LIMITED DEPENDENT VARIABLES

PHOEBUS J. DHRYMES

Columbia University

Contents

0. Introduction 1568
1. Logit and probit 1568
   1.1. Generalities 1568
   1.2. Why a general linear model (GLM) formulation is inappropriate 1570
   1.3. A utility maximization motivation 1572
   1.4. Maximum likelihood estimation 1575
   1.5. Goodness of fit 1579
2. Truncated dependent variables 1585
   2.1. Generalities 1585
   2.2. Why simple OLS procedures fail 1586
   2.3. Estimation of parameters by ML methods 1589
   2.4. An initial consistent estimator 1590
   2.5. Limiting properties and distribution of the ML estimator 1595
   2.6. Goodness of fit 1603
3. Sample selectivity 1604
   3.1. Generalities 1604
   3.2. Inconsistency of least squares procedures 1606
   3.3. The LF and ML estimation 1610
   3.4. An initial consistent estimator 1613
   3.5. Limiting distribution of the ML estimator 1619
   3.6. A test for selectivity bias 1625
References 1626

Handbook of Econometrics, Volume III, Edited by Z. Griliches and M.D. Intriligator
© Elsevier Science Publishers BV, 1986

0. Introduction

This is intended to be an account of certain salient themes of the Limited Dependent Variable (LDV) literature. The object will be to acquaint the reader with the nature of the basic problems and the major results rather than recount just who did what when. An extended bibliography is given at the end, that attempts to list as many papers as have come to my attention – even if only by title.

By LDV we will mean instances of (dependent) variables – i.e. variables to be explained in terms of some economic model or rationalizing scheme for which (a) their range is intrinsically a finite discrete set and any attempt to extend it to the real line (or the appropriate multivariable generalization) not only does not lead to useful simplification, but befouls any attempt to resolve the issues at hand; (b) even though their range may be the real (half) line (or the appropriate multivariable generalization) their behavior is conditioned on another process(es).

Examples of the first type are models of occupational choice, entry into labor force, entry into college upon high school graduation, utilization of recreational facilities, utilization of modes of transport, childbearing, etc.

Examples of the latter are models of housing prices and wages in terms of the relevant characteristics of the housing unit or the individual – what is commonly referred to as hedonic price determination. Under this category we will also consider the case of truncated dependent observations.

In examining these issues we shall make an attempt to provide an economic rationalization for the model considered, but our main objective will be to show why common procedures such as least squares fail to give acceptable results; how one approaches these problems by maximum likelihood procedures and how one can handle problems of inference – chiefly by determining the limiting distributions of the relevant estimators. An attempt will be made to handle all problems in a reasonably uniform manner and by relatively elementary means.

1. Logit and probit

1.1. Generalities

Consider first the problem faced by a youth completing high school; or by a married female who has attained the desired size of her family. In the instance of the former the choice to be modelled is going to college or not; in the case of the latter we need to model the choice of entering the labor force or not.

Suppose that as a result of a properly conducted survey we have observations on T individuals, concerning their socioeconomic characteristics and the choices they have made.

In order to free ourselves from dependence on the terminology of a particular subject when discussing these problems, let us note that, in either case, we are dealing with binary choice; let us denote this by

Alternative 1 Going to College or Entering Labor Force

Alternative 2 Not Going to College or Not Entering Labor Force

Since the two alternatives are exhaustive we may make alternative 1 correspond to an abstract event $\mathscr{E}$ and alternative 2 correspond to its complement $\bar{\mathscr{E}}$. In this context it will be correct to say that what we are interested in is the set of factors affecting the occurrence or nonoccurrence of $\mathscr{E}$. What we have at our disposal is some information about the *attributes of these alternatives* and the (*socioeconomic*) *attributes of the individual exercising choice*. Of course we also observe the choices of the individual agent in question. Let

$$\begin{aligned} y_t &= 1 \quad \text{if individual } t \text{ chooses in accordance with event } \mathscr{E}, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Let

$$w = (w_1, w_2, \ldots, w_s),$$

be a vector of characteristics relative to the alternatives corresponding to the events $\mathscr{E}$ and $\bar{\mathscr{E}}$; finally, let

$$r_{t\cdot} = (r_{t1}, \ldots, r_{tm}),$$

be the vector describing the socioeconomic characteristics of the tth individual economic agent.

We may be tempted to model this phenomenon as

$$y_t = x_{t\cdot}\beta + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{1}$$

where

$$x_{t\cdot} = (w, r_{t\cdot}).$$

β is a vector of unknown constants and

$$\varepsilon_t : t = 1, 2, \ldots, T,$$

is a sequence of suitably defined error terms.

The formulation in (1) and subsequent estimation by least squares procedures was a common occurrence in the empirical research of the sixties.

1.2. *Why a general linear model (GLM) formulation is inappropriate*

Although the temptation to think of LDV problems in a GLM context is enormous a close examination will show that this is also fraught with considerable problems. At an intuitive level, we seek to approximate the dependent variable by a linear function of some other observables; the notion of approximation is based on ordinary Euclidean distance. That is quite sensible, in the usual GLM context, since no appreciable violence is done to the essence of the problem by thinking of the dependent variable as ranging without restriction over the real line – perhaps after suitably centering it first.

Since the linear function by which we approximate it is similarly unconstrained, it is not unreasonable to think of Euclidean distance as a suitable measure of proximity. Given these considerations we proceed to construct a logically consistent framework in which we can optimally apply various inferential procedures.

In the present context, however, it is not clear whether the notion of Euclidean distance makes a great deal of sense as a proximity measure. Notice that the dependent variable can only assume two possible values, while no comparable restrictions are placed on the first component of the right hand side of (1). Second, note that if we insist on putting this phenomenon in the GLM mold, then for observations in which

$$y_t = 1,$$

we must have

$$\varepsilon_t = 1 - x_{t\cdot}\beta, \tag{2}$$

while for observations in which

$$y_t = 0,$$

we must have

$$\varepsilon_t = -x_{t\cdot}\beta. \tag{3}$$

Thus, the error term can only assume two possible values, and we are immediately led to consider an issue that is important to the proper conceptualization of such models, viz., that what we need is *not* a linear model "explaining" the choices individuals make, but rather a model of the probabilities corresponding to the choices in question. Thus, if we ask ourselves: what is the expectation of ε_t, we shall be forced to think of the probabilities attaching to the relations described in (2) and (3) and thus conclude that

$$\varepsilon_t = 1 - x_{t\cdot}\beta,$$

with probability equal to

$$p_{t1} = P(y_t = 1), \tag{4}$$

and

$$\varepsilon_t = -x_{t\cdot}\beta,$$

with probability

$$p_{t2} = P(y_t = 0) = 1 - p_{t1}. \tag{5}$$

What we really should be asking is: what determines the probability that the tth economic agent chooses in accordance with event $\mathscr{E}$, and eq. (1) should be viewed as a clumsy way of going about it. We see that putting

$$p_{t1} = F(x_{t\cdot}\beta) = \int_{-\infty}^{x_{t\cdot}\beta} f(\xi)\,\mathrm{d}\xi, \tag{6}$$

$$p_{t2} = 1 - F(x_{t\cdot}\beta) = \int_{x_{t\cdot}\beta}^{\infty} f(\xi)\,\mathrm{d}\xi, \tag{7}$$

where $f(\cdot)$ is a suitable density function with known parameters, formalizes the dependence of the probabilities of choice on the observable characteristics of the individual and/or the alternatives.

