

Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model

William C. Horrace ^{a,*}, Ronald L. Oaxaca ^b

^a Department of Economics, Syracuse University, Syracuse, NY 13244, USA and NBER

^b Department of Economics, University of Arizona, Tucson, AZ 85721, USA and IZA

June, 2005

Abstract

This note formalizes bias and inconsistency results for ordinary least squares (OLS) on the linear probability model and provides sufficient conditions for unbiasedness and consistency to hold. The conditions suggest that a "trimming estimator" may reduce OLS bias.

Key Words: Consistency, Unbiased, LPM, OLS

JEL classification: C25

* Tel: 1-315-443-9016, fax: 1-315-443-1081, email: whorrace@maxwell.syr.edu.

1. Introduction

Limitations of the Linear Probability Model (LPM) are well-known. OLS estimated probabilities are not bounded on the unit interval, and OLS estimation implies that heteroscedasticity exists. Conventional advice points to probit or logit as the standard remedy, which bound the maximum likelihood estimated probabilities on the unit interval. However, the fact that consistent estimation of the LPM may be difficult does not imply that either probit or logit is the correct specification of the probability model; it may be reasonable to assume that probabilities are generated from bounded linear decision rules. Theoretical rationalizations for the LPM are in Rosenthal (1989) and Heckman and Snyder (1997).

Despite the attractiveness of logit and probit for estimating binary dependent variable models, OLS on the LPM is still used. Recent applications include Klaassen and Magnus (2001), Bettis and Farlie (2001), Lukashin (2000), McGarry (2000), Fairlie and Sundstrom (1999), Reiley (2005), and Currie and Gruber (1996). Empirical rationales for the LPM specification are plentiful. McGarry appeals to ease of interpretation of estimated marginal effects, while Reiley cites a perfect correlation problem associated with the probit model. Fairlie and Sundstrom prefer LPM because it implies a simple expression for the change in unemployment rate between two censuses. Bettis and Farlie choose LPM because of an extremely large sample size and other simplifications implied by it. Lukashin uses the LPM, because it lends itself to a model selection algorithm based on an adaptive gradient criterion. Currie and Gruber state that logit, probit, and OLS are similar for their data and only report LPM results.

Other rationales for the OLS on the LPM are complications of probit/logit models in certain contexts. Klaassen and Magnus cite panel data complications in their tennis example and select OLS. OLS is perhaps justified in simultaneous equations/instrumental variable methods. The presence of dummy endogenous regressors

is problematic if the DGP is assumed to be probit or logit; these problems were first considered by Heckman (1978). While perhaps less popular than logit and probit, OLS on the LPM model still finds its way into the literature for various reasons.

Some well-known LPM theorems are provided in Amemiya (1977). Econometrics textbooks (e.g., Greene, 2000), acknowledge complications leading to biased and inconsistent OLS estimates. Nevertheless, the literature is not clear on the precise conditions when OLS is problematic. This note rigorously lays out these conditions, derives the finite-sample and asymptotic biases of OLS, and provides additional results that highlight the appropriateness or inappropriateness of OLS estimation of the LPM. Finally, we suggest a trimmed sample estimator that could reduce OLS bias.

2. Results

Let y_i be a discrete random variable, taking on the values 0 or 1. Let x_i be a $1 \times k$ vector of explanatory variables on \mathfrak{R}^k , β be a $k \times 1$ vector of coefficients, and ε_i be a random error. Define probabilities over the random variable $x_i\beta \in \mathfrak{R}$.

$$\Pr(x_i\beta > 1) = \pi,$$

$$\Pr(x_i\beta \in [0, 1]) = \gamma,$$

$$\Pr(x_i\beta < 0) = \rho,$$

where $\pi + \gamma + \rho = 1$. Consider a random sample of data: $(y_i, x_i); i \in N; N = \{1, \dots, n\}$.

