#F 201661

ON MODELLING TRUNCATION VIA THE LOGISTIC DISTRIBUTION, WITH AN APPLICATION TO LAND MARKETS

1991

by

Feng Xu, Ron C. Mittelhammer, and Paul W. Barkley

Postdoctoral Fellow of Agricultural Economics at University of Missouri, and Professors of Agricultural Economics at Washington State University, respectively.

Land - Prices

Selected paper submitted for the annual meeting of the Western Agricultural Economics Association
Portland, Oregon, July 7-10, 1991

AAEA 1991

#20166

ON MODELLING TRUNCATION VIA THE LOGISTIC DISTRIBUTION, WITH AN APPLICATION TO LAND MARKETS

Feng Xu, Ron. C. Mittelhammer, and Paul W. Barkley
Postdoctoral Fellow of Agricultural Economics at University of Missouri, and Professors of Agricultural Economics at Washington State University, respectively.

This study presents an empirical method of modelling truncation using the logistic distribution. A simple test for the significance of truncation is provided. The method is applied in estimating land market hedonic price functions. Results show that truncation is significant in half of the cases analyzed.

Introduction

In many applied econometric analyses in agricultural economics it is well-known, and well-ignored, that the dependent variables being modelled are nonnegative random variables. In particular, in models of the form

$$(1)\quad Y = g(X;\beta) + \mu,$$

if Y refers to price or quantity of a market good demanded or supplied, or if Y refers to the sale price of a parcel of land, as in the application contained in this paper, the constraint $Y\geq 0$ applies. Any distribution assumed for μ in the model that does not assign probability zero to the event $\mu<-g(X;\beta)$ is logically a model misspecification. Explicit or implicit justification used by researchers for avoiding the nonnegativity issue relies on the assumption that the probability of $\mu<-g(X;\beta)$ is negligible and that the specification bias introduced by the truncation of μ is therefore also negligible. The assumption is rarely, if ever, tested in empirical work despite its critical importance to the validity of the model specification used in a particular econometric analysis.

In this paper, a relatively straightforward procedure for modelling and testing the significance of the truncation effect is presented. Given the advent of powerful and readily available nonlinear parameter estimation techniques, application of the truncation procedure is straightforward enough to be considered routine on current-day microcomputers. The empirical results of the paper provide a warning that assuming away the truncation effect may not be as innocuous as most researchers would like to believe.

Truncation Via the Logistic Distribution

The distribution of the proposed procedure begins with the nonlinear model (1), where $g(X;\beta)$ is a differentiable function of the explanatory variable X (which of course could be linear), and β is a vector of unknown parameters to be estimated. The logistic distribution with mean zero and scale parameter τ is assumed for the disturbance term μ, as

$$(2) \quad \mu \sim \tau^{-1}\exp[-\mu/\tau]/(1+\exp[-\mu/\tau])^2.$$

Note that the logistic distribution is assumed here as a compromise between the popular assumption of normality for the distribution term and computational tractability and stability of the truncation model evolving from the logistic distribution. Amemiya (1985, p.269) notes that a major justification for using the logistic distribution is that "the logistic distribution function is similar to a normal distribution, but has a much simpler form." In fact, it would be difficult to distinguish between graphs of the standard normal and the standard logistic density functions, with the only notable difference between the densities being that the logistic density has slightly heavier tails (Amemiya, p.269).

Given the distribution (2) for μ, and given that the dependent variable Y is truncated at zero, i.e., $Y \geq 0$, it can be shown that the expected value of the truncated distribution for Y is given by[a]

$$(3) \quad E(Y|Y\geq 0) = \tau(1+\exp[-g(X;\beta)/\tau])\ln(1+\exp[g(X;\beta)/\tau])$$

where $\exp(z) = e^z$ and $\ln(\cdot)$ refers to the natural logarithmic function. Thus, a specification of the truncated version of model (1) in a form suitable for nonlinear least squares estimation of the parameter vector β is

$$(4) \quad Y = E(Y|Y\geq 0) + \varepsilon = H(X;\beta,\tau) + \varepsilon$$

where $E(\varepsilon)=0$ and $Var(\varepsilon)=\sigma^2$, say (the variance of ε is a parametric constant that is not functionally related to β, but its value is a rather complicated infinite sum of incomplete gamma functions — see Maddala (1983, p.369)).