To complete the argumentation about why the GLM is inapplicable in the present context we note further

$$E(\varepsilon_t) = F(x_{t\cdot}\beta)(1 - x_{t\cdot}\beta) + [1 - F(x_{t\cdot}\beta)](-x_{t\cdot}\beta) = F(x_{t\cdot}\beta) - x_{t\cdot}\beta, \tag{8}$$

$$\mathrm{Var}(\varepsilon_t) = F(x_{t\cdot}\beta)[1 - F(x_{t\cdot}\beta)]. \tag{9}$$

Hence, prima facie, least squares techniques are not appropriate, even if the formulations in (1) made intuitive sense.

We shall see that similar situations arise in other LDV contexts in which the absurdity of least squares procedures is not as evident as it is here.

Thus, to recapitulate, least squares procedures are inapplicable

i. because we should be interested in estimating the probability of choice; however, we are using a linear function to predict actual choices, without ensuring that the procedure will yield "predictions" satisfying the conditions that probabilities ought to satisfy
ii. on a technical level the conditions on the error term that are compatible with the desirable properties of least squares estimators in the context of the GLM are patently false in the present case.

1.3. *A utility maximization motivation*

As before, consider an individual, t, who is faced with the choice problem as in the preceding section but who is also hypothesized to behave so as to maximize his utility in choosing between the two alternatives. In the preceding it is assumed that the individual's utility contains a random component. It involves little loss in relevance to write the utility function as

$$U_t = u(w, r_{t\cdot}; \theta) + \varepsilon_t, \qquad t = 1,2,\dots,T,$$

where

$$u(w, r_{t\cdot}; \theta) \equiv E(U|w, r_{t\cdot}), \qquad \varepsilon_{t\cdot} \equiv U_t - u(w, r_{t\cdot}; \theta).$$

For the moment we shall dispense with the subscript t referring to the tth individual.

If the individual chooses according to event $\mathscr{E}$, his utility is (where now any subscripts refer to alternatives),

$$U_1 = u(w, r; \theta_1) + \varepsilon_1. \tag{10}$$

The justification for the parameter vector θ being subscripted is that, since w is constant across alternatives, θ must vary. While this may seem unnatural to the reader it is actually much more convenient, as the following development will make clear.

If the individual chooses in accordance with $\bar{\mathscr{E}}$, then

$$U_2 = u(w, r; \theta_2) + \varepsilon_2. \tag{11}$$

Hence, choice is in accordance with event $\mathscr{E}$ if, say,

$$U_1 \geq U_2. \tag{12}$$

But (12) implies

Alternative 1 is chosen or choice is made in accordance with event $\mathscr{E}$ if

$$\varepsilon_2 - \varepsilon_1 \leq u(w, r; \theta_1) - u(w, r; \theta_2), \tag{13}$$

which makes it abundantly clear that we can speak unambiguously *only* about the probabilities of choice. To "predict" choice we need an additional "rule" – such as, for example,

Alternative 1 is chosen when the probability attaching to event $\mathscr{E}$ is 0.5 or higher.

If the functions $u(\cdot)$ in (13) are linear, then the tth individual will choose Alternative 1 if

$$\varepsilon_{t2} - \varepsilon_{t1} \leq x_{t\cdot}\beta, \tag{14}$$

where

$$x_{t\cdot} = (w, r_{t\cdot}), \qquad \beta = \theta_1 - \theta_2. \tag{15}$$

Hence, in the notation of the previous section

$$P(y_t = 1) = P(\varepsilon_{t2} - \varepsilon_{t1} \leq x_{t\cdot}\beta) = \int_{-\infty}^{x_{t\cdot}\beta} f_t(\xi)\,\mathrm{d}\xi = F_t(x_{t\cdot}\beta), \tag{16}$$

where now f_t is the density function of $\varepsilon_{t2} - \varepsilon_{t1}$.

If

$$f_t(\cdot) = f_{t'}(\cdot), \qquad t \neq t',$$

then we have a basis for estimating the parametric structure of our model. Before we examine estimation issues, however, let us consider some possible distribution for the errors, i.e. the random variables ε_{t1}, ε_{t2}.

Thus, suppose

$$\varepsilon_{t'\cdot} \sim N(0, \Sigma), \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \qquad \Sigma > 0,$$

and the $\varepsilon_{t\cdot}$'s are independent identically distributed (i.i.d.). We easily find that

$$\varepsilon_{t2} - \varepsilon_{t1} \sim N(0, \sigma^2), \qquad \sigma^2 = \sigma_{22} - 2\sigma_{12} + \sigma_{11}.$$

Hence

$$\Pr\{y_t = 1\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x_{t\cdot}\beta} \mathrm{e}^{-1/2\sigma^2\xi^2}\,\mathrm{d}\xi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\nu_t} \mathrm{e}^{-(1/2)\zeta^2}\,\mathrm{d}\zeta = F(\nu_t),$$

where

$$\nu_t = \frac{x_{t\cdot}\beta}{\sqrt{\sigma^2}},$$

and $F(\cdot)$ is the c.d.f. of the unit normal. Notice that in this context it is not possible to identify separately β and σ^2 by observing solely the choices individuals make; we can only identify β/σ.

For reasons that we need not examine here, analysis based on the assumption that errors in (10) and (11) are normally distributed is called *Probit Analysis.*

We shall now examine another specification that is common in applied research, which is based on the logistic distribution. Thus, let q be an exponentially distributed random variable so that its density is

$$g(q) = \mathrm{e}^{-q} \qquad q \in (0,\infty), \tag{17}$$

and consider the distribution of

$$v = \ln(q)^{-1} = -\ln q. \tag{18}$$

The Jacobian of this transformation is

$$J(r \to q) = \mathrm{e}^{-v}.$$

Hence, the density of v is

$$h(v) = \exp - v \exp - \mathrm{e}^{-v} \qquad v \in (-\infty, \infty). \tag{19}$$

If the ε_{ti}, $i = 1,2$ of (14) are mutually independent with density as in (19), then the joint density is

$$h(\varepsilon_1, \varepsilon_2) = \exp - (\varepsilon_1 + \varepsilon_2)\exp - (\mathrm{e}^{-\varepsilon_1} + \mathrm{e}^{-\varepsilon_2}). \tag{20}$$

Put

$$\begin{aligned} v_1 + v_2 &= \varepsilon_2, \\ v_2 &= \varepsilon_1. \end{aligned} \tag{21}$$