Define the data partition:

$$\kappa_\gamma = \{i \mid x_i\beta \in [0, 1]\}, \tag{1}$$

$$\kappa_\pi = \{i \mid x_i\beta > 1\},$$

implying

$$\begin{aligned}
\Pr(i \in \kappa_\pi) &= \pi, \\
\Pr(i \in \kappa_\gamma) &= \gamma, \\
\Pr(i \notin \kappa_\gamma \cup \kappa_\pi) &= \rho.
\end{aligned} \tag{2}$$

The LPM DGP is:

$$\begin{aligned}
y_i &= 1 \text{ for } i \in \kappa_\pi, \\
&= x_i\beta + \varepsilon_i \text{ for } i \in \kappa_\gamma, \\
&= 0 \text{ otherwise.}
\end{aligned} \tag{3}$$

The conditional probability of y_i is:

$$\begin{aligned}
\Pr(y_i = 1 \mid x_i, i \in \kappa_\pi) &= 1, \\
\Pr(y_i = 1 \mid x_i, i \in \kappa_\gamma) &= x_i\beta, \\
\Pr(y_i = 0 \mid x_i, i \in \kappa_\gamma) &= 1 - x_i\beta, \\
\Pr(y_i = 0 \mid x_i, i \notin \kappa_\gamma \cup \kappa_\pi) &= 1.
\end{aligned} \tag{4}$$

Therefore, y_i traces the familiar ramp function on $x_i\beta$ with error process:

$$\begin{aligned}
\varepsilon_i &= 0 \text{ for } i \in \kappa_\pi, \\
&= y_i - x_i\beta, \text{ } i \in \kappa_\gamma, \\
&= 0 \text{ for } i \notin \kappa_\gamma \cup \kappa_\pi,
\end{aligned}$$

and probabilities

$$\begin{aligned}
\Pr(\varepsilon_i = 0 \mid x_i, i \in \kappa_\pi) &= 1, \\
\Pr(\varepsilon_i = 1 - x_i\beta \mid x_i, i \in \kappa_\gamma) &= x_i\beta, \\
\Pr(\varepsilon_i = -x_i\beta \mid x_i, i \in \kappa_\gamma) &= 1 - x_i\beta, \\
\Pr(\varepsilon_i = 0 \mid x_i, i \notin \kappa_\gamma \cup \kappa_\pi) &= 1.
\end{aligned} \tag{5}$$

OLS proceeds as:

$$y_i = x_i\beta + u_i, \quad i \in N,$$

where u_i is a zero-mean random variable, independent of the x_i . Notice that the OLS error term, u_i , differs from ε_i :

$$\begin{aligned} u_i &= 1 - x_i\beta \text{ for } i \in \kappa_\pi, \\ &= y_i - x_i\beta \text{ for } i \in \kappa_\gamma, \\ &= -x_i\beta \text{ for } i \notin \kappa_\gamma \cup \kappa_\pi, \end{aligned}$$

with probability function:

$$\begin{aligned} \Pr(u_i = 1 - x_i\beta \mid x_i, i \in \kappa_\pi) &= 1, \\ \Pr(u_i = 1 - x_i\beta \mid x_i, i \in \kappa_\gamma) &= x_i\beta, \\ \Pr(u_i = -x_i\beta \mid x_i, i \in \kappa_\gamma) &= 1 - x_i\beta, \\ \Pr(u_i = -x_i\beta \mid x_i, i \notin \kappa_\gamma \cup \kappa_\pi) &= 1. \end{aligned} \tag{6}$$

The distinction between u_i and ε_i induces problems in OLS.

Theorem 1 *If $\gamma < 1$, then Ordinary Least Squares estimation of the Linear Probability Model is generally biased and inconsistent.*

Proof. Equation (6) implies:

$$\begin{aligned} E(u_i \mid x_i, i \in \kappa_\pi) &= 1 - x_i\beta, \\ E(u_i \mid x_i, i \in \kappa_\gamma) &= 0, \\ E(u_i \mid x_i, i \notin \kappa_\gamma \cup \kappa_\pi) &= -x_i\beta. \end{aligned}$$

Therefore, the conditional expectation of the OLS error, u_i , is a function of x_i with probability $(1 - \gamma)$. Hence, OLS is biased and inconsistent, if $\gamma < 1$. ■

Hence, only observations $i \in \kappa_\gamma$ possess mean-zero errors, so OLS with $i \notin \kappa_\gamma$ is problematic.

Remark 2 If $\kappa_\gamma \neq N$, then OLS estimation is biased and inconsistent. That is, if the sample used to estimate β contains any $i \notin \kappa_\gamma$, then $\gamma < 1$, so OLS is problematic.

