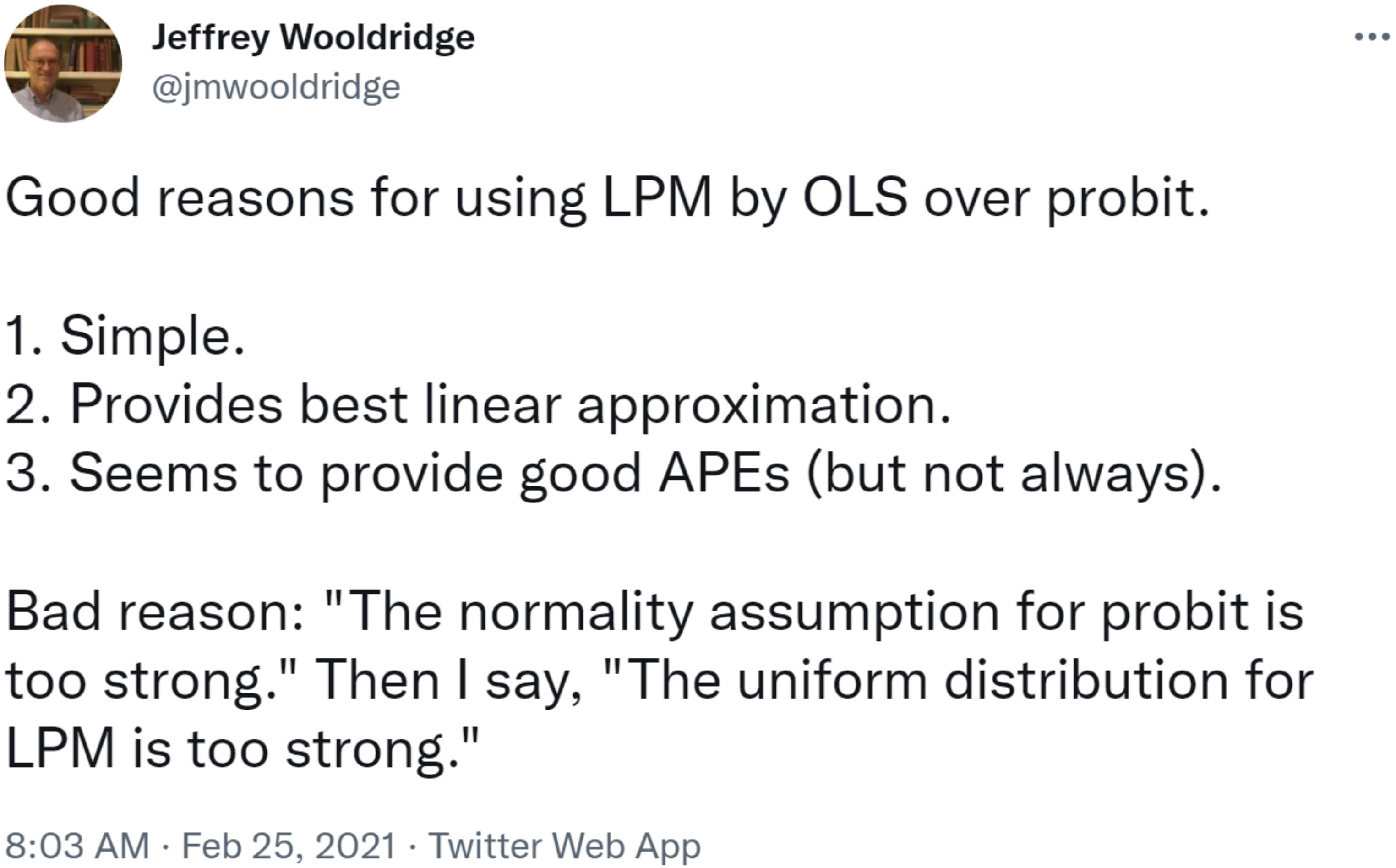
- **The nature and role of ε make no sense**
- **The constraint is impossible for any nontrivial γ'x.**
- **y is not equal to $\alpha + \beta T + \gamma'x + \varepsilon$ for any coherent specification.**

From MHE – On Being Wrong

Obviously, the LPM won’t give the true marginal effects from the right nonlinear model.

But then, the same is true for the wrong nonlinear model!

*The fact that we have a probit, a logit, and the LPM is just a statement to the fact that **we don’t know what the right model is**. Hence, **there is a lot to be said for sticking to a linear regression function** as compared to a fairly arbitrary choice of a non-linear one! Nonlinearity per se is a red herring**.***



Jeffrey Wooldridge
@jmwooldridge

Good reasons for using LPM by OLS over probit.

1. Simple.
2. Provides best linear approximation.
3. Seems to provide good APEs (but not always).

Bad reason: "The normality assumption for probit is too strong." Then I say, "The uniform distribution for LPM is too strong."

8:03 AM · Feb 25, 2021 · Twitter Web App

On The Wrong Model

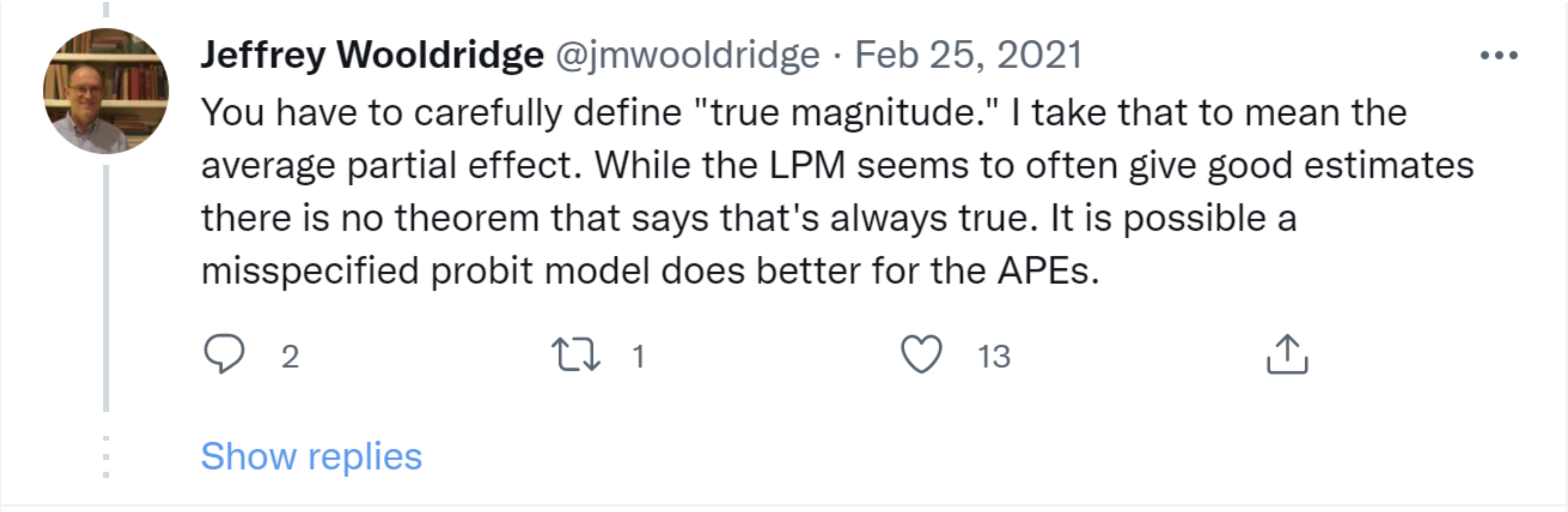
“The same is true for the wrong nonlinear model.”

The wrong nonlinear model won’t give the true marginal effects from the right nonlinear model*.*

[T]he LPM is just a statement to the fact that ***we don’t know what the right model is****.*

We should use the linear model, known to be “wrong,” instead of a nonlinear one that might be wrong.

What Is the True Model?



Jeffrey Wooldridge @jmwooldridge · Feb 25, 2021

You have to carefully define "true magnitude." I take that to mean the average partial effect. While the LPM seems to often give good estimates there is no theorem that says that's always true. It is possible a misspecified probit model does better for the APEs.

Greene's Aphorism #2: With the right parameterization, a nonlinear function can mimic a linear one. A linear function cannot mimic a nonlinear function.

If you believe that neither the linear $\beta'\mathbf{x}$ nor the nonlinear $\Phi(\beta'\mathbf{x})$ is the right model, then by what construction is the linear function a better approximation than the nonlinear sigmoid function?

ANSWER If you believe the wrong nonlinear model is very far from the right nonlinear model and the linear model is always close.

About the “Wrong” Model

A George Box Aphorism

What did George Box mean by “wrong?”
What do Angrist and Pischke mean by “wrong?”
In what way is the model wrong?

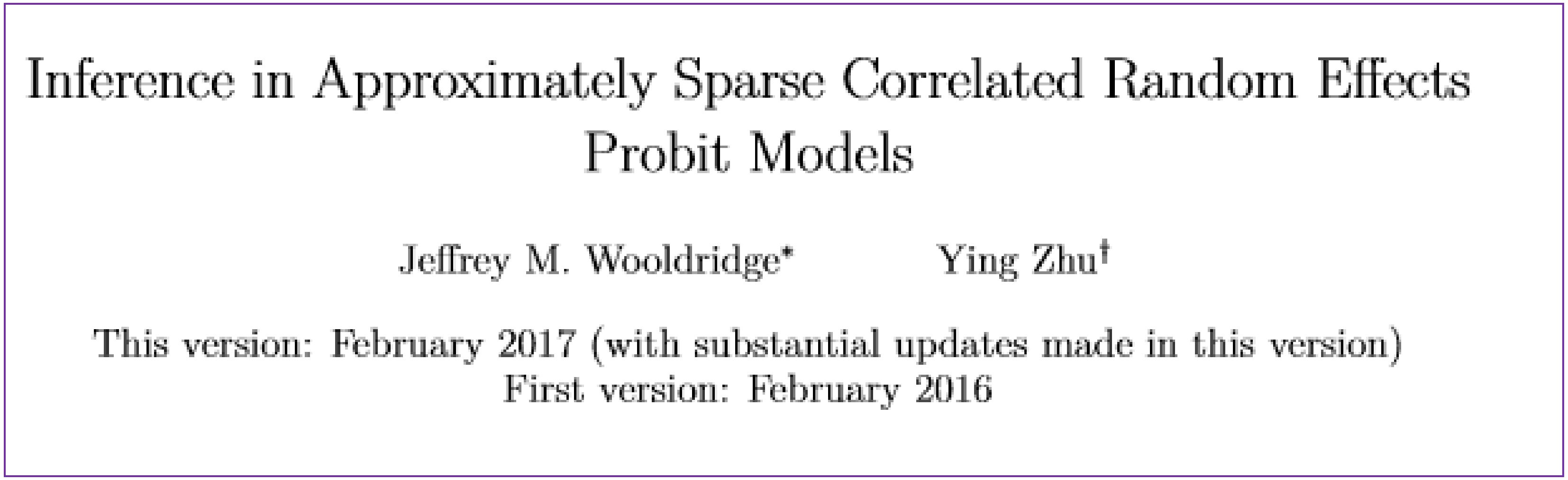
This is a specious argument.