The Jacobian of this transformation is 1; hence the joint density of the v_i, $i = 1,2$, is given by

$$\exp - (v_1 + 2v_2)\exp - (\mathrm{e}^{-v_2} + \mathrm{e}^{-v_1 - v_2}).$$

Since

$$v_1 = \varepsilon_2 - \varepsilon_1,$$

the desired density is found as

$$g(v_1) = \int_{-\infty}^{\infty} \exp -(v_1 + 2v_2) \exp -(\mathrm{e}^{-v_2} + \mathrm{e}^{-v_1 - v_2}) \mathrm{d}v_2.$$

To evaluate this put

$$1 + \mathrm{e}^{-v_1} = t, \qquad s = t\mathrm{e}^{-v_2},$$

to obtain

$$g(v_1) = \frac{\mathrm{e}^{-v_1}}{t^2} \int_0^{\infty} s\mathrm{e}^{-s} \mathrm{d}s = \frac{\mathrm{e}^{-v_1}}{(1 + \mathrm{e}^{-v_1})^2}.$$

Hence, in this case the probability of choosing Alternative 1 is given by

$$P(y_t = 1) = \int_{-\infty}^{x_{t\cdot}\beta} \frac{\mathrm{e}^{-\zeta}}{(1 - \mathrm{e}^{-\zeta})^2} \mathrm{d}\zeta = \frac{1}{1 + \mathrm{e}^{-\zeta}} \Bigg|_{-\infty}^{x_{t\cdot}\beta} = \frac{1}{1 + \mathrm{e}^{-x_{t\cdot}\beta}},$$

$$P(y_t = 0) = 1 - F(x_{t\cdot}\beta) = \frac{\mathrm{e}^{-x_{t\cdot}\beta}}{1 + \mathrm{e}^{-x_{t\cdot}\beta}}.$$

This framework of binary or *dichotomous* choice easily generalizes to the case of *polytomous* choice, without any appreciable complication – see, e.g. Dhrymes (1978a).

1.4. Maximum likelihood estimation

Although alternative estimation procedures are available we shall examine only the maximum likelihood (ML) estimator, which appears to be the most appropriate, given the sorts of data typically available to economists.

To recapitulate: we have the problem of estimating the parameters in a dichotomous choice context, characterized by a density function $f(\cdot)$; we shall deal with the case where $f(\cdot)$ is the *unit normal* and the *logistic*.

As before we define

$$\begin{aligned} y_t &= 1 \quad \text{if choice corresponds to event } \mathscr{E} \\ &= 0 \quad \text{if choice corresponds to event } \bar{\mathscr{E}} \end{aligned}$$

The event $\mathscr{E}$ may correspond to entering the labor force or going to college in the examples considered earlier.

$$P(y_t=1)=F(x_{t\cdot}\beta),$$

where

$$x_{t\cdot}=(w, r_{t\cdot}),$$

w is the s-element row vector describing the relevant attributes of the alternatives and $r_{t\cdot}$ is the m-element row vector describing the relevant socioeconomic characteristics of the tth individual.

We recall that a likelihood function may be viewed in two ways: for purposes of estimation we take the sample as given (here the y_t's and $x_{t\cdot}$'s) and regard it as a function of the unknown parameters (here the vector β) with respect to which it is to be maximized; for purposes of deriving the limiting distribution of estimators it is appropriate to think of it as a function of the dependent variable(s) – and hence as one that encompasses the probabilistic structure imposed on the model. This dual view of the likelihood function (LF) will become evident below.

The LF is easily determined to be

$$L^*=\prod_{t=1}^{T}F(x_{t\cdot}\beta)^{y_t}[1-F(x_{t\cdot}\beta)]^{1-y_t}. \tag{22}$$

As usual, we find it more convenient to operate with its logarithm

$$\ln L^*=L=\sum_{t=1}^{T}\{y_t\ln F(x_{t\cdot}\beta)+(1-y_t)\ln[1-F(x_{t\cdot}\beta)]\}. \tag{23}$$

For purposes of estimation, this form is unduly complicated by the presence of the random variables, y_t's. Given the sample, we will know that some of the y_t's assume the value one and others assume the value zero. We can certainly rearrange the observations so that the first $T_1\le T$ observations correspond to

$$y_t=1,\qquad t=1,2,\dots,T_1,$$

while the remaining $T_2<T$ correspond to

$$y_{T_1+t}=0,\qquad t=1,2,\dots,T_2,$$

If we give effect to these statements the log likelihood function becomes

$$L=\sum_{t=1}^{T_1}\ln F(x_{t\cdot}\beta)+\sum_{t=T_1+1}^{T_1+T_2}\ln[1-F(x_{t\cdot}\beta)], \tag{24}$$

and as such it does not contain any random variables[1] – *even symbolically*! Thus, it is rather easy for a beginning scholar to become confused as to how, solving

$$\frac{\partial L}{\partial \beta} = 0,$$

will yield an estimator, say $\hat{\beta}$, with any probabilistic properties. At least the analogous situation in the GLM

$$y = X\beta + u,$$

using the standard notation yields

$$\hat{\beta} = (X'X)^{-1}X'y,$$

and y is recognized to be a random variable with a probabilistic structure induced by our assumption on the structural error vector u.

Thus, we shall consistently avoid the use of the form in (24) and use instead the form in (23). As is well known, the ML estimator is found by solving

$$\frac{\partial L}{\partial \beta} = \sum_{t=1}^{T}\left[y_t \frac{f(x_{t\cdot}\beta)}{F(x_{t\cdot}\beta)} - (1-y_t)\frac{f(x_{t\cdot}\beta)}{1-F(x_{t\cdot}\beta)} \right] x_{t\cdot} = 0. \tag{25}$$

We note that, in general, (25) is a highly nonlinear function of the unknown parameter β and, hence, can only be solved by iteration.

Since by definition a ML estimator, $\hat{\beta}$, is one obeying

$$L(\hat{\beta}) \geq L(\beta), \text{ for all admissible } \beta, \tag{26}$$

it is important to ensure that solving (25) does, indeed, yield a maximum in the form of (26) and not merely a local stationary point – at least asymptotically.

The assumptions under which the properties of the ML estimator may be established are partly motivated by the reason stated above. These assumptions are

Assumption A.1.1.

The explanatory variables are uniformly bounded, i.e. $x_{t\cdot} \varepsilon H_*$, for all t, where H_* is a closed bounded subset of R_{s+m}, i.e. the $(s+m)$-dimensional Euclidean space.

Assumption A.1.2.

The (admissible) parameter space is, similarly, a closed bounded subset of R_{s+m}, say, P_* such that $P_* \supset N(\beta^0)$, where $N(\beta^0)$ is an open neighborhood of the true parameter point β^0.

[1]For any sample, of course, the choice of T_1 is random.

Remark 1

Assumption (A.1.1.) is rather innocuous and merely states that the socioeconomic variables of interest are bounded. Assumption (A.1.2.) is similarly innocuous. The technical import of these assumptions is to ensure that, at least asymptotically, the maximum maximorum of (24) is properly located by the calculus methods of (25) and to also ensure that the equations in (25) are well defined by precluding a singularity due to

$$F(x_{t\cdot}\beta)=0 \qquad \text{or} \qquad 1-F(x_{t\cdot}\beta)=0.$$

Moreover, these assumptions also play a role in the argument demonstrating the consistency of the ML estimator.