It can be shown that the truncation model (4) subsumes the truncated model (1) as a special (limiting) case. In particular[b]

$$(5) \quad \lim_{\tau\to 0} E(Y|Y\geq 0) = \lim_{\tau\to 0} H(X;\beta,\tau) = g(X;\beta).$$

This nesting of the truncated model within the specification (4) allows a common Wald-test (effectively, an asymmetric one-side t-test) to be used to test the significance of the null hypothesis H_0: $\tau=0$ versus the alternative hypothesis H_a: $\tau>0$, thereby testing the statistical significance of the truncation effect. A significant Wald-statistic (one sided t-test) would provide statistical evidence in form of accounting for truncation in the specification of the model. The test statistic would be calculated in the usual way as $\hat{\tau}/(Var(\hat{\tau}))^{1/2}$, where $Var(\hat{\tau})$ is the

[a,b] Derivations can be obtained from the authors.

estimate of the asymptotic variance of $\hat{\tau}$ provided by the nonlinear least squares estimation procedure.

Straightforward differentiation of $H(X;\beta,\tau)$ with respect to the i^{th} explanatory variable, X_i, yields a functional relationship between $\partial H(X;\beta,\tau)/\partial X_i$ and $\partial g(X;\beta)/\partial X_i$. In particular, defining $D(g(X;\beta)/\tau) = 1-\exp[-g(X;\beta)/\tau]\ln(1+\exp[g(X;\beta)/\tau])$, it follows that

$$(6)\quad \partial H(X;\beta,\tau)/\partial X_i = D(g(X;\beta)/\tau)\partial g(X;\beta)/\partial X_i.$$

Regarding the range of the function D, note that for $(g(X;\beta)/\tau) \in [0,\infty)$, $D(g(X;\beta)/\tau)$ is a monotonically increasing function such that $D(g(X;\beta)/\tau) \in [.3069,1]$. Then the closer $g(X;\beta)/\tau$ is to zero, the larger $|\partial g(X;\beta)/\partial X_i|$ is relative to $|\partial H(X;\beta,\tau)/\partial X_i|$; and as $g(X;\beta)/\tau \to \infty$, $|\partial g(X;\beta)/\partial X_i| \to |\partial H(X;\beta,\tau)/\partial X_i|$. This is in accordance with the fact that $P(Y\geq 0) = (1+\exp[-g(X;\beta)/\tau])^{-1}$ is monotonically increasing in $g(X;\beta)/\tau$, so that as $P(Y\geq 0) \to 1$, and thus as $g(X;\beta)/\tau \to \infty$, the truncation effect eventually vanishes.

The preceding discussion indicates that for the truncation effect to be negligible, it must be the case that the value of $g(X;\beta)$ relative to the scale parameter τ is sufficiently large so that $D(g(X;\beta)/\tau) \approx 1$ for all values of the explanatory variables X in the data being analyzed. The assumption would appear to be quite stringent, especially since at the outset of the analysis, neither $g(X;\beta)$ nor τ are known to the researchers. A prudent research strategy would be to test the significance of the truncation effect, as we illustrate in the next section.

An Application to Land Market Hedonic Price Function

The data were gathered using a telephone survey of all buyers of agricultural land in 25 rural Washington counties who purchased a land parcel of 10 or more acres in the years 1980 through 1987. The 25 counties were aggregated into six reasonably homogeneous regions with respect to types of agricultural production. The resulting number of useable observations were 137, 120, 81, 224, 184 and 159 for regions one through six, respectively. Variable definition are provided in Table 1.