Greene’s Aphorism #3: Continuous distributions do not occur in nature (in social science data).

There is no “right” nonlinear or linear model.

3.1 OLS and Probit/Logit

Does the probit model give the right answer even if the random elements are nonnormal?



Inference in Approximately Sparse Correlated Random Effects Probit Models

Jeffrey M. Wooldridge* Ying Zhu†

This version: February 2017 (with substantial updates made in this version)
First version: February 2016

“**Simulation** results in Li and Zheng (2008) suggest that, for obtaining partial effects, the estimates are not overly sensitive to the normality assumption.”

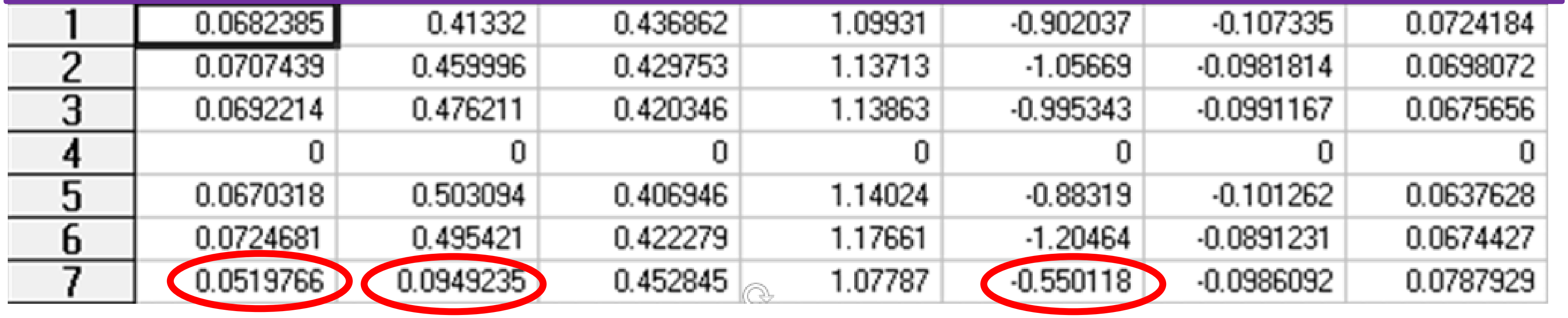
Some Simulations to Explore Wrongness and the Arbitrariness of the Model

(1) Compare 7 different model results using a published real data set.
There is no specified “right” model. (All models may be wrong.)

(2) Compare 7 models when the **X** data are realistic, the true β is known, ε and y are simulated. True model is known. (Some models are right.)

ε are generated by a known “right” model.
y is simulated by threshold DGP for several cases
Average of partial effects in 10 repetitions.

Searching for the Right Model
Partial Effects in German Innovation Data



	logSales	SP	IMUM	FDIUM	PROD	RawMtl	InvGood
1	0.0682385	0.41332	0.436862	1.09931	-0.902037	-0.107335	0.0724184
2	0.0707439	0.459996	0.429753	1.13713	-1.05669	-0.0981814	0.0698072
3	0.0692214	0.476211	0.420346	1.13863	-0.995343	-0.0991167	0.0675656
4	0	0	0	0	0	0	0
5	0.0670318	0.503094	0.406946	1.14024	-0.88319	-0.101262	0.0637628
6	0.0724681	0.495421	0.422279	1.17661	-1.20464	-0.0891231	0.0674427
7	0.0519766	0.0949235	0.452845	1.07787	-0.550118	-0.0986092	0.0787929

1 = Probit
2 = Logit
3 = Burr
4 = Complementary log log
5 = Extreme value
6 = Arc tangent
7 = Linear

They all look “right” except for the linear model which seems to miss the mark in the two cases we saw earlier. (Model 4 inestimable with these data.)

Health Care: GSOEP, 1994 Data

True ε is unknown: Actual Data: Doctor Visits > 0

	FEMALE	AGE	EDUC	MARRIED	INCOME
Probit	0.12941	0.00507	-0.00172	-0.02159	-0.06561
Logit	0.12991	0.00512	-0.00158	-0.02296	-0.06705
Burr	0.13072	0.00522	-0.00129	-0.02502	-0.06764
Comp Log Log	0.12759	0.00490	-0.00218	-0.01769	-0.06180
Extreme Value	0.13079	0.00522	-0.00125	-0.02518	-0.06762
ArcTangent	0.13033	0.00516	-0.00146	-0.02411	-0.06821
Linear	0.12965	0.00502	-0.00147	-0.01996	-0.06516

They all look “right” with minor differences.

Simulation: Probit is the Right Model

Simulated True ε: Standard Normal Random Terms

	FEMALE	AGE	EDUC	MARRIED	INCOME
Probit	0.12941	0.00507	-0.00172	-0.02159	-0.06561
Logit	0.12991	0.00512	-0.00158	-0.02296	-0.06705
Burr	0.13072	0.00522	-0.00129	-0.02502	-0.06764
Comp Log Log	0.12759	0.00490	-0.00218	-0.01769	-0.06180
Extreme Value	0.13079	0.00522	-0.00125	-0.02518	-0.06762
ArcTangent	0.13033	0.00516	-0.00146	-0.02411	-0.06821
Linear	0.12965	0.00502	-0.00147	-0.01996	-0.06516

They all look “right” with minor differences.

Simulation: Logit is the Right Model

Simulated True ε : Standard Logistic Random Terms

	FEMALE	AGE	EDUC	MARRIED	INCOME
Probit	0.11765	0.00607	-0.00424	-0.00937	-0.03086
Logit	0.11787	0.00614	-0.00427	-0.01024	-0.03289
Burr	0.11794	0.00627	0.00440	-0.01078	-0.03541
Comp Log Log	0.11638	0.00586	-0.00410	-0.00754	-0.02537
Extreme Value	0.11797	0.00629	-0.00440	-0.01110	-0.03573
ArcTangent	0.11791	0.00619	-0.00432	-0.01088	-0.03432
Linear	0.11769	0.00602	-0.00418	-0.00609	-0.02903

They all look “right” with minor differences. Coefficient on MARRIED is a bit erratic. Model 4 seems a bit erratic. Linear seems worse than complementary log log.

Simulation: Finite Mixture of Normals

Simulated True ε: .5/.5 Mixture of N[-5,1] and N[+5,2]

	FEMALE	AGE	EDUC	MARRIED	INCOME
Probit	0.05217	0.00421	0.00313	-0.06660	-0.05412
Logit	0.06182	0.00420	0.00314	-0.06638	-0.05338
Burr	0.06973	0.00436	0.00267	-0.06699	-0.07141
Comp Log Log	0.06506	0.00427	0.00297	-0.06718	-0.05960
Extreme Value	0.05926	0.00415	0.00328	-0.06547	-0.04951
ArcTangent	0.06147	0.00418	0.00315	-0.06620	-0.05266
Linear	0.06166	0.00418	0.00315	-0.06613	-0.05287

They all look “right” with minor differences.

Simulation: Implausible Mixture

True ε : Mixture of Standard Normal and Chi Squared[1]

	FEMALE	AGE	EDUC	MARRIED	INCOME
Probit	0.21147	0.00875	-0.01035	-0.01430	-0.03875
Logit	0.20984	0.00865	-0.01021	-0.01385	-0.03927
Burr	0.21198	0.00830	-0.00991	-0.00732	-0.02583
Comp Log Log	0.21460	0.00873	-0.01038	-0.01139	-0.03757
Extreme Value	0.20562	0.00863	-0.01012	-0.01713	-0.03625
ArcTangent	0.20849	0.00856	-0.01009	-0.01344	-0.03967
Linear	0.20942	0.00859	-0.01034	-0.00770	-0.03309

They all look "right" with minor differences.

Simulation: Does the LPM Work If It Is Right?