To the above we add another condition, well known in the context of the general linear model (GLM).

Assumption A.1.3.

Let

$$X=(x_{t\cdot}) \qquad t=1,2,\dots,T,$$

where the elements of $x_{t\cdot}$ are nonstochastic. Then

$$\operatorname{rank}(X)=s+m, \qquad \lim_{T\to\infty}\frac{X'X}{T}=M>0.$$

With the aid of these assumptions we can easily demonstrate (the proof will not be given here) the validity of the following

Theorem 1

Given assumption A.1.1. through A.1.3. the log likelihood function, L of (24), is concave in β, whether $F(\cdot)$ is the unit normal or the logistic c.d.f..

Remark 2

The practical implication of Theorem 1 is that, at any sample size, if we can satisfy ourselves that the LF of (24) does not attain its maximum on the boundary of the parameter space, then a solution to (25), say $\hat{\beta}$, obeys

$$L(\hat{\beta})\geq L(\beta) \qquad \text{for all admissible } \beta.$$

On the other hand as the sample size tends to infinity then with probability one the condition above is satisfied.

The (limiting) properties of the ML estimator necessary for carrying out tests of hypotheses are given in

Theorem 2

The ML estimator, $\hat{\beta}$, in the logistic as well as the normal case is strongly consistent and moreover it obeys

$$\sqrt{T}(\hat{\beta}-\beta) \sim N(0,C),$$

when

$$C^{-1} = -\lim_{T\to\infty} \frac{1}{T} E\left[\frac{\partial^2 L}{\partial\beta\partial\beta}(\beta^0)\right].$$

Corollary 1

A consistent estimator of the covariance matrix of the limiting distribution is given, in the case of normal density, f, and c.d.f., F, by

$$\hat{C} = \left\{\frac{1}{T}\sum\left[\frac{f^2(x_{t\cdot}\hat{\beta})}{F(x_{t\cdot}\hat{\beta})[1-F(x_{t\cdot}\hat{\beta})]}x'_{t\cdot}x_{t\cdot}\right]\right\}^{-1}, \tag{27}$$

For the logistic c.d.f. (logit) this reduces to

$$\hat{C} = \left[\frac{1}{T}\sum_{t=1}^{T} f(x_{t\cdot}\hat{\beta})x'_{t\cdot}x_{t\cdot}\right]^{-1}. \tag{28}$$

1.5. Goodness of fit

In the context of the GLM the coefficient of determination of multiple regression (R^2) has at least three useful interpretations.

i. it stands in a one-to-one relation to the F-statistic for testing the hypothesis that the coefficients of the *bona fide* explanatory variables are zero;
ii. it is a measure of the reduction of the variability of the dependent variable through the *bona fide* explanatory variables;
iii. it is the square of the simple correlation coefficient between predicted and actual values of the dependent variable within the sample.

Unfortunately, in the case of the discrete choice models under consideration we do not have a statistic that fits all three characterizations above. We can, on the other hand, define one that essentially performs the first two functions.

In order to demonstrate these facts it will be convenient to represent the maximized (log) LF more informatively. Assuming that the ML estimator corresponds to an interior point of the admissible parameter space we can write

$$L(\hat{\beta}) = L(\beta^0) + \frac{\partial L}{\partial \beta}(\beta^0)(\hat{\beta}-\beta^0) + \frac{1}{2}(\hat{\beta}-\beta^0)'\frac{\partial^2 L}{\partial \beta \partial \beta}(\beta^0)(\hat{\beta}-\beta^0) + \text{third order terms.} \tag{29}$$

The typical third order term involves

$$\phi_T = \frac{1}{6}\frac{1}{T^{3/2}}\frac{\partial^3 L}{\partial \beta_i \partial \beta_j \partial \beta_k}(\beta^*)\sqrt{T}(\hat{\beta}_i-\beta_i^0)\sqrt{T}(\hat{\beta}_j-\beta_j^0)\sqrt{T}(\hat{\beta}_k-\beta_k^0).$$

It is our contention that

$$\operatorname*{plim}_{T\to\infty} \phi_T = 0. \tag{30}$$

Now,

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\frac{\partial^3 L}{\partial \beta_i \partial \beta_j \partial \beta_k}(\beta^*) = \bar{L}_{ijk},$$

is a well defined, finite quantity, where

$$|\beta^* - \beta^0| \leq \hat{\beta} - \beta^0|.$$

But then, (30) is obvious since it can be readily shown that

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T^{3/2}}\frac{\partial^3 L}{\partial \beta_i \partial \beta_j \partial \beta_k} = 0,$$

and moreover that

$$\sqrt{T}(\hat{\beta}_i-\beta_i^0),\ \sqrt{T}(\hat{\beta}_j-\beta_j^0),\ \sqrt{T}(\hat{\beta}_k-\beta_k^0),$$

are a.c. finite. Hence, for large samples, approximately

$$L(\hat{\beta}) \sim L(\beta^0) + \frac{\partial L}{\partial \beta}(\beta^0)(\hat{\beta} - \beta^0) + \frac{1}{2}(\hat{\beta} - \beta^0)' \frac{\partial^2 L}{\partial \beta \partial \beta}(\beta^0)(\hat{\beta} - \beta^0).$$

On the other hand, expanding $\frac{\partial L}{\partial \beta}$ by Taylor series we find

$$\frac{1}{\sqrt{T}} \frac{\partial L}{\partial \beta}(\beta^0) \sim -\left[\frac{1}{T} \frac{\partial^2 L}{\partial \beta \partial \beta}(\beta^0)\right] \sqrt{T}(\hat{\beta} - \beta^0).$$

Thus,

$$\frac{\partial L}{\partial \beta}(\beta^0)(\hat{\beta} - \beta^0) \sim -(\hat{\beta} - \beta^0)' \frac{\partial^2 L}{\partial \beta \partial \beta}(\beta^0)(\hat{\beta} - \beta^0),$$

and, consequently, for large samples

$$L(\hat{\beta}) \sim L(\beta^0) - \frac{1}{2}(\hat{\beta} - \beta^0)' \frac{\partial^2 L}{\partial \beta \partial \beta}(\beta^0)(\hat{\beta} - \beta^0).$$

Hence

$$2[L(\hat{\beta}) - L(\beta^0)] \sim -(\hat{\beta} - \beta^0)' \frac{\partial^2 L}{\partial \beta \partial \beta}(\beta^0)(\hat{\beta} - \beta^0) \sim \chi^2_{s+m}. \tag{31}$$