A hedonic model of per acre sale price (SALEPR) was specified for each of the six regions as follows:

$$(7)\quad g(X;\beta) = TOTACRES^{\alpha}e^{\delta TIME}\,[\,\Sigma_{k=1}^{K} C_iCNTY_i + B_0 + B_1GI + B_2DTOWN + B_3\ln(LNDCAP)$$
$$+ B_4WDBK + \Sigma_i B_5SIZEB_i\ln(B_6 - AGEB_i) + B_7NUMP(B_8 - AGEP) + B_9IRRICP + B_{10}IRRISP$$
$$+ B_{11}IRRIR + B_{12}PASTURE\,] + B_{13}SIZEH\ln(B_{14} - AGEH) + B_{15}MACH,$$

with the functional form, incorporating the truncation effect, being,

$$(8)\quad SALEPR = h(X;\beta,\tau) = \tau(1+\exp[-g(X;\beta)/\tau])\ln(1+\exp[g(X;\beta)/\tau]) + \varepsilon,$$

where $\beta = \{\alpha, \delta, B_0, B_1, B_2, \ldots, B_{15}, C_i\text{'s}\}$ and τ are parameters to be estimated. The final model for each region contains different sets of explanatory variables, and is presented in general algebraic form below in (9) to (14) for regions one through six, respectively.

$$(9)\quad SALEPR = TOTACRES^{\alpha} e^{\delta TIME} [B_0 + C_1CNTY_1 + B_1GI + B_2DTOWN + B_3\ln(LNDCAP) + \Sigma_{i=1,5} B_5SIZEB_i\ln(B_6 - AGEB_i)] + B_{13}SIZEH\ln(B_{14} - AGEH)$$

$$(10)\quad SALEPR = TOTACRES^{\alpha} e^{\delta TIME} [B_0 + B_1GI + B_2DTOWN + B_3\ln(LNDCAP) + \Sigma_{i=1,4} B_5SIZEB_i + B_7 NUMP(B_8 - AGEP) + B_{10}IRRISP + B_{12}PASTURE] + B_{13}SIZEH\ln(B_{14} - AGEH)$$

$$(11)\quad SALEPR = TOTACRES^{\alpha} e^{\delta TIME} [B_0 + C_1CNTY_1 + B_1GI + B_2DTOWN + B_3\ln(LNDCAP) + \Sigma_i B_5SIZEB_i + B_{10}IRRISP] + B_{13}SIZEH\ln(B_{14} - AGEH) + B_{15}MACH$$

$$(12)\quad SALEPR = TOTACRES^{\alpha} e^{\delta TIME} [B_0 + B_1GI + B_2DTOWN + B_3\ln(LNDCAP) + B_4WDBK + \Sigma_{i=1,5} B_5SIZEB_i\ln(B_6 - AGEB_i) + B_{10}IRRISP + B_{11}IRRIR] + B_{13}SIZEH\ln(B_{14} - AGEH) + B_{15}MACH,$$

$$(13)\quad SALEPR = TOTACRES^{\alpha} e^{\delta TIME} [B_0 + B_1GI + B_2DTOWN + B_3\ln(LNDCAP) + B_4WDBK + \Sigma_{i=1,3} B_5SIZEB_i\ln(B_6 - AGEB_i) + B_9IRRICP + B_{10}IRRISP + B_{11}IRRIR] + B_{13}SIZEH\ln(B_{14} - AGEH) + B_{15}MACH$$

$$(14)\quad SALEPR = TOTACRES^{\alpha} e^{\delta TIME} [B_0 + C_1CNTY_1 + C_2CNTY_2 + C_3CNTY_3 + B_1GI + B_2DTOWN + B_3\ln(LNDCAP) + \Sigma_{i=1,5} B_5SIZEB_i\ln(B_6 - AGEB_i)] + B_{13}SIZEH\ln(B_{14} - AGEH) + B_{15}MACH$$

Models (7) and (8) were estimated using the SYSNLIN procedure in SAS/ETS (Statistical Analysis System / Econometric and Time Series) package. The results of the nonlinear least squares estimation for each of the six regions are given in Table 2.