Also:

Remark 3 If $\gamma = 1$, then OLS is unbiased and consistent, because $\pi = \rho = 0$, $E(u_i | x_i) = 0$ for all $i \in N$, and:

$$E(y_i | x_i) = \Pr(y_i = 1 | x_i) = x_i\beta, \quad i \in N.$$

Define random variables z_i and w_i :

$$z_i = 1 \text{ for } i \in \kappa_\gamma,$$

$$= 0 \text{ otherwise.}$$

$$w_i = 1 \text{ for } i \in \kappa_\pi,$$

$$= 0 \text{ otherwise.}$$

Hence, $\Pr(z_i = 1) = \gamma$ and $\Pr(w_i = 1) = \pi$. Alternative representation of equation (3) is:

$$y_i = w_i + z_i x_i \beta + u_i z_i; \quad i \in N, \tag{7}$$

making explicit that u_i is not the correct OLS error. Notice,

$$u_i z_i = 0 \text{ for } i \notin \kappa_\gamma,$$

$$= 1 - x_i \beta \text{ for } y_i = 1, \quad i \in \kappa_\gamma,$$

$$= -x_i \beta \text{ for } y_i = 0, \quad i \in \kappa_\gamma,$$

so the conditional probability function of $u_i z_i$ is the same as that of ε_i . Therefore, $E(u_i z_i | x_i) = 0$, and equation (7) has a zero-mean error, independent of x_i . Taking

the unconditional mean of equation (7):

$$\begin{aligned}
E(y_i) &= \pi + E(z_i x_i) \beta + E(u_i z_i) \\
&= \pi + \gamma E(z_i x_i \mid z_i = 1) \beta + \gamma E(z_i u_i \mid z_i = 1) \\
&= \pi + \gamma \mu_{x\gamma} \beta,
\end{aligned} \tag{8}$$

where $\mu_{x\gamma} = E(x_i \mid z_i = 1)$. Equation (8) will be used in the sequel. The OLS estimator is:

$$\hat{\beta}_n = \left[\sum_{i \in N} x'_i x_i \right]^{-1} \sum_{i \in N} x'_i y_i.$$

Substituting equation (7):

$$\hat{\beta}_n = \left[\sum_{i \in N} x'_i x_i \right]^{-1} \sum_{i \in N} x'_i (w_i + z_i x_i \beta + u_i z_i). \tag{9}$$

Partitioning the data by κ_γ and κ_π , and taking into consideration z_i and w_i in each regime:

$$\begin{aligned}
\hat{\beta}_n &= \left[\sum_{i \in N} x'_i x_i \right]^{-1} \left[\sum_{i \notin \kappa_\gamma \cup \kappa_\pi} x'_i(0) + \sum_{i \in \kappa_\gamma} x'_i(x_i \beta + u_i) + \sum_{i \in \kappa_\pi} x'_i(1) \right] \\
&= \left[\sum_{i \in N} x'_i x_i \right]^{-1} \left[\sum_{i \in \kappa_\gamma} x'_i x_i \beta + \sum_{i \in \kappa_\gamma} x'_i u_i + \sum_{i \in \kappa_\pi} x'_i \right].
\end{aligned}$$

Hence:

$$\begin{aligned}
E(\hat{\beta}_n \mid x_i) &= \left[\sum_{i \in N} x'_i x_i \right]^{-1} \left[\sum_{i \in \kappa_\gamma} x'_i x_i \beta + \sum_{i \in \kappa_\gamma} x'_i E(u_i \mid x_i, i \in \kappa_\gamma) + \sum_{i \in \kappa_\pi} x'_i \right] \\
&= \left[\sum_{i \in N} x'_i x_i \right]^{-1} \left[\sum_{i \in \kappa_\gamma} x'_i x_i \beta + \sum_{i \in \kappa_\gamma} x'_i(0) + \sum_{i \in \kappa_\pi} x'_i \right] \\
&= \left[\sum_{i \in N} x'_i x_i \right]^{-1} \sum_{i \in \kappa_\gamma} x'_i x_i \beta + \left[\sum_{i \in N} x'_i x_i \right]^{-1} \sum_{i \in \kappa_\pi} x'_i \neq \beta,
\end{aligned} \tag{10}$$

which is generally biased and asymptotically biased, because $\gamma < 1$. When $\gamma = 1$, $\kappa_\gamma = N$, the first term on the RHS is β , the second term is 0, and $\widehat{\beta}_n$ is unbiased.