True ε : Linear Probability Model

	FEMALE	AGE	EDUC	MARRIED	INCOME
Probit	0.23247	0.00882	-0.00227	-0.00048	-0.18921
Logit	0.23197	0.00877	-0.00235	-0.00132	-0.18723
Burr	0.23263	0.00855	-0.00138	+0.00427	-0.18902
Comp Log Log	0.23247	0.00854	-0.00109	+0.00600	-0.19055
Extreme Value	0.22905	0.00892	-0.00327	-0.00603	-0.18350
ArcTangent	0.23148	0.00876	-0.00241	-0.00200	-0.18563
Linear	0.23176	0.00878	-0.00237	+0.02401	-0.14620

$Y = \mathbf{1}[U \leq \beta'\mathbf{x}]$

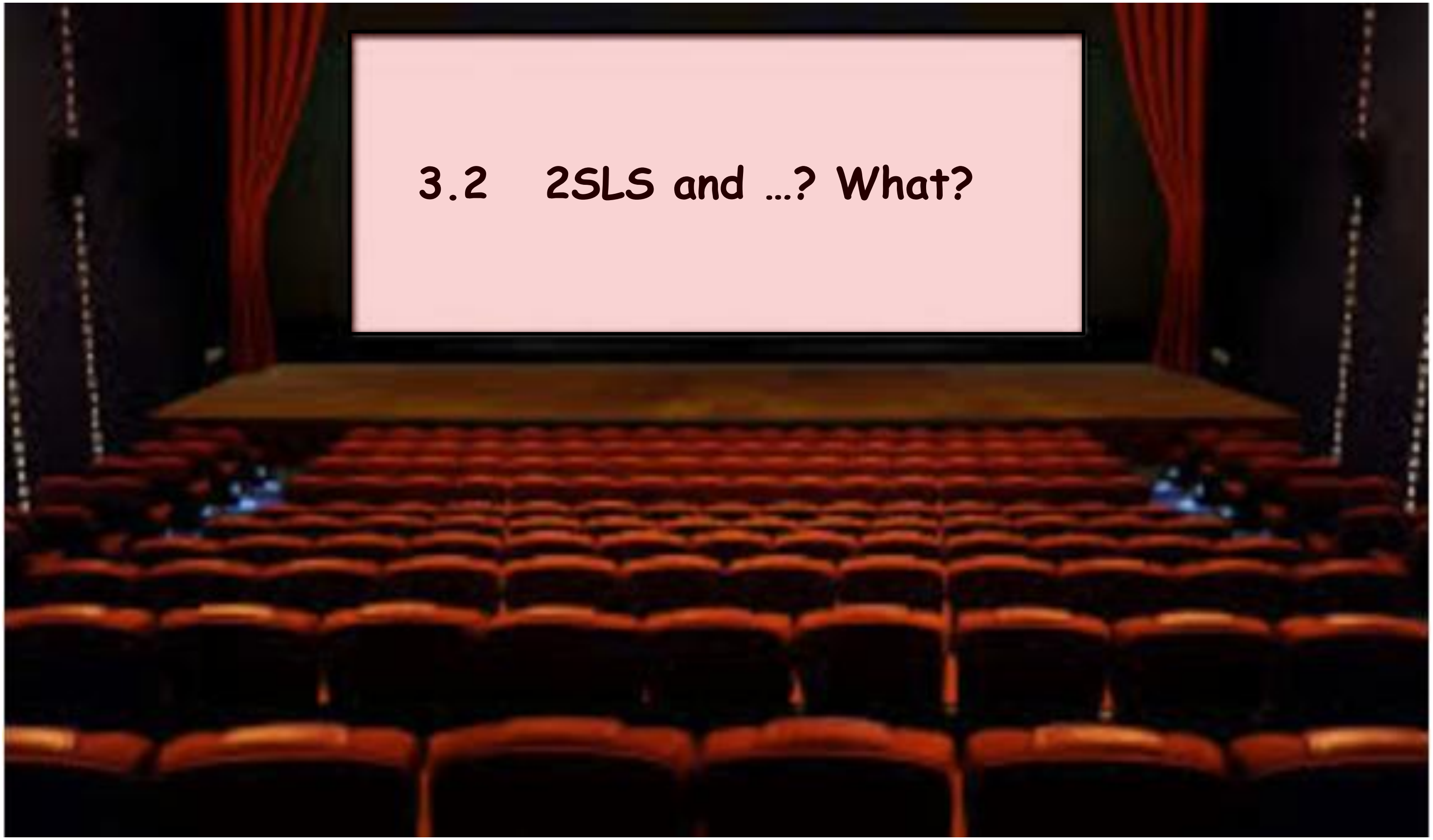
Data are scaled to force $0 < \beta'\mathbf{x} < 1$.

Results are erratic. LPM is not best when it is right.

The Wrong (and Arbitrary) Model

The choice of nonlinear model is arbitrary and likely to be ‘wrong,’ whereas the LPM seems to be robust.

- The claim exaggerates the differences in partial effects across functional forms.
- The implication of ‘wrongness’ appears to be exaggerated.
- Arguments about the ‘right’ vs. ‘wrong’ model are specious and misleading in the use of the term “wrong,”



3.2 2SLS and …? What?

The centerpiece of this entire exercise is IV (2SLS) estimation, not OLS.

- Credible econometrics is about causal inference, treatment effects and the awesome power of instrumental variables.
- OLS seems to ‘work’ for the base case (no endogeneity on RHS)
- 2SLS is motivated by the logic of IVs. (Method of moments)
- Does 2SLS ‘work’ in the same way for the treatment effect case?
    - What is the underlying model?
    - What is the treatment effect?
- **Are there demonstrations that 2SLS is “robustly true?”**

Is 2SLS Fairly Robustly True?

"… instrumental variables methods methods estimate an average causal effect for a well-defined population even if the instrument does not affect everyone." (Angrist and Pischke (2005), p. xiii.)

Compared to what? Not a simple probit. It is now a 2 equation model. (Endogeneity of the treatment dummy.) A&P comparisons:

- A Variety of models and strategies that avoid distributions.
  A&P state that this work should all be ignored because they do not pursue partial effects
- Recursive bivariate probit (RBP): Maddala (1983), Greene (2018), others
- A&P: RBP + an exclusion restriction (IV in 2nd equation)
  (This model is "harmless." p. 201 – see slide 74)
- Has it been shown that 2SLS estimates the partial effects of the "right" probability model counterpart?

Comparisons of 2SLS vs. Bivariate Probit

http://people.stern.nyu.edu/wgreene/DiscreteChoice/2015/ME-2-2-NonlinearEffects-Endogeneity.pptx

Angrist and Pischke (p. 203).

- ATET requires a distribution assumption (bivariate probit)
- 2SLS Estimates LATE.
- Comparison 2SLS to a bivariate normal CEF.
  - Difference between ATET and LATE
  - Difference between linear and true nonlinear model

Treatment Effects in Binary Choice

T is a “treatment”

$$\text{Prob}(Y=1|\mathbf{x},T) = \Phi(\boldsymbol{\beta}'\mathbf{x} + \theta T)$$

Treatment effect of T on y?

$$\text{Prob}(Y = 1|\mathbf{x},T=1) – \text{Prob}(Y = 1|\mathbf{x},T=0)$$

$$= \Phi(\boldsymbol{\beta}'\mathbf{x} + \theta) - \Phi(\boldsymbol{\beta}'\mathbf{x})$$

Treatment effect on the treated.

Compare being treated to being untreated for someone who was actually treated. (COUNTERFACTUAL)

$$\text{Prob}(Y = 1|\mathbf{x},T = 1)_{|T=1} - \text{Prob}(Y = 1|\mathbf{x},T=0)_{|T=1}$$

Measuring ATET

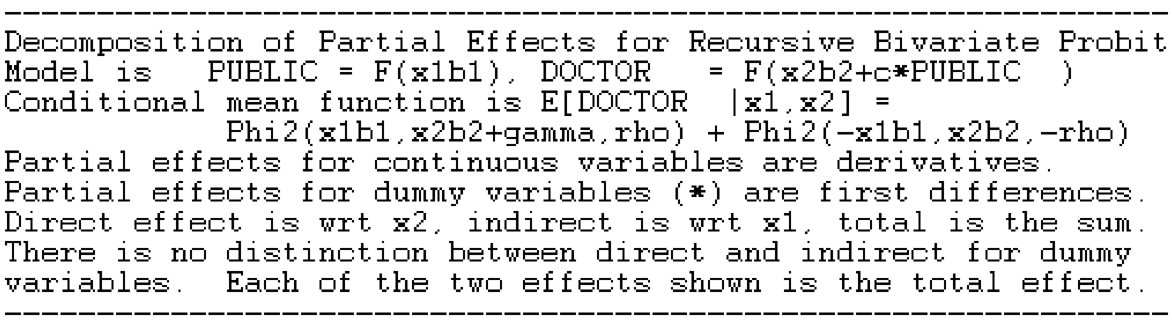
This result appears in full in A&P (p. 201, (4.6.16)). They then argue that it’s OK to settle for LATE based on a linear approach (2SLS), with the **benefit of dropping the normality assumption**. A&P argue that 2SLS would be preferred to the BVP. But, ATET can only be obtained with a distributional assumption.

$$\text{E}[\text{Y}_{1\text{i}} - \text{Y}_{0\text{i}} \mid \text{D} = 1]$$

$$= \text{E}\left\{ \frac{\Phi_b\left(\dfrac{\text{x}'_i\beta_0^* + \beta_1^*}{\sigma_\varepsilon}, \dfrac{\text{x}'_i\gamma_0^* + \gamma_1^* z_i}{\sigma_\varepsilon}; \rho_{\varepsilon v} \right) - \Phi_b\left(\dfrac{\text{x}'_i\beta_0^*}{\sigma_\varepsilon}, \dfrac{\text{x}'_i\gamma_0^* + \gamma_1^* z_i}{\sigma_\varepsilon}; \rho_{\varepsilon v} \right)}{\Phi\left(\dfrac{\text{x}'_i\gamma_0^* + \gamma_1^* z_i}{\sigma_\varepsilon} \right)} \right\}$$

(σ_ε is obviously not identified (not estimable).)