The null hypothesis that $\tau=0$ (i.e., the truncation effect is significant) was rejected at the .06 level of type I error in half of the regions analyzed, and was insignificant at the .10 level in the remaining regions. Regarding the practical significance of the truncation effect in the three regions where the effect was statistically significant, note that the value of $D(g(X;\hat{\beta})/\hat{\tau})$ was .9992, .9997 and .9372 evaluated at the mean level of predicted sale prices ($\widehat{SALEPR}$) for each region, and .9690, .9788 and .8022 at .5$\widehat{SALEPR}$. The effect became substantially more pronounced the lower the predicted sale price.

Concluding Comments

Given the prevalence of nonnegatively constrained dependent variables in the econometric models of various aspects of the agricultural economy, and given the increasingly available computational ability to perform nonlinear parameter estimation, the time may be right for a more systematic evaluation of the need for explicitly incorporating the effects of the nonnegativity constraint in the specification of model structure. One method of explicitly modelling the effect of truncation is provided by the truncated logistic distribution approach presented in this paper. The approach is straightforward to implement using a nonlinear least squares algorithm, and allows a direct test of the significance of the truncation effect.

In a seeming routine hedonic analysis of land values, the routine assumption that the truncation effect induced by the nonnegativity of land values could be ignored in the specification of the model was rejected in half of the cases analyzed. One wonders how many other analyses have routinely dismissed the truncation effect without test.

TABLE 1. Definitions of Variables as Used in the Statistical Analysis

Variable	Definition and Source
SALEPR	Per acre sale price excluding payment for crops, but including the value of any machinery included in the transaction, divided by the CPI (original data)
TOTACRES	Parcel size in acres (original data)
TIME	Monthly time index, where January 1980=1, ..., and December 1987=96 (calculated)
GI	Gross income per acre, divided by the CPI (calculated)
PASTURE	Proportion of total acres that is pasture (original data)
$CNTY_i$	Dummy for county i, or for several counties in a region that are tested to have no county-specific difference in values (original data)
DTOWN	Distance to the nearest town with a service station and grocery store in miles (original data)
LNDCAP	Land capability class, ranging from 1 to 7, where 1 is the best-quality land, and 7 the poorest (calculated)
WDBK	Length of windbreak, in feet, per acre in parcel (original data)
NUMP	Number of stalls in the milking parlor, on a per acre basis (original data)
AGEP	Age of the milking parlor in years (original data)
IRRICP	Proportion of total acres irrigated by center pivot irrigation (original data)
IRRISP	Proportion of total acres irrigated by sprinkler irrigation (original data)
IRRIR	Proportion of total acres irrigated by rill irrigation (original data)
$SIZEB_i$	Size of barn i in square feet per acre in parcel (original data)
$AGEB_i$	Age of barn i in years (original data)
SIZEH	House size in square feet per acre in parcel (original data)
AGEH	House age in years (original data)
MACH	Assessed value of machinery per acre in parcel, divided by the CPI (original data)
C	A constant value (original data)

Source: The CPI refers to Western United States consumers who resided in areas having a population of less than 75,000. See United States Labor Statistics Bureau: CPI Deflated Report. The index value for 1980 is 1.00.