Consider now the hypothesis

$$H_0: \qquad \beta^0 = 0, \tag{32}$$

as against

$$H_1: \qquad \beta^0 \neq 0.$$

Under H_0

$$L(\beta^0) = \sum_{t=1}^{T} \{ y_t \ln F(0) + (1 - y_t) \ln[1 - F(0)] \} = T \ln(\tfrac{1}{2}),$$

and

$$2[L(\hat{\beta}) - T \ln \tfrac{1}{2}] \sim \chi^2_{s+m},$$

is a test statistic for testing the null hypothesis in (32). On the other hand, this is not a useful basis for defining an R^2 statistic, for it implicitly juxtaposes the economically motivated model that defines the probability of choice as a function of

$$x_{t\cdot}\beta,$$

and the model based on the *principle of insufficient reason* which states that the probability to be assigned to choice corresponding to the event $\mathscr{E}$ and that corresponding to its complement $\bar{\mathscr{E}}$ are both $\frac{1}{2}$. It would be far more meaningful to consider the null hypothesis to be

$$\beta^0 = \begin{pmatrix} \beta_0^0 \\ 0 \end{pmatrix},$$

i.e. to follow for a nonzero constant term, much as we do in the case of the GLM. The null hypothesis as above would correspond to assigning a probability to choice corresponding to event $\mathscr{E}$ by

$$\bar{y} = F(\tilde{\beta}_0) \qquad \text{or} \qquad \tilde{\beta}_0 = F^{-1}(\bar{y}),$$

where

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t.$$

Thus, for some null hypothesis H_0, let

$$L(\tilde{\beta}) = \sup_{H_0} L(\beta).$$

By an argument analogous to that leading to (31) we conclude that

$$\begin{aligned} 2[L(\hat{\beta}) - L(\tilde{\beta})] \sim -(\hat{\beta}-\beta^0)'\frac{\partial^2 L}{\partial\beta\partial\beta}(\beta^0)(\hat{\beta}-\beta^0) \\ +(\tilde{\beta}-\beta^0)'\frac{\partial^2 L}{\partial\beta\partial\beta}(\beta^0)(\tilde{\beta}-\beta^0). \end{aligned} \tag{33}$$

In fact, (33) represents a transform of the likelihood ratio (LR) and as such it is a LR test statistic. We shall now show that in the case where

$$H_0\colon \ \beta_{(2)}^0 = 0, \qquad \beta^0 = \begin{pmatrix} \beta_{(1)}^0 \\ \beta_{(2)}^0 \end{pmatrix},$$

the quantity in the right member of (33) reduces to a test[2] based on the marginal (limiting) distribution of

$$\sqrt{T}\left(\hat{\beta}_{(2)}-\beta^0_{(2)}\right).$$

To this effect put

$$\tilde{C}_* = \frac{1}{T}\frac{\partial^2 L}{\partial\beta\partial\beta}(\beta^0),$$

and note that

$$\frac{1}{\sqrt{T}}\frac{\partial L}{\partial\beta}(\beta^0) \sim -\tilde{C}_*\sqrt{T}(\hat{\beta}-\beta^0). \tag{34}$$

Partitioning

$$\tilde{C}_* = \begin{bmatrix} C_{*11} & C_{*12} \\ C_{*21} & C_{*22} \end{bmatrix},$$

conformably with

$$(\hat{\beta}-\beta^0) = \begin{bmatrix} \hat{\beta}_{(1)}-\beta^0_{(1)} \\ \hat{\beta}_{(2)}-\beta^0_{(2)} \end{bmatrix},$$

we find

$$\frac{1}{\sqrt{T}}\frac{\partial L}{\partial\beta_{(1)}}(\beta^0) \sim -\left[\tilde{C}_{*11}\sqrt{T}\left(\hat{\beta}_{(1)}-\beta^0_{(1)}\right)+\tilde{C}_{*12}\sqrt{T}\left(\hat{\beta}_{(2)}-\beta^0_{(2)}\right)\right],$$

$$\frac{1}{\sqrt{T}}\frac{\partial L}{\partial\beta_{(2)}}(\beta^0) \sim -\left[\tilde{C}_{*21}\sqrt{T}\left(\hat{\beta}_{(1)}-\beta^0_{(1)}\right)+\tilde{C}_{*22}\sqrt{T}\left(\hat{\beta}_{(2)}-\beta^0_{(2)}\right)\right]. \tag{35}$$

Using (34) we can rewrite (33) as

$$\begin{aligned} -2[L(\hat{\beta})-L(\tilde{\beta})] &\sim -(\hat{\beta}-\beta^0)'\frac{\partial L}{\partial\beta}(\beta^0)+(\tilde{\beta}-\beta^0)'\frac{\partial L}{\partial\beta}(\beta^0) \\ &= -\left\{\left[\left(\hat{\beta}_{(1)}-\beta^0_{(1)}\right)-\left(\tilde{\beta}_{(1)}-\beta^0_{(1)}\right)\right]'\frac{\partial L}{\partial\beta_{(1)}}(\beta^0)\right. \\ &\qquad \left.+\left(\hat{\beta}_{(2)}-\beta^0_{(2)}\right)'\frac{\partial L}{\partial\beta_{(2)}}(\beta^0)\right\}. \end{aligned}$$

[2] It should be remarked that a similar result in the context of the GLM is called, somewhat redundantly, a Chow test.

From (34) we find, bearing in mind that under H_0 we estimate

$$\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_{(1)} \\ 0 \end{pmatrix},$$

$$\sqrt{T}\left(\tilde{\beta}_{(1)} - \beta^0_{(1)}\right) \sim -\tilde{C}^{-1}_{*11}\frac{1}{\sqrt{T}}\frac{\partial L}{\partial \beta_{(1)}}(\beta^0). \tag{36}$$

From (35) we find

$$-\sqrt{T}\left(\hat{\beta}_{(1)} - \beta^0_{(1)}\right) \sim \tilde{C}^{-1}_{*11}\left[\frac{1}{\sqrt{T}}\frac{\partial L}{\partial \beta_{(1)}}(\beta^0) + \tilde{C}_{*12}\sqrt{T}\left(\hat{\beta}_{(2)} - \beta^0_{(2)}\right)\right]. \tag{37}$$

Hence

$$\left[\sqrt{T}\left(\tilde{\beta}_{(1)} - \beta^0_{(1)}\right) - \sqrt{T}\left(\hat{\beta}_{(1)} - \beta^0_{(1)}\right)\right] \sim \tilde{C}^{-1}_{*11}\tilde{C}_{*12}\sqrt{T}\left(\hat{\beta}_{(2)} - \beta^0_{(2)}\right),$$

and thus (33) may be further rewritten as

$$-2[L(\hat{\beta}) - L(\tilde{\beta})] \sim \left[\hat{\beta}_{(2)} - \beta^0_{(2)}\right]'\left[\tilde{C}_{*21}C^{-1}_{*11}\frac{\partial L}{\partial \beta_{(1)}}(\beta^0) - \frac{\partial L}{\partial \beta_{(2)}}(\beta^0)\right]. \tag{38}$$

Again, from (35) we see that

$$\tilde{C}_{*21}\tilde{C}^{-1}_{*11}\frac{1}{\sqrt{T}}\frac{\partial L}{\partial \beta_{(1)}}(\beta^0) - \frac{1}{\sqrt{T}}\frac{\partial L}{\partial \beta_{(2)}}(\beta^0)$$
$$\sim \left[\tilde{C}_{*22} - \tilde{C}_{*21}\tilde{C}^{-1}_{*11}\tilde{C}_{*12}\right]\cdot\sqrt{T}\left(\hat{\beta}_{(2)} - \beta^0_{(2)}\right),$$

and thus (38) reduces to

$$-2[L(\hat{\beta}) - L(\tilde{\beta})] \sim T\left(\hat{\beta}_{(2)} - \beta^0_{(2)}\right)'\left[\tilde{C}_{*22} - \tilde{C}_{*21}\tilde{C}^{-1}_{*11}\tilde{C}_{*12}\right]\left(\hat{\beta}_{(2)} - \beta^0_{(2)}\right). \tag{39}$$