FIML Partial Effects

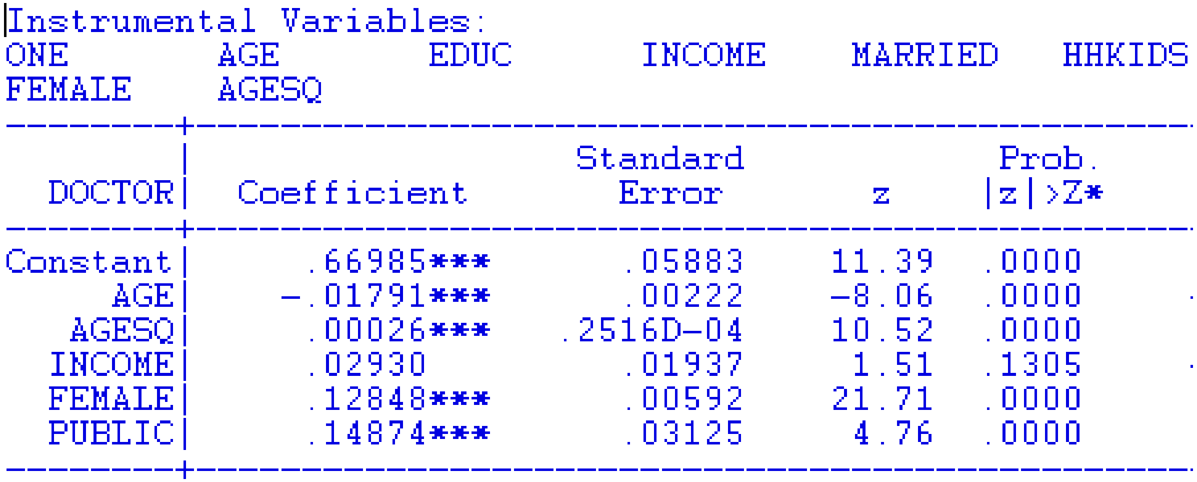


```
-----------------------------------------------------------------
Decomposition of Partial Effects for Recursive Bivariate Probit
Model is   PUBLIC = F(x1b1), DOCTOR   = F(x2b2+c*PUBLIC  )
Conditional mean function is E[DOCTOR  |x1,x2] =
            Phi2(x1b1,x2b2+gamma,rho) + Phi2(-x1b1,x2b2,-rho)
Partial effects for continuous variables are derivatives.
Partial effects for dummy variables (*) are first differences.
Direct effect is wrt x2, indirect is wrt x1, total is the sum.
There is no distinction between direct and indirect for dummy
variables.  Each of the two effects shown is the total effect.
-----------------------------------------------------------------
```

Partial Effects

	FIML (ATET)	2SLS (LATE)
Income	0.02349	0.02930
Female	0.13059	0.12848

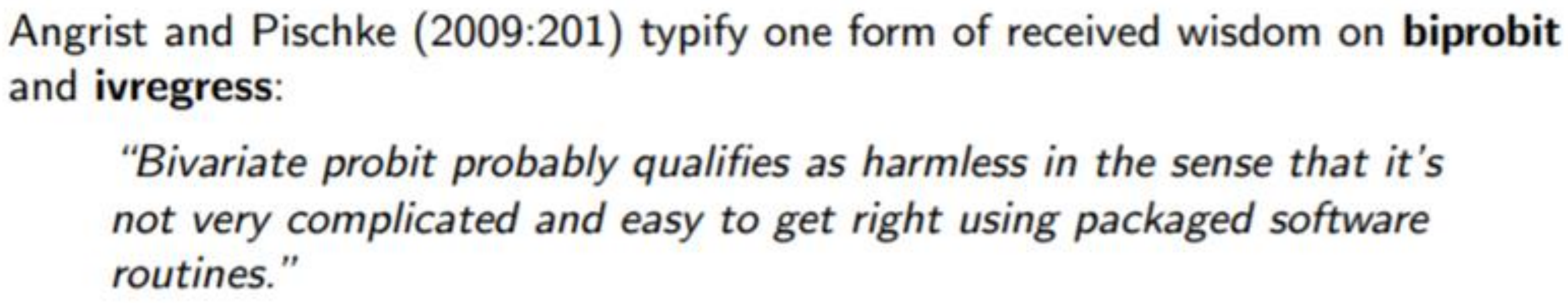
Two Stage Least Squares Effects



```
Instrumental Variables:
ONE          AGE          EDUC         INCOME       MARRIED      HHKIDS
FEMALE       AGESQ
---------+------------------------------------------------------------
         |                  Standard                  Prob.
   DOCTOR|  Coefficient       Error           z       |z|>Z*
---------+------------------------------------------------------------
 Constant|      .66985***      .05883       11.39    .0000
      AGE|     -.01791***      .00222       -8.06    .0000
    AGESQ|      .00026***   .2516D-04       10.52    .0000
   INCOME|      .02930         .01937        1.51    .1305
   FEMALE|      .12848***      .00592       21.71    .0000
   PUBLIC|      .14874***      .03125        4.76    .0000
---------+------------------------------------------------------------
```

(Politically) “Harmless” Econometrics

From Dave Giles’s *Econometrics Beat*, blog.



Angrist and Pischke (2009:201) typify one form of received wisdom on **biprobit** and **ivregress**:

> *“Bivariate probit probably qualifies as harmless in the sense that it’s not very complicated and easy to get right using packaged software routines.”*

What defines a method/model as “harmful?”
Why is complexity harmful?
What does “get right” mean?

Why is the univariate probit model harmful?
(“dangerous” (MHE: cover))

4. Econometric Methodology

Econometric Methodology for Binary Choice

- The credibility revolution prescribes research design based strictly on the orthodoxy of causal inference.

- Credible research design seems comfortable with unclear econometric execution (OLS, 2SLS) justified by observed experience and speculative approximations, not by derived econometric theory.

  - A&P suggest that bivariate probit is the natural approach then argue that 2SLS seems more “robust”

  - Avoiding nonlinear functional form trumps econometric validity. … ***there is a lot to be said for sticking to a linear regression function*** *as compared to a fairly arbitrary choice of a non-linear one! Nonlinearity per se is a red herring.”*

On Least Squares for Binary Choice

OLS is not an econometric 'estimator.' It is a statistic.

- "Not a theorem" means we have no theory to use to establish econometric properties of OLS
- There is no theory to establish a claim that OLS estimates partial effects. It only "seems to work."
- There is no theoretical basis (formula) for computing standard errors of any sort, robust, clustered or otherwise. Possibly bootstrapping might be useful.

On the Right Model

- There is no “right” binary choice model. Arguments that the probit, logit or arctangent model might be “wrong” are specious.

- The probit model seems as close as any for a useful binary choice model that reveals useful information about the world.

- Use the recursive bivariate probit model for endogenous binary treatment effects. There is no way to claim 2SLS gives a ‘right’ answer.

On Robustness to Distribution

Nonlinearity is not a red herring (a clue meant deliberately to deceive or mislead). The world is nonlinear.

- We know how to do this. In 2022, methods are not non-credible or harmful because they are complex.
- Nonlinear models are not more wrong than LPM. There is no coherent model that supports OLS or 2SLS
- If we are going to rely on an approximation of unknown validity, why is linear better than nonlinear?

On Going Linear

Is this all about the linear binary choice model?*

Apparently. There seems to be no constituency for a linear ordered choice model or a linear multinomial choice or count data model.

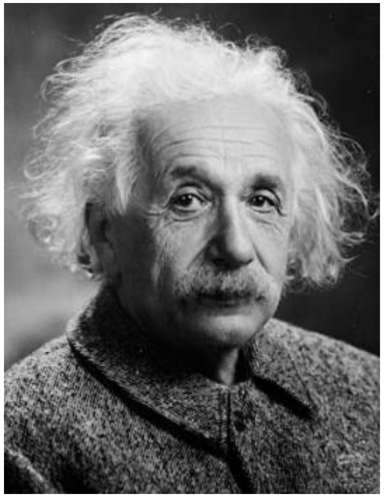
Not quite: Compare the rigorous orthodoxy of modern theoretical econometrics to the “robust,” model free methodology of harmless econometrics.

***See, e.g.,** https://davegiles.blogspot.com/2012/06/another-gripe-about-linear-probability.html

Wisdom

“If the estimates you get are not the estimates you want, the fault lies in the econometrician and not the econometrics.”
… Angrist and Pischke, p. xii.

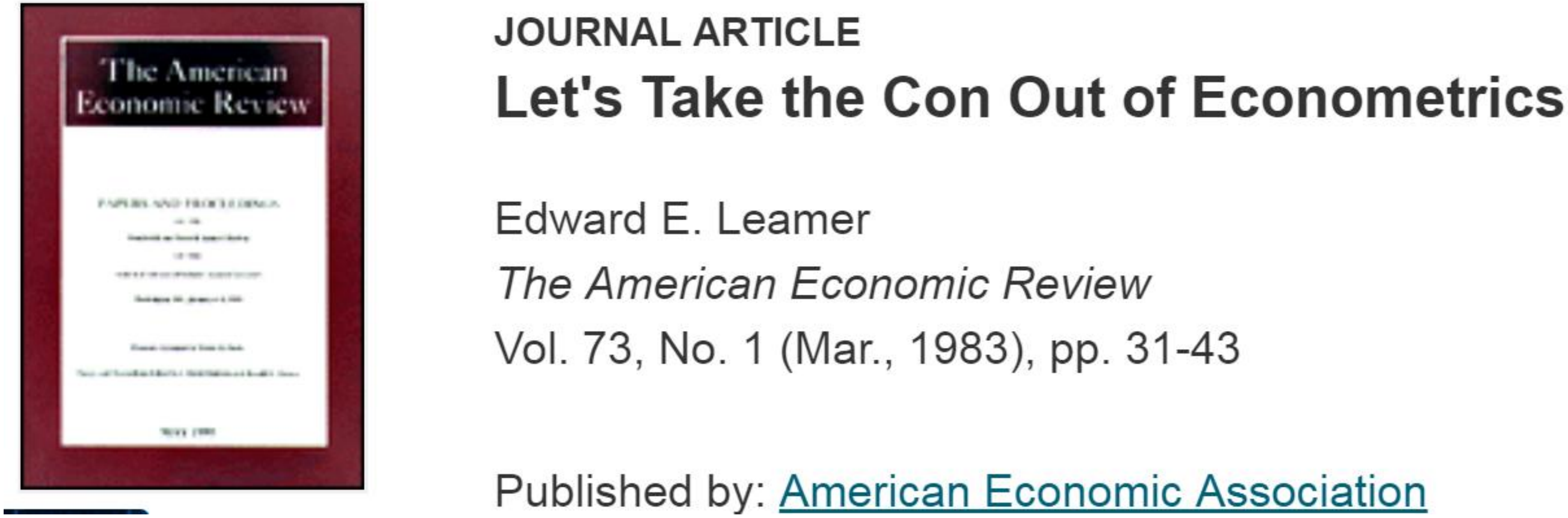
It’s not just nitpicking about functional form. … W. Greene



“Everything should be as simple as can be, but not simpler.”