The inconsistency of $\widehat{\beta}_n$ follows in a similar fashion. Letting C denote the cardinality operator, define $n_\pi = C(\kappa_\pi)$, $n_\gamma = C(\kappa_\gamma)$ and $n_\rho = n - n_\pi - n_\gamma$. Let plim denote the probability limit operator as $n \rightarrow \infty$. Assume $\text{plim}[n^{-1} \sum_{i \in N} x'_i x_i] = Q$ and $\text{plim}[n_\gamma^{-1} \sum_{i \in \kappa_\gamma} x'_i x_i] = Q_\gamma$ where Q and Q_γ are finite, (non-singular) positive definite. Assume $\text{plim}[n_\pi^{-1} \sum_{i \in \kappa_\pi} x'_i] = \mu'_{x\pi}$, $\text{plim}[n^{-1} \sum_{i \in N} x'_i] = \mu'_x$ and $\text{plim}[n_\gamma^{-1} \sum_{i \in \kappa_\gamma} x'_i u_i] = 0$, where $\mu'_{x\pi}$ and μ'_x are finite vectors. Assume $\text{plim}[n^{-1} n_\pi] = \pi$ and $\text{plim}[n_\gamma n^{-1}] = \gamma$. Then:

$$\text{plim}(\widehat{\beta}_n) = Q^{-1} (Q_\gamma \beta \gamma + \pi \mu'_{x\pi}) \neq \beta.$$

Even if γ and π were known, $\widehat{\beta}_n$ could not be bias corrected, yet equation (8) seems to imply that if γ and π were known, an OLS regression of $(y_i - \pi)$ on (γx_i) might be unbiased. Define transformed OLS estimator:

$$\widehat{\beta}_n^* = \left[\sum_{i \in N} \gamma^2 x'_i x_i \right]^{-1} \sum_{i \in N} \gamma x'_i (y_i - \pi). \quad (11)$$

Theorem 4 $\widehat{\beta}_n^*$ is biased and inconsistent for β .

Proof. Equation (11) implies

$$\begin{aligned} \widehat{\beta}_n^* &= \frac{1}{\gamma} \left[\sum_{i \in N} x'_i x_i \right]^{-1} \sum_{i \in N} x'_i y_i - \frac{1}{\gamma} \left[\sum_{i \in N} x'_i x_i \right]^{-1} \sum_{i \in N} x'_i \pi \\ &= \frac{1}{\gamma} \widehat{\beta}_n - \frac{\pi}{\gamma} \left[\sum_{i \in N} x'_i x_i \right]^{-1} \sum_{i \in N} x'_i. \end{aligned}$$

Hence,

$$E(\widehat{\beta}_n^* | x_i) = \frac{1}{\gamma} E(\widehat{\beta}_n | x_i) - \frac{\pi}{\gamma} \left[\sum_{i \in N} x'_i x_i \right]^{-1} \sum_{i \in N} x'_i \neq \beta.$$

■

Thus, knowledge of π and γ does not ensure an unbiased OLS estimator of β , and the bias will persist asymptotically. Moreover, it does not facilitate consistent estimation. The problem with $\widehat{\beta}_n$ and $\widehat{\beta}_n^*$ is not that γ and π are unknown but that κ_γ is unknown. If we knew κ_γ , we could perform OLS only on observations $i \in \kappa_\gamma$. Therefore:

Remark 5 *Sufficient information for unbiased and consistent OLS estimation is knowledge of κ_γ .*

Also, if $\kappa_\gamma = N$, then:

$$\sum_{i \in \kappa_\gamma} x'_i x_i = \sum_{i \in N} x'_i x_i, \text{ and } \sum_{i \in \kappa_\pi} x'_i = 0.$$

Therefore, equation (10) becomes:

$$E(\widehat{\beta}_n | x_i) = \left[\sum_{i \in N} x'_i x_i \right]^{-1} \sum_{i \in N} x'_i x_i \beta + \left[\sum_{i \in N} x'_i x_i \right]^{-1} (0) = \beta,$$

unbiased for $\kappa_\gamma = N$. A similar argument can be made for consistency. If $\gamma = 1$, then $\kappa_\gamma = N$. Therefore:

Remark 6 *Without knowledge of κ_γ and κ_π , a sufficient condition for unbiased OLS when $\gamma < 1$ is $\kappa_\gamma = N$.*

$\kappa_\gamma = N$ is a weaker sufficient condition than $\gamma = 1$, but probably unlikely in large samples. For any given random sample, $\Pr[\kappa_\gamma = N] = \gamma^n$, so

$$\lim_{n \rightarrow \infty} \Pr[\kappa_\gamma \neq N] = \lim_{n \rightarrow \infty} (1 - \gamma^n) = 1.$$

Remark 7 *Without knowledge of κ_γ and κ_π , if $\gamma < 1$ and $\kappa_\gamma = N$, then as $n \rightarrow \infty$, $\kappa_\gamma \neq N$ with probability approaching 1, and $\widehat{\beta}_n$ is asymptotically biased and inconsistent.*

Therefore, as N grows, once the first observation $x_i\beta \notin [0, 1]$ appears, then $\kappa_\gamma \neq N$ and unbiasedness is lost. Oddly, the estimator $\widehat{\beta}_n$ could be reliable in small samples yet unreliable in large samples.

3. Conclusions

Although it is theoretically possible for OLS on the LPM to yield unbiased estimation, this generally would require fortuitous circumstances. Furthermore, consistency seems to be an exceedingly rare occurrence as one would have to accept extraordinary restrictions on the joint distribution of the regressors. Therefore, OLS is frequently a biased estimator and almost always an inconsistent estimator of the LPM. If we had knowledge of the sets κ_γ and κ_π , then a consistent estimate of β could be based on the sub-sample $i \in \kappa_\gamma$. This is tantamount to removing observations $i \notin \kappa_\gamma$, suggesting that trimming observations violating the rule $\widehat{y}_i = x_i\widehat{\beta}_n \in [0, 1]$ and re-estimating the OLS model (based on the trimmed sample) may reduce finite sample bias. This seems to hold in simulations, but formal proof of this result is left for future research.

Acknowledgements

We gratefully acknowledge valuable comments by Seung Ahn, Badi Baltagi, Gordon Dahl, Dan Houser, Price Fishback, Art Gonzalez, Shawn Kantor, Alan Ker, Paul Ruud and Peter Schmidt. Capable research assistance was provided by Nidhi Thakur.

REFERENCES

- Amemiya, T., 1977, Some theorems in the linear probability model, *International Economic Review*, 18, 645-650.
- Bettis, J.R. and R.W. Fairlie, 2001, Explaining ethnic, racial, and immigrant differences in private school attendance, *Journal of Urban Economics*, 50, 26-51.

- Currie, J. and J. Gruber, 1996, Health insurance eligibility, utilization of medical care, and child health, *Quarterly Journal of Economics*, 111, 431-466.
- Fairlie, R.W. and W.A. Sundstrom, 1999, The emergence, persistence, and recent widening of the racial unemployment gap, *Industrial and Labor Relations Review*, 52, 252-270.
- Greene, W.H., 2000, *Econometric analysis* (Upper Saddle River, NJ: Prentice-Hall).
- Heckman, J.J., 1978, Dummy endogenous variables in a simultaneous equation system, *Econometrica*, 46, 931-59.
- Heckman, J.J. and J.M. Snyder, Jr., 1977, Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators, *Rand Journal of Economics*, 28, S142-S189.
- Klaassen, F.J.G.M. and J.R. Magnus, 2001, Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model, *Journal of the American Statistical Association*, 96, 500-509.
- Lukashin, Y.P., 2000, Econometric analysis of managers' judgements on the determinants of the financial situation in Russia, *Economics of Planning*, 33, 85-101.
- McGarry, K. 2000, Testing parental altruism: Implications of a dynamic model," NBER Working Paper 7593.
- Reiley, D.H., 2005, Field experiments on the effects of reserve prices in auctions: More magic on the internet, mimeo, University of Arizona.
- Rosenthal, R.W., 1989, A bounded-rationality approach to the study of noncooperative games, 18, 273-292.