But under H_0, (39) is exactly the test statistic based on the (limiting) marginal distribution of

$$\sqrt{T}\left(\hat{\beta}_{(2)} - \beta^0_{(2)}\right) \sim N(0, 1 - C_{22}), \tag{40}$$

where

$$C_{22} = \operatorname*{plim}_{T \to \infty}\left[\tilde{C}_{*22} - \tilde{C}_{*21}\tilde{C}^{-1}_{*11}\tilde{C}_{*12}\right]^{-1}. \tag{41}$$

In the special case where

$$\beta_{(1)} = \beta_0,$$

i.e. it is the constant term in the expression

$$x_{t\cdot}\beta,$$

so that no bona fide explanatory variables "explain" the probability of choice, we can define R^2 by

$$R^2 = 1 - \frac{L(\hat{\beta})}{L(\tilde{\beta})}. \tag{42}$$

The quantity in (42) has the property

i. $R^2 \in [0,1)$
ii. the larger the contribution of the bona fide variables to the maximum of the LF the closer is R^2 to 1
iii. R^2 stands in a one-to-one relation to the chi-square statistic for testing the hypothesis that the coefficients of the bona fide variables are zero. In fact, under H_0

$$-2L(\tilde{\beta})R^2 \sim \chi^2_{s+m-1}.$$

It is desirable, in empirical practice, that a statistic like R^2 be reported and that a constant term be routinely included in the specification of the linear functional

$$x_{t\cdot}\beta,$$

Finally, we should also stress that R^2 as in (42) *does not* have the interpretation as the square of the correlation coefficient between "predicted" and "actual" observations.

2. Truncated dependent variables

2.1. *Generalities*

Suppose we have a sample conveying information on consumer expenditures; in particular, suppose we are interested in studying household expenditures on consumer durables. In such a sample survey it would be routine that many households report zero expenditures on consumer durables. This was, in fact, the situation faced by Tobin (1958) and he chose to model household expenditure on consumer durables as

$$\begin{aligned} y_t &= x_{t\cdot}\beta + u_t, && \text{if } x_{t\cdot}\beta + u_t > 0 \\ &= 0 && \text{otherwise} \end{aligned} . \tag{43}$$

The same model was later studied by Amemiya (1973). We shall examine below the inference and distributional problem posed by the manner in which the model's dependent variable is truncated.

2.2. *Why simple OLS procedures fail*

Let us append to the model in (43) the standard assumptions that

(A.2.1.) The $\{u_t\colon t=1,2,\ldots\}$ is a sequence of i.i.d. random variables with

$$u_t \sim N(0,\sigma^2), \qquad \sigma^2 \in (0,\infty).$$

(A.2.2.) The elements of $x_{t\cdot}$ are bounded for all t, i.e.

$$|x_{ti}| < k_i, \qquad \text{for all } t, \qquad i=1,2,\ldots,n,$$

are linearly independent and

$$(p)\lim_{T\to\infty} \frac{X'X}{T} = M,$$

exists as a nonsingular nonstochastic matrix.

(A.2.3.) If the elements of $x_{t\cdot}$ are stochastic, then $x_{t\cdot}, u_{t'}$ are mutually independent for all t, t', i.e. the error and data generating processes are mutually independent.

(A.2.4.) The parameter space, say $H \subset R_{n+2}$, is compact and it contains an open neighborhood of the true parameter point $(\beta^{0\prime}, \sigma_0^2)'$.

The first question that occurs is why not use the entire sample to estimate β? Thus, defining

$$X = (x_{t\cdot}), \qquad t=1,2,\ldots,T,$$

$$u = (u_1,u_2,\ldots,u_T)', \qquad y^{(1)} = (y_1,y_2,\ldots,y_{T1})', \qquad y^{(2)} = (0,\ldots,0)',$$

$$y = \left(y^{(1)\prime},\ y^{(2)\prime}\right)',$$

we may write

$$y = X\beta + u,$$

and estimate β by

$$\tilde{\beta} = (X'X)^{-1}X'y. \tag{44}$$

A little reflection will show, however, that this leads to serious and palpable specification error *since in (43) we do not assert that the zero observations are generated by the same process that generates the positive observations.* Indeed, a little further reflection would convince us that it would be utterly inappropriate to insist that the same process that generates the zero observations should also generate the nonzero observations, since for the zero observations we should have that

$$u_t = -x_{t\cdot}\beta, \qquad t = T_{1+1},\dots,T_1 + T_2,$$

and this would be inconsistent with assumption (A.1.1.).

We next ask, why not confine our sample solely to the nonzero observations,

$$y^{(1)} = X_1\beta + u_{(1)},$$

and thus estimate β by

$$\tilde{\beta} = \left(X_1'X_1\right)^{-1}X_1'y^{(1)}.$$

This may appear quite reasonable at first, even though it is also apparent that we are ignoring some (perhaps considerable) information. Deeper probing, however, will disclose a much more serious problem. After all, ignoring some sample elements would affect only the degrees of freedom and the t- and F-statistics alone. If we already have a large sample, throwing out even a substantial part of it will not affect matters much. But now it is in order to ask: What is the process by which some dependent variables are assigned the value zero? A look at (43) convinces us that it is a random process governed by the behavior of the error process and the characteristics relevant to the economic agent, $x_{t\cdot}$. Conversely, the manner in which the sample on the basis of which we shall estimate β is selected is governed by some aspects of the error process. In particular we note that for us to observe a positive y_t, according to

$$y_t = x_{t\cdot}\beta + u_t, \tag{45}$$

the error process should satisfy

$$u_t > -x_{t\cdot}\beta. \tag{46}$$

Thus, for the positive observations we should be dealing with the *truncated* distribution function of the error process. But, what is the mean of the truncated distribution? We have, if $f(\cdot)$ is the density and $F(\cdot)$ the c.d.f. of u_t

$$E(u_t|u_t > -x_{t\cdot}\beta) = \frac{1}{1-F(-x_{t\cdot}\beta)}\int_{-x_{t\cdot}\beta}^{\infty} \xi f(\xi)\,\mathrm{d}\xi.$$

If $f(\cdot)$ is the $N(0,\sigma^2)$ density the integral can be evaluated as

$$f(x_{t\cdot}\beta),$$

and, in addition, we also find

$$1-F(-x_{t\cdot}\beta) = F(x_{t\cdot}\beta).$$

Moreover, if we denote by $\phi(\cdot)$, $\Phi(\cdot)$ the $N(0,1)$ density and c.d.f., respectively, and by

$$\nu_t = \frac{x_{t\cdot}\beta}{\sigma}, \tag{47}$$

then

$$E(u_t|u_t > +x_{t\cdot}\beta) = \sigma\frac{\phi(\nu_t)}{\Phi(\nu_t)} = \sigma\psi_t. \tag{48}$$

Since the mean of the error process in (45) is given by (48) we see that we are committing a misspecification error by leaving out the "variable" $\phi(\nu_t)/\Phi(\nu_t)$ [see Dhrymes (1978a)].