OUTTAKES

The Con in Econometrics?
An Unfortunate Econometric Mea Culpa.



JOURNAL ARTICLE

Let's Take the Con Out of Econometrics

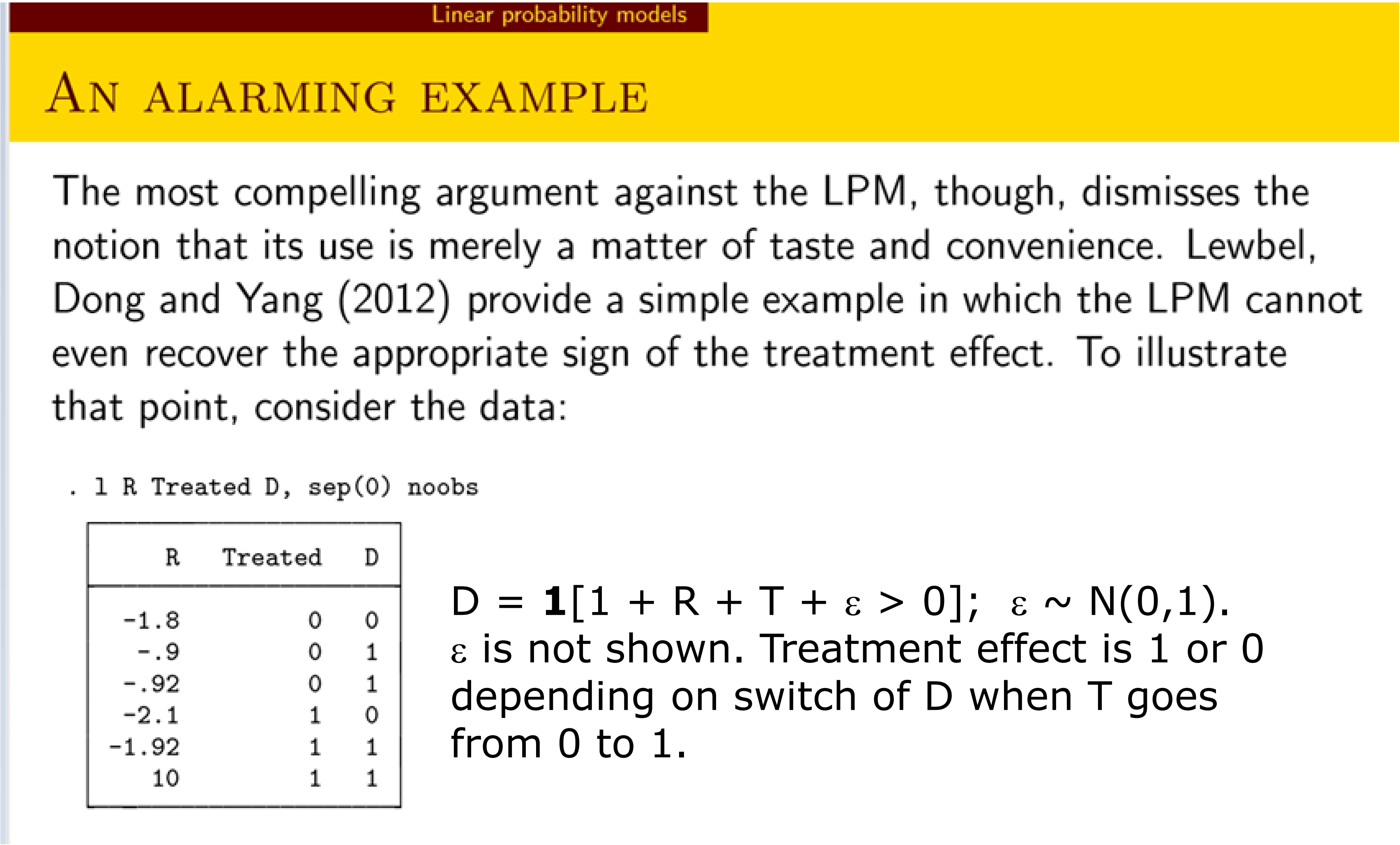
Edward E. Leamer

The American Economic Review

Vol. 73, No. 1 (Mar., 1983), pp. 31-43

Published by: American Economic Association

Leamer argued that sometimes misguided specifications needed to be more detailed and deal more effectively with assumptions and identification. He suggested ways that econometric inferences were fragile and influenced by assumptions. He was critical of the way specification searches tainted econometric inference. He did not suggest that econometrics was a dishonest confidence game.



Linear probability models

An alarming example

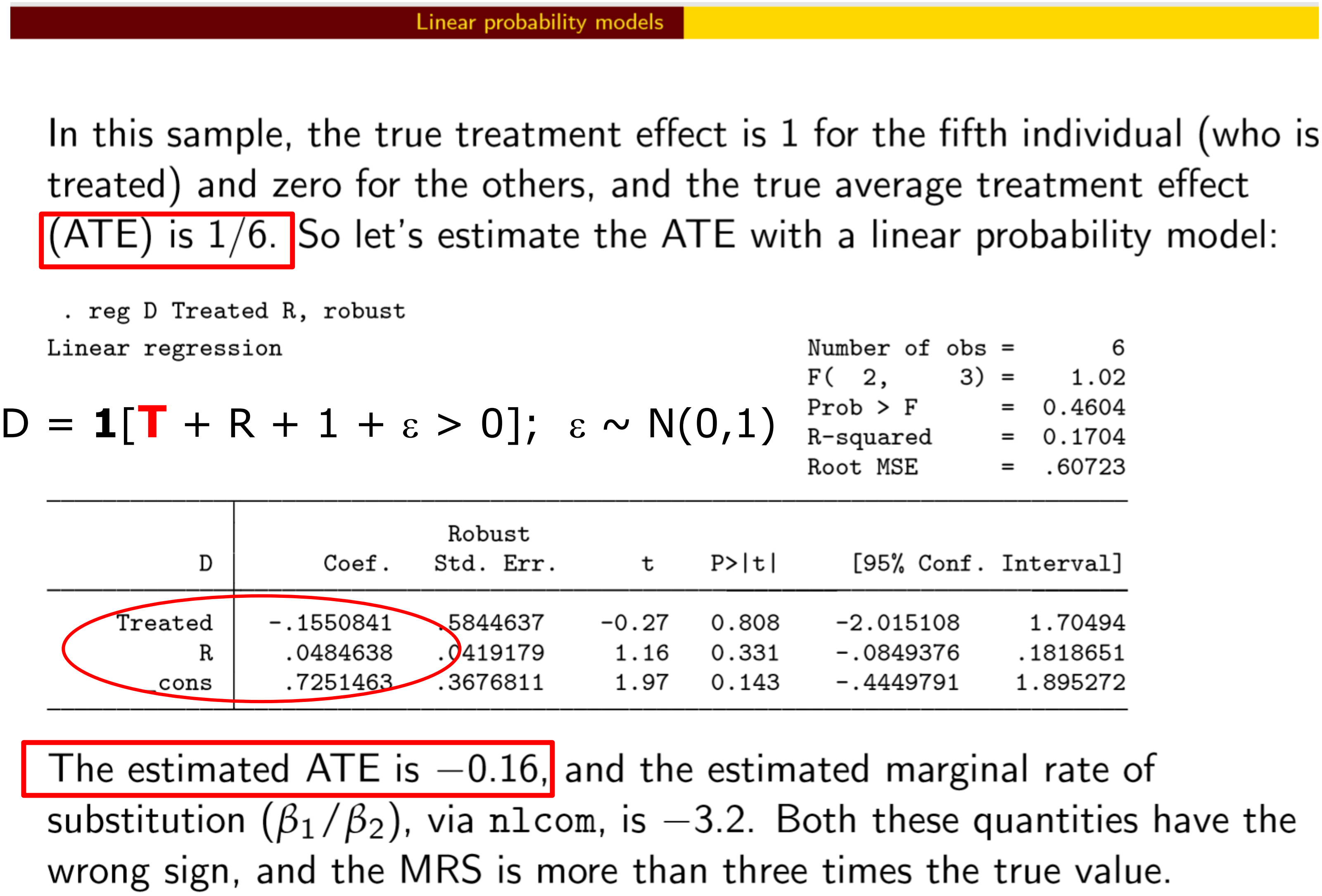
The most compelling argument against the LPM, though, dismisses the notion that its use is merely a matter of taste and convenience. Lewbel, Dong and Yang (2012) provide a simple example in which the LPM cannot even recover the appropriate sign of the treatment effect. To illustrate that point, consider the data:

```
. l R Treated D, sep(0) noobs

  +---------------------+
  |     R   Treated   D |
  |---------------------|
  |  -1.8         0   0 |
  |   -.9         0   1 |
  |  -.92         0   1 |
  |  -2.1         1   0 |
  | -1.92         1   1 |
  |    10         1   1 |
  +---------------------+
```

$D = \mathbf{1}[1 + R + T + \varepsilon > 0]$; $\varepsilon \sim N(0,1)$.
ε is not shown. Treatment effect is 1 or 0 depending on switch of D when T goes from 0 to 1.

Thanks to Arthur Lewbel for this example. Lewbel, Dong & Yang, "Comparing features of Convenient Estimators for Binary Choice Models With Endogenous Regressors", ***Canadian Journal of Economics, 2012***



In this sample, the true treatment effect is 1 for the fifth individual (who is treated) and zero for the others, and the true average treatment effect (ATE) is 1/6. So let's estimate the ATE with a linear probability model:

```
. reg D Treated R, robust

Linear regression                                      Number of obs =       6
                                                       F(  2,      3) =    1.02
                                                       Prob > F      =  0.4604
                                                       R-squared     =  0.1704
                                                       Root MSE      =  .60723

------------------------------------------------------------------------------
             |               Robust
           D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     Treated |  -.1550841    .5844637    -0.27   0.808    -2.015108     1.70494
           R |   .0484638    .0419179     1.16   0.331    -.0849376    .1818651
        cons |   .7251463    .3676811     1.97   0.143    -.4449791    1.895272
------------------------------------------------------------------------------
```

$$D = \mathbf{1}[T + R + 1 + \varepsilon > 0]; \quad \varepsilon \sim N(0,1)$$

The estimated ATE is -0.16, and the estimated marginal rate of substitution (β_1/β_2), via `nlcom`, is -3.2. Both these quantities have the wrong sign, and the MRS is more than three times the true value.