TABLE 2. The Estimated Parameters of Selected Models by Region

Para-meter	Charact-eristic	Region 1	2	3	4	5	6
		Coefficient (t-Value)					
α	TOTACRES	-.2056 (4.75)	-.1199 (3.54)	-.3881 (5.08)	-.0851 (2.72)	-.3494 (7.86)	-.1405 (4.68)
β	TIME	-.0114 (6.97)	-.0074 (5.33)	-.0156 (5.73)	-.0041 (4.48)	-.0146 (8.48)	-.0081 (5.45)
B_0	CONSTANT	8628.01 (5.64)	6431.59 (5.74)	38525.32 (2.97)	5024.02 (7.26)	9541.42 (4.68)	3646.12 (4.56)
C_1	$CNTY_1$	1848.01 (4.04)		18939.93 (2.48)			-584.54 (2.21)
C_2	$CNTY_2$						616.89 (2.10)
C_3	$CNTY_3$						1271.05 (4.12)
B_1	GI	1.2206 (1.68)	1.3591 (2.30)	3.0906 (2.96)	1.3299 (8.46)	10.9868 (4.68)	2.1315 (2.25)
B_2	DTOWN	-125.07 (2.01)	-133.38 (2.13)	-731.78 (1.99)	-74.0216 (2.01)	-483.15 (2.70)	-56.57 (3.55)
B_3	LNDCAP	-1560.16 (2.14)	-2075.44 (3.26)	-9848.49 (2.46)	-2338.98 (5.00)	-3967.81 (2.74)	-1185.24 (2.93)
B_4	WDBK				9.2574 (1.65)	173.97 (4.07)	
B_5	BARN	1.9145 (3.13)	2.6700 (2.52)	56.4500 (2.19)	3.1433 (6.69)	10.3576 (2.10)	8.7392 (1.35)
B_6	BARN	85.0238 (1749.46)			73.6780 (10.05)	55.0014 (6254.46)	55.0004 (22334.18)
B_7	PARLOR		217.12 (2.65)				
B_8	AGEP		26.2183 (8.56)				
B_9	IRRICP					5352.67 (2.61)	
B_{10}	IRRISP		1645.69 (2.35)	2701.07 (0.99)	681.07 (2.45)	4806.49 (3.22)	
B_{11}	IRRIR				528.05 (2.00)	4310.15 (3.10)	
B_{12}	PASTURE		1463.69 (3.62)				
B_{13}	HOUSE	4.3564 (15.31)	5.7513 (8.98)	3.8209 (3.28)	2.9466 (5.96)	4.3516 (6.05)	7.4103 (21.94)

B_{14}	HOUSE	95.6002 (36.67)	93.6590 (8.16)	75.5404 (140.58)	92.6361 (3.42)	79.1130 (2.74)	85.1139 (251.31)
B_{15}	MACH			1.7379 (3.57)	1.3098 (4.78)	1.2322 (3.95)	0.9122 (2.08)
n		137	120	81	225	184	159
R^2		.8367	.8045	.9409	.8673	.9039	.9015
Root MSE		692.76	710.41	775.09	721.40	449.15	301.65
Mean SALEPR		2612.37	3329.94	2444.91	2802.40	1518.73	915.08
τ		ns	ns	310.01 (2.51)	320.28 (1.56)	487.56 (9.30)	ns

Note: Values in parentheses are absolute t-values.
ns = not significant.
For regions one, four, five and six, barn variables are in the form $\Sigma_{i=1}^{I} B_5 SIZEB_i \ln(B_6 - AGEB_i)$, while for regions two and three, the variable is $\Sigma_{i=1}^{I} B_5 SIZEB_i$. County dummies are defined as follows: $CNTY_1$ is 1 if Grays Harbor, Lewis or Pacific, and 0 otherwise for region one; $CNTY_1$ is 1 if Stevens, and 0 otherwise for region three. For region six, $CNTY_1$ is 1 if Lincoln, and 0 otherwise; $CNTY_2$ is 1 if Walla Walla, and 0 otherwise; $CNTY_3$ is if Whitman, and 0 otherwise.

REFERENCES

Amemiya, Takeshi. Advanced Econometrics. Harvard University Press, Cambridge, Massachusetts. 1985.

Maddala, G.S. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press. 1983.

SAS/ETS User's Guide, Version 6. Cary, NC: SAS Institute Inc., 1988.

Xu, Feng. "An Econometric Study of Contributions of Parcel Characteristics to Agricultural Land Values in Washington: A Hedonic Approach." Unpublished Ph.D. Dissertation. Department of Agricultural Economics, Washington State University, Pullman. December 1990.