Defining

$$v_t = u_t - \sigma\frac{\phi(\nu_t)}{\Phi(\nu_t)}, \tag{49}$$

we see that $\{v_t\colon\ t=1,2,\dots\}$ is a sequence of independent *but non-identically distributed random variables*, since

$$\operatorname{Var}(v_t) = \sigma^2(1-\nu_t\psi_t-\psi_t^2). \tag{50}$$

Thus, there is no simple procedure by which we can obtain efficient and/or consistent estimators by confining ourselves to the positive subsample; consequently, we are forced to revert to the entire sample and employ ML methods.

2.3. Estimation of parameters with ML methods

We are operating with the model in (43), subject to (A.2.1.) through (A.2.4.) and the convention that the first T_1 observations correspond to positive dependent variables, while the remaining T_2, $(T_1 + T_2 = T)$, correspond to zero observations.

Define

$$c_t = 1 \quad \text{if } y_t > 0, \qquad = 0 \quad \text{otherwise}, \tag{51}$$

and note that the (log) LF can be written as

$$L = \sum_{t=1}^{T}\left\{(1-c_t)\ln\Phi(-\nu_t) - c_t\left[\frac{1}{2}\ln(2\pi) + \frac{1}{2}\ln\sigma^2 + \frac{1}{2\sigma^2}(y_t - x_{t\cdot}\beta)^2\right]\right\}. \tag{52}$$

Differentiating with respect to $\gamma = (\beta', \sigma^2)'$, we have

$$\frac{\partial L}{\partial \beta} = -\frac{1}{\sigma}\sum_{t=1}^{T}\left\{(1-c_t)\frac{\phi(\nu_t)}{\Phi(-\nu_t)} - c_t\left(\frac{y_t - x_{t\cdot}\beta}{\sigma}\right)\right\}x_{t\cdot} = 0, \tag{53}$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left\{c_t\left[1 - \frac{1}{\sigma^2}(y_t - x_{t\cdot}\beta)^2\right] - (1-c_t)\frac{\nu_t\phi(\nu_t)}{\Phi(-\nu_t)}\right\} = 0,$$

and these equations have to be solved in order to obtain the ML estimator. It is, first, interesting to examine how the conditions in (53) differ from the equations to be satisfied by simple OLS estimators applied to the positive component of the sample. By simple rearrangement we obtain, using the convention alluded to above,

$$X_1'X_1\beta = X_1'y^{(1)} - \sigma\sum_{t=T_1+1}^{T}\psi(-\nu_t)x_{t\cdot}', \tag{54}$$

$$\sigma^2 = \frac{1}{T_1}\left(y^{(1)} - X_1\beta\right)'\left(y^{(1)} - X_1\beta\right) + \frac{\sigma^2}{T_1}\sum_{t=T_1+1}^{T}\psi(-\nu_t)\nu_t, \tag{55}$$

where

$$\psi(\nu_t) = \frac{\phi(\nu_t)}{\Phi(\nu_t)}, \qquad \psi(-\nu_t) = \frac{\phi(\nu_t)}{\Phi(-\nu_t)}. \tag{56}$$

Since these expressions occur very frequently, we shall often employ the abbreviated notation

$$\psi_t = \psi(\nu_t), \qquad \psi_t^* = \psi(-\nu_t).$$

Thus, if in some sense

$$z'_{T\cdot} = \sum_{t=T_1+1}^{T} \psi_t^* x'_{t\cdot}, \tag{57}$$

is negligible, the ML estimator, say $\hat{\beta}$, could yield results that are quite similar, from an applications point of view, to those obtained through the simple OLS estimator, say $\tilde{\beta}$, as applied to the positive component of the sample. From (54) it is evident that if $z'_{T\cdot}$ of (57) is small then

$$\sigma^2 \sum_{t=T_1+1}^{T} \psi_t^* \nu_t = \sigma z_{T\cdot}\beta,$$

is also small. Hence, under these circumstances

$$\hat{\beta} = \tilde{\beta}, \quad \hat{\sigma}^2 = \tilde{\sigma}^2$$

which explains the experience occasionally encountered in empirical applications.

The eqs. (53) or (54) and (55) are highly nonlinear and can only be solved by iterative methods. In order to ensure that the root of

$$\frac{\partial L}{\partial \gamma} = 0, \qquad \gamma = (\beta', \sigma^2)',$$

so located is the ML estimator it is necessary to show either that the equation above has only one root – which is difficult – or that we begin the iteration with an initial consistent estimator.

2.4. *An initial consistent estimator*

Bearing in mind the development in the preceding section we can rewrite the model describing the positive component of the sample as

$$y_t = x_{t\cdot}\beta + \sigma\psi_t + v_t = \sigma(\nu_t + \psi_t) + v_t, \tag{58}$$

such that

$$\{v_t\colon t = 1, 2, \dots\},$$

is a sequence of mutually independent random variables with

$$E(v_t)=0, \qquad \mathrm{Var}(v_t)=\sigma^2(1-\nu_t\psi_t-\psi_t^2), \tag{59}$$

and such that they are independent of the explanatory variables x_t.

The model in (58) cannot be estimated by simple means owing to the fact that ψ_t is not directly observable; thus, we are forced into nonstandard procedures.

We shall present below a modification and simplification of a consistent estimator due to Amemiya (1973). First we note that, confining our attention to the positive component of the sample

$$y_t^2=\sigma^2(\nu_t+\psi_t)^2+v_t^2+2v_t(\nu_t+\psi_t)\sigma. \tag{60}$$

Hence

$$\begin{aligned} E(y_t^2|x_{t\cdot},u_t>-x_{t\cdot}\beta)&=\sigma^2(\nu_t^2+\nu_t\psi_t)+\sigma^2\\ &=x_{t\cdot}\beta E(y_t|x_{t\cdot},u_t>-x_{t\cdot}\beta)+\sigma^2. \end{aligned} \tag{61}$$

Defining

$$\varepsilon_t\equiv y_t^2-E(y_t^2|x_{t\cdot},u_t>-x_{t\cdot}\beta), \tag{62}$$

we see that $\{\varepsilon_t\colon t=1,2,\dots\}$ is a sequence of independent random variables with mean zero and, furthermore, we can write

$$w_t=y_t^2=x_{t\cdot}\beta y_t+\sigma^2+\varepsilon_t, \qquad t=1,2,\dots,T_1. \tag{63}$$

The problem, of course, is that y_t is correlated with ε_t and hence simple regression will not produce a consistent estimator for β and σ^2.