The ML estimated ATE from the true probit model is +.33.

Convenience of Computation Is Not a Virtue (In 2022)

- Bypassing the complexity of nonlinear models is not a worthy objective. The world is already full of great programmers and great software.

- It is easy to include fixed effects.

Fixed Effects in nonlinear models are indeed complicated.
But this is a solvable problem. See Greene (2004)

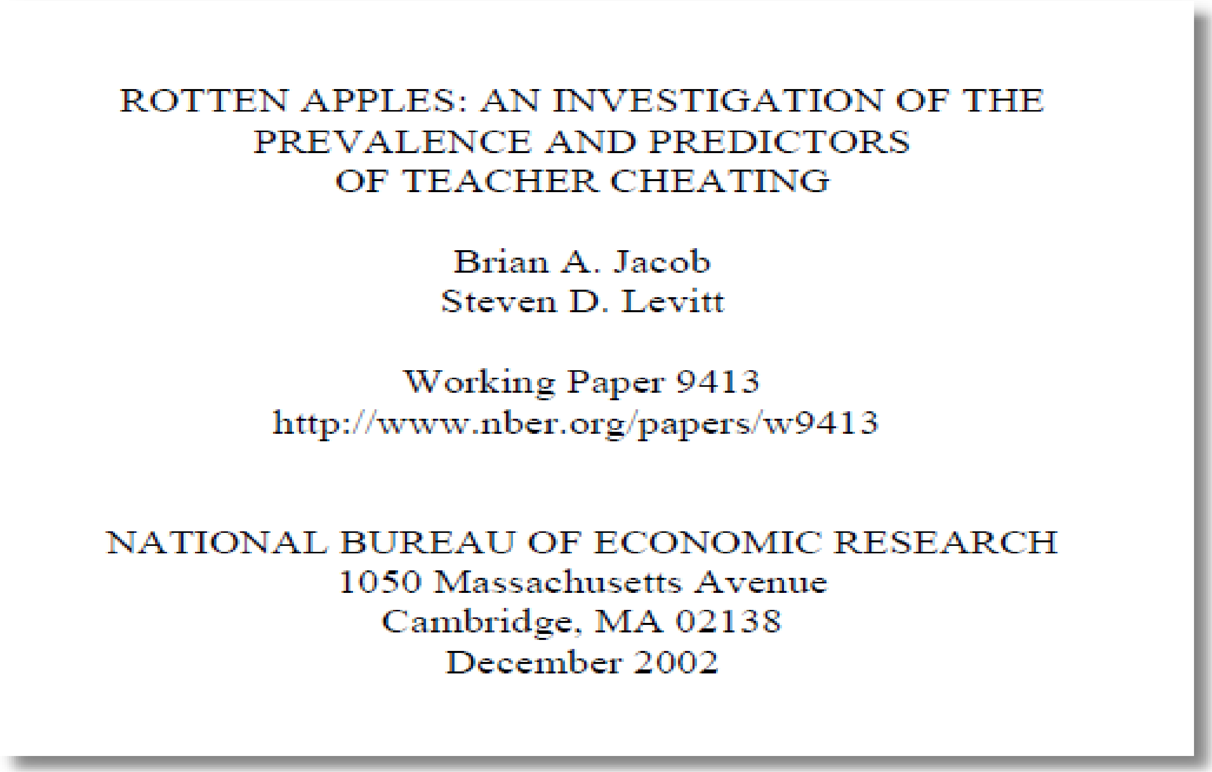
“I don’t have to worry about ‘incidental parameters problems.”
You usually don’t anyway. See Greene (2004,2005) on the IP problem.

Modern researchers are comfortable with correlated random effects as a very useful approach to this problem.

Panel data may tempt the analyst to use diff-in-diff. That makes less sense here. The outcome is not a quantity so differences are meaningless.

Erroneously Interpreting the LPM: It’s not just the magnitudes.

Teachers in Chicago



ROTTEN APPLES: AN INVESTIGATION OF THE
PREVALENCE AND PREDICTORS
OF TEACHER CHEATING

Brian A. Jacob
Steven D. Levitt

Working Paper 9413
http://www.nber.org/papers/w9413

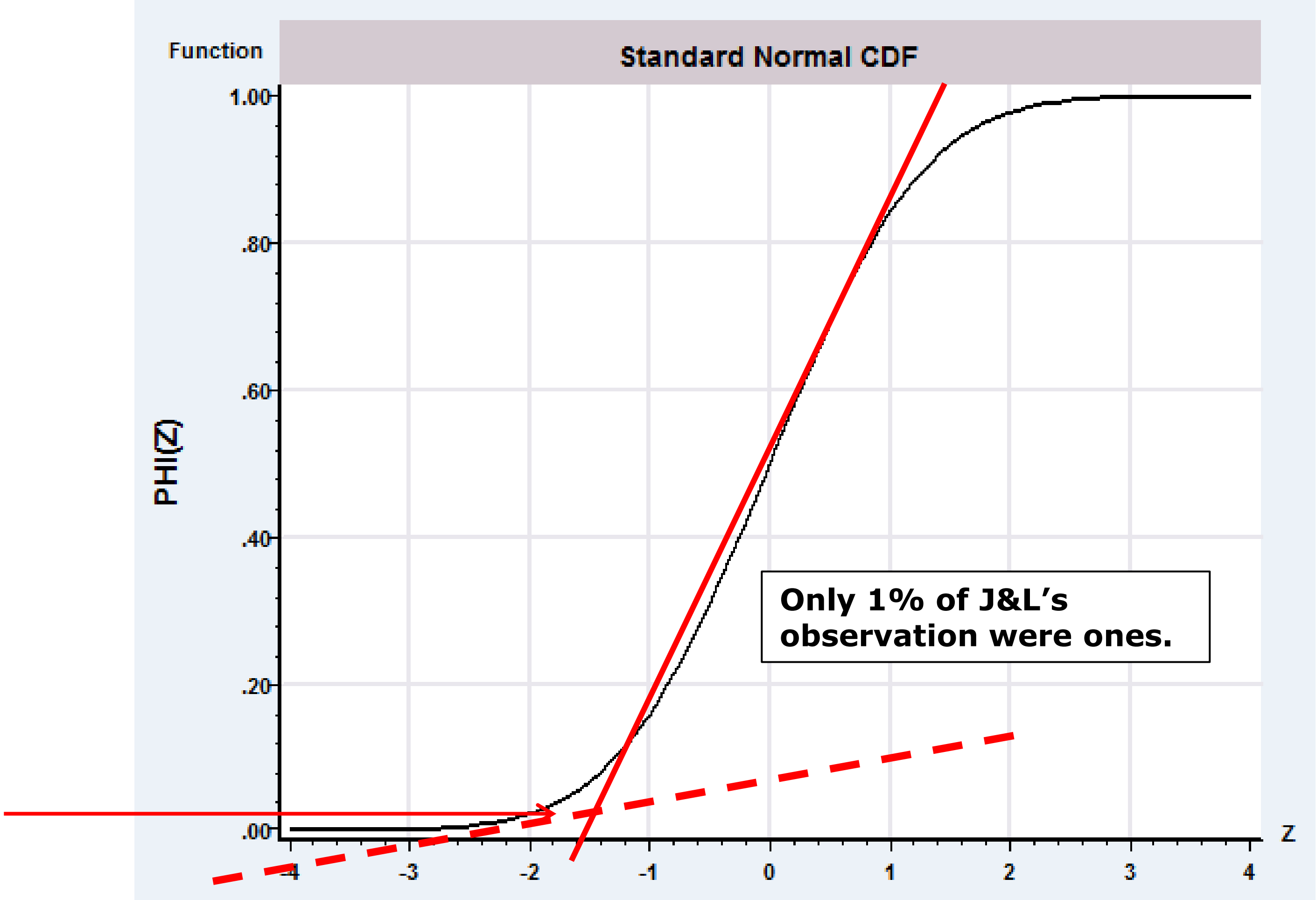
NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2002

Cheating in Chicago

From Jacob and Levitt
P. 32. About Table 7. OLS Used. “Probits yield similar marginal effects.”
P. 37. About Table 9. OLS Used. “Logit models evaluated at the mean yield comparable results.” OLS results presented “for ease of comparison.”
P. 41. About Table 10. The main results… “[e]quations are estimated using 2SLS.” No mention of logit or probit.

Finding that 2SLS resembles probit or logit would not be good news. See A&P, Table 4.6.1, columns (2),(3) p. 203. Probit should give the wrong answer because it ignores the endogeneity of the treatment.

The obvious problem with Jacob/Levitt's LPM with highly unbalanced data. Bliss (1934) anticipated this.



Only 1% of J&L's observation were ones.

Nonlinearity is needed in the tails of the distribution.

(Jacob and Levitt) "*In column 1, teachers are roughly 6 percentage points more likely to cheat for students who scored in the second quartile (between the 25th and 50th percentile) in the prior year…*" (p. 41)

According to the model, Prob(Y=1|D=0) is .01 and Prob(Y=1|D=1) is .01+.06 = .07. The treatment effect, the change in the probability, is 600%, not 6%!

Sample ~ 40,000
Responses ~ 400.
Mean Y ~ 0.01.
Is a partial effect of 0.06 on a dummy variable moderate?"
It's only "6%."

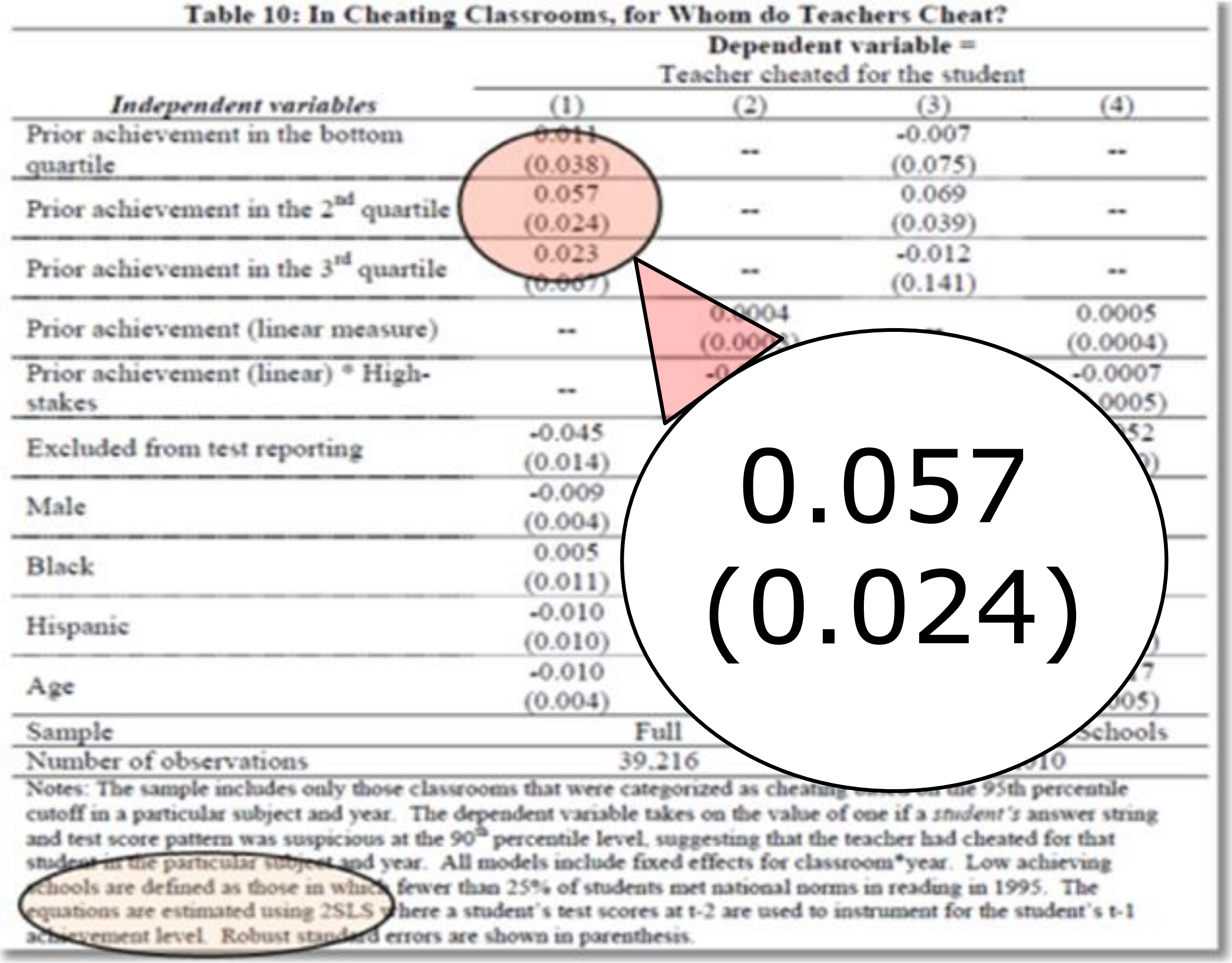


Table 10: In Cheating Classrooms, for Whom do Teachers Cheat?

Independent variables	Dependent variable = Teacher cheated for the student (1)	(2)	(3)	(4)
Prior achievement in the bottom quartile	0.011 (0.038)	--	-0.007 (0.075)	--
Prior achievement in the 2nd quartile	0.057 (0.024)	--	0.069 (0.039)	--
Prior achievement in the 3rd quartile	0.023 (0.067)	--	-0.012 (0.141)	--
Prior achievement (linear measure)	--	0.0004 (0.00[illegible])	[illegible]	0.0005 (0.0004)
Prior achievement (linear) * High-stakes	--	-0[illegible]	[illegible]	-0.0007 (0.0005)
Excluded from test reporting	-0.045 (0.014)	[illegible]	[illegible]	[illegible]52 [illegible])
Male	-0.009 (0.004)	[illegible]	[illegible]	[illegible]
Black	0.005 (0.011)	[illegible]	[illegible]	[illegible]
Hispanic	-0.010 (0.010)	[illegible]	[illegible]	[illegible])
Age	-0.010 (0.004)	[illegible]	[illegible]	[illegible]7 [illegible]05)
Sample	Full		[illegible]	Schools
Number of observations	39,216		[illegible]010	

Notes: The sample includes only those classrooms that were categorized as cheating [illegible] the 95th percentile cutoff in a particular subject and year. The dependent variable takes on the value of one if a *student's* answer string and test score pattern was suspicious at the 90th percentile level, suggesting that the teacher had cheated for that student in the particular subject and year. All models include fixed effects for classroom*year. Low achieving schools are defined as those in which fewer than 25% of students met national norms in reading in 1995. The equations are estimated using 2SLS where a student's test scores at t-2 are used to instrument for the student's t-1 achievement level. Robust standard errors are shown in parenthesis.

0.057
(0.024)

2SLS for a binary dependent variable.

A Common Misconception

- (Jacob and Levitt) “*In column 1, teachers are roughly 6 percentage points more likely to cheat for students who scored in the second quartile (between the 25th and 50th percentile) in the prior year…*” (p. 41)

- According to the model, Prob(Y=1|D=0) is .01 and Prob(Y=1|D=1) is .01+.06 = .07. It makes no sense to suggest that this treatment effect is 6%. The treatment effect, the change in the probability, is 600%, not 6%!

“For instance , if equation (3) yields β = .01, we immediately understand that the treatment caused an increase of 1 percentage point in the probability to observe Y = 1....”

***Gomila, R., “Logistic or Linear? Estimating Causal Effects of Treatments on Binary Outcomes Using Regression Analysis,” Msp. Department of Psychology, Princeton, 2019.**

An Endogenous Binary Variable – IVProbit: Angrist and Pischke agree with this.

$$
\begin{aligned}
y^* &= \boldsymbol{\beta}'\mathbf{x} + \theta T + \varepsilon \\
y &= 1[y^* > 0] \\
T^* &= \boldsymbol{\alpha}'\mathbf{z} \qquad\quad + u \\
T &= 1[T^* > 0]
\end{aligned}
$$

⇐ **Correlation = ρ.**

$E[\varepsilon|T,\mathbf{x}] \neq 0 \Leftrightarrow \text{Cov}[u, \varepsilon] \neq 0$

Additional Assumptions:

$(u,\varepsilon) \sim N[(0,0),(\sigma_u^2, \rho\sigma_u, 1)]$

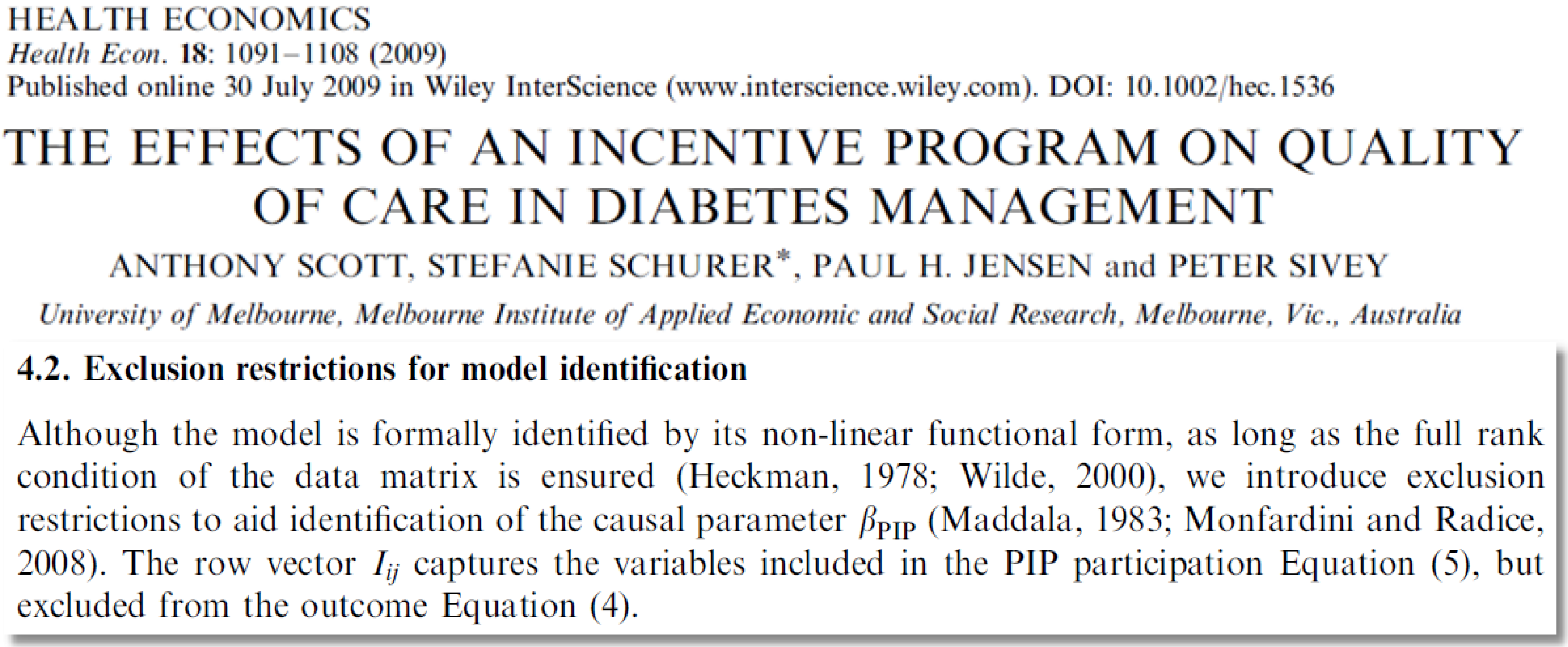
$\mathbf{z}$ = a valid set of exogenous (and excluded) variables

Estimation:

Harmless: 2SLS **y** on **x** using **z** (**z** contains variables not in **x**.)

Harmful: FIML; recursive bivariate probit.
This is what Stata calls IVProbit**.**

Identification by IV & Functional Form



HEALTH ECONOMICS
Health Econ. **18**: 1091–1108 (2009)
Published online 30 July 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hec.1536

THE EFFECTS OF AN INCENTIVE PROGRAM ON QUALITY OF CARE IN DIABETES MANAGEMENT

ANTHONY SCOTT, STEFANIE SCHURER*, PAUL H. JENSEN and PETER SIVEY

University of Melbourne, Melbourne Institute of Applied Economic and Social Research, Melbourne, Vic., Australia

4.2. Exclusion restrictions for model identification

Although the model is formally identified by its non-linear functional form, as long as the full rank condition of the data matrix is ensured (Heckman, 1978; Wilde, 2000), we introduce exclusion restrictions to aid identification of the causal parameter β_{PIP} (Maddala, 1983; Monfardini and Radice, 2008). The row vector I_{ij} captures the variables included in the PIP participation Equation (5), but excluded from the outcome Equation (4).

Angrist and Pischke, (4.6.16), p. 201

$$\mathrm{E}[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} \mid \mathrm{D} = 1]$$

$$= \mathrm{E}\left\{ \frac{\Phi_b\left(\frac{x_i'\beta_0^* + \beta_1^*}{\sigma_\varepsilon}, \frac{x_i'\gamma_0^* + \gamma_1^* z_i}{\sigma_\varepsilon}; \rho_{\varepsilon v} \right) - \Phi_b\left(\frac{x_i'\beta_0^*}{\sigma_\varepsilon}, \frac{x_i'\gamma_0^* + \gamma_1^* z_i}{\sigma_\varepsilon}; \rho_{\varepsilon v} \right)}{\Phi\left(\frac{x_i'\gamma_0^* + \gamma_1^* z_i}{\sigma_\varepsilon} \right)} \right\}$$

FIML Estimates

```
-----------------------------------------------------------------------------
FIML - Recursive Bivariate Probit Model
Dependent variable               PUBDOC
Log likelihood function    -25671.32339
Estimation based on N =  27326, K =  14
Inf.Cr.AIC  =  51370.6 AIC/N =    1.880
--------+--------------------------------------------------------------------
  PUBLIC|                 Standard            Prob.       95% Confidence
  DOCTOR|  Coefficient      Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Index    equation for PUBLIC.........................................
Constant|    3.55056***     .07446    47.68  .0000     3.40462   3.69650
     AGE|     .00067        .00115      .58  .5626     -.00159    .00293
    EDUC|    -.16835***     .00416   -40.48  .0000     -.17650   -.16020
 MARRIED|    -.00997        .02922     -.34  .7329     -.06724    .04729
  HHKIDS|    -.08094***     .02510    -3.22  .0013     -.13014   -.03174
  INCOME|    -.98735***     .05172   -19.09  .0000    -1.08872   -.88598
  FEMALE|     .12140***     .02231     5.44  .0000      .07768    .16512
        |Index    equation for DOCTOR.........................................
Constant|     .58983***     .14474     4.08  .0000      .30615    .87351
     AGE|    -.05740***     .00601    -9.56  .0000     -.06917   -.04563
   AGESQ|     .00082***   .6817D-04   12.10  .0000      .00069    .00096
  INCOME|     .08900*       .05097     1.75  .0808     -.01091    .18890
  FEMALE|     .34580***     .01629    21.22  .0000      .31386    .37773
  PUBLIC|     .43595***     .07358     5.92  .0000      .29174    .58016
        |Disturbance correlation..............................................
RHO(1,2)|    -.17317***     .04075    -4.25  .0000     -.25303   -.09330
--------+--------------------------------------------------------------------
```

Treatment

Outcome

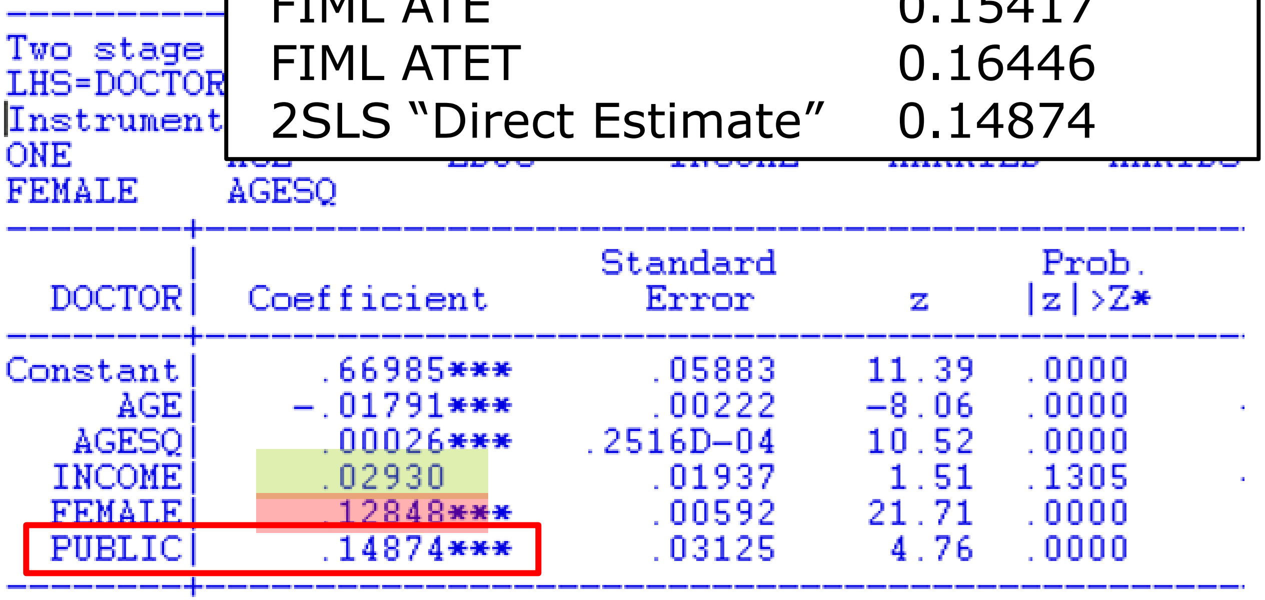
z

Treatment Effects

```
----------------------------------------------------------------------
Partial Effects  Analysis for RcrsvBvProb: Effect of PUBLIC on DOCTOR
----------------------------------------------------------------------
df/dPUBLIC             Partial    Standard
(Delta Method)          Effect     Error       |t|   95% Confidence Interval
----------------------------------------------------------------------
ATET Function           .16446     .02820     5.83      .10920        .21973
ATE  Function           .15417     .02482     6.21      .10553        .20282
```

FIML ATE 0.15417
FIML ATET 0.16446
2SLS “Direct Estimate” 0.14874

Two Stage Least Squares Effects



```
----------
Two stage
LHS=DOCTOR
Instrument
ONE
FEMALE     AGESQ
---------+---------------------------------------------------------
         |                      Standard                  Prob.
   DOCTOR|  Coefficient          Error          z        |z|>Z*
---------+---------------------------------------------------------
 Constant|      .66985***        .05883       11.39     .0000
      AGE|    -.01791***         .00222       -8.06     .0000
    AGESQ|      .00026***     .2516D-04       10.52     .0000
   INCOME|      .02930           .01937        1.51     .1305
   FEMALE|      .12848***        .00592       21.71     .0000
   PUBLIC|      .14874***        .03125        4.76     .0000
---------+---------------------------------------------------------
```

Partial Effects for IVProbit

Functions of Interest

$$\text{Prob}[y=1\,|\,\mathbf{x},T]=\Phi(\boldsymbol{\beta}'\mathbf{x}+\theta T)$$

$$\begin{aligned}\text{Prob}[y=1\,|\,\mathbf{x},\mathbf{z}] &= E_{T|\mathbf{z}}\text{Prob}[y=1\,|\,\mathbf{x},T]\\ &= \text{Prob}(T=0\,|\,\mathbf{z})\text{Prob}[y=1\,|\,\mathbf{x},T=0]+\text{Prob}(T=1\,|\,\mathbf{z})\text{Prob}[y=1\,|\,\mathbf{x},T=1]\\ &= \Phi(-\boldsymbol{\alpha}'\mathbf{z})\Phi(\boldsymbol{\beta}'\mathbf{x})+\Phi(\boldsymbol{\alpha}'\mathbf{z})\Phi(\boldsymbol{\beta}'\mathbf{x}+\theta)\end{aligned}$$

Partial Effects

Direct Effects - Primary Interest

$$\frac{\partial\text{Prob}[y=1\,|\,\mathbf{x},\mathbf{z}]}{\partial\mathbf{x}}=\left[\Phi(-\boldsymbol{\alpha}'\mathbf{z})\phi(\boldsymbol{\beta}'\mathbf{x})+\Phi(\boldsymbol{\alpha}'\mathbf{z})\phi(\boldsymbol{\beta}'\mathbf{x}+\theta)\right]\boldsymbol{\beta}$$

Indirect Effects - Secondary Interest

$$\frac{\partial\text{Prob}[y=1\,|\,\mathbf{x},\mathbf{z}]}{\partial\mathbf{z}}=\phi(\boldsymbol{\alpha}'\mathbf{z})\left[\Phi(\boldsymbol{\beta}'\mathbf{x}+\theta)-\Phi(\boldsymbol{\beta}'\mathbf{x})\right]\boldsymbol{\alpha}$$

On Avoiding Strong (Heroic) Assumptions

By what construction is a distributional assumption “strong?”

For better or worse, I tend to divide the estimation literature into methods that attempt to control for unobserved confounders and methods that don't. What I find disturbing is the number of articles we see at […] that purport to explore causal relationships, but begin with "we assume that all potential confounders are observed in the data." One author recently acknowledged that there might be unobserved confounders in his model, but **that problem could be dealt with only by making strong assumptions**. That's certainly true, but what assumption could be stronger than "all potential confounders are observed in the data?"