However, we can employ an instrumental variables (I.V.) estimator[3]

$$\bar{\gamma}=(\tilde{X}_*'X_*)^{-1}X_*'w, \qquad w=(w_1,w_2,\dots,w_{T_1})', \tag{64}$$

[3] It is here that the procedure differs from that suggested by Amemiya (1973). He defines

$$\tilde{y}_t=x_{t\cdot}(X_1'X_1)^{-1}X_1'y^{(1)},$$

while we define

$$\tilde{y}_t=x_{t\cdot}a,$$

for nontrivial vector a.

where

$$X_* = (D_y X_1, e), \qquad \tilde{X}_* = (D_{\tilde{y}} X_1, e), \tag{65}$$

and

$$\tilde{y}_t = x_{t\cdot} a, \qquad D_{\tilde{y}} = \operatorname{diag}(\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_{T_1}), \qquad D_y = (y_1, y_2, \dots, y_{T_1}), \tag{66}$$

for an arbitrary nontrivial vector a.

It is clear that by substitution we find

$$\tilde{\gamma} = \gamma + (\tilde{X}_*' X_*)^{-1} \tilde{X}_*' \varepsilon. \tag{67}$$

We easily establish that

$$\tilde{X}_*' X_* = \begin{bmatrix} X_1' D_{\tilde{y}} D_y X_1 & X_1' \tilde{y} \\ y' X_1 & e'e \end{bmatrix}.$$

Clearly

$$\lim_{T\to\infty} \frac{e'e}{T_1} = 1, \qquad \lim_{T\to\infty} \frac{1}{T_1} X_1' \tilde{y} = \left(\lim_{T\to\infty} \frac{X_1' X_1}{T} \right) a,$$

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} X_1' y = \left(\lim_{T\to\infty} \frac{X_1' X_1}{T_1} \right) \beta + \operatorname*{plim}_{T\to\infty} \frac{X_1' u_{\cdot 1}}{T_1}.$$

Now

$$\frac{1}{T_1} X_1' u_{\cdot 1} = \frac{1}{T_1} \sum_{t=1}^{T_1} x_{t\cdot}' u_t,$$

and

$$\{ x_{t\cdot}' u_t \colon t = 1, 2, \dots \},$$

is a sequence of independent random variables with mean

$$E(x_{t\cdot}' u_t) = \sigma x_{t\cdot}' \psi_t, \tag{68}$$

and covariance matrix

$$\operatorname{Cov}(x_{t\cdot}' u_t) = \sigma^2 (1 - \nu_t \psi_t - \psi_t) x_{t\cdot}' x_{t\cdot} = \omega_t x_{t\cdot}' x_{t\cdot}, \tag{69}$$

where

$$\omega_t = \sigma^2(1 - \nu_t\psi_t - \psi_t^2),$$

is uniformly bounded by assumption (A.2.2) and (A.2.4). Hence, by (A.2.2)

$$\lim_{T_1 \to \infty} \frac{1}{T_1} \sum_{t=1}^{T_1} \omega_t x'_{t\cdot} x_{t\cdot},$$

converges to a matrix with finite elements. Further and similar calculations will show that

$$\frac{\tilde{X}'_* X_*}{T},$$

converges a.c. to a nonsingular matrix. Thus, we are reduced to examining the limiting behavior of

$$\frac{1}{\sqrt{T_1}} \tilde{X}'_* \varepsilon = \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \begin{pmatrix} x_{t\cdot} a x'_{t\cdot} \\ 1 \end{pmatrix} \varepsilon_t. \tag{70}$$

But this is a sequence of independent nonidentically distributed random variables with mean zero and uniformly bounded (in $x_{t\cdot}$ and β) moments to any finite order. Now for any arbitrary $(n+2\times 1)$ vector α^* consider

$$\frac{1}{\sqrt{T_1}} \alpha^{*\prime} \tilde{X}'_* \varepsilon = \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \alpha_t \varepsilon_t, \tag{71}$$

where

$$\alpha_t = (x_{t\cdot}\alpha)(x_{t\cdot}a), \qquad \alpha^* = (\alpha', \alpha_{n+2}),$$

and note that

$$\lim_{T_1 \to \infty} \frac{S_{T_1}^{*2}}{T_1} = S,$$

is well defined where

$$S_{T_1}^{*2} = \sum_{t=1}^{T_1} \alpha_t^2 \operatorname{Var}(\varepsilon_t). \tag{72}$$

Define, further

$$S_{T_1}^2 = \frac{S_{T_1}^{*2}}{T_1},$$

and note that

$$S_{T_1}^* = T_1^{1/2} S_{T_1}.$$

But then it is evident that Liapounov's condition is satisfied, i.e. with K a uniform bound on $E|\alpha_t \varepsilon_t|^{2+\delta}$

$$\lim_{T_1 \to \infty} \frac{\sum_{t=1}^{T_1} E|\alpha_t \varepsilon_t|^{2+\delta}}{S_{T_1}^{*2+\delta}} \le K \lim_{T \to \infty} \frac{T_1}{T_1^{1+\delta/2} S_{T_1}^{2+\delta}} = \lim_{T_1 \to \infty} \frac{K}{T_1^{\delta/2} S^{2+\delta}} = 0.$$

By a theorem of Varadarajan, see Dhrymes (1970), we conclude that

$$\frac{1}{\sqrt{T_1}} \tilde{X}_*' \varepsilon \sim N(0, H),$$

where

$$H = \lim_{T \to \infty} \frac{1}{T_1} \begin{bmatrix} \sum_{t=1}^{T_1} (x_{t\cdot} a)^2 x_{t\cdot}' x_{t\cdot} \operatorname{Var}(\varepsilon_t) & \sum_{t=1}^{T_1} (x_{t\cdot} a) x_{t\cdot}' \operatorname{Var}(\varepsilon_t) \\ \sum_{t=1}^{T_1} (x_{t\cdot} a) x_{t\cdot} \operatorname{Var}(\varepsilon_t) & \sum_{t=1}^{T_1} \operatorname{Var}(\varepsilon_t) \end{bmatrix}. \tag{73}$$

Consequently we have shown that

$$\sqrt{T_1}(\tilde{\gamma} - \gamma) \sim N(0, Q^{-1} H Q^{-1}),$$

where

$$Q = \lim_{\text{a.c.}} \frac{(\tilde{X}_*' X_*)}{T_1}. \tag{74}$$

Moreover since

$$\sqrt{T_1}(\tilde{\gamma} - \gamma) \sim \zeta,$$

where ζ is an a.c. finite random vector it follows that

$$\tilde{\gamma}-\gamma_0 \sim \frac{\zeta}{\sqrt{T_1}},$$

which shows that $\tilde{\gamma}$ converges a.c. to γ_0.

We may summarize the development above in

Lemma 1

Consider the model in (43) subject to assumptions (A.2.1.) through (A.2.4.); further consider the I.V. estimator of the parameter vector γ in

$$w_t = (x_{t\cdot}y_t, 1)\gamma + \varepsilon_t, \qquad w_t = y_t^2,$$

given by

$$\tilde{\gamma} = (\tilde{X}_*'X_*)^{-1}\tilde{X}_*'w,$$

where $\tilde{X}_*$, X_* and w are as defined in (65) and (66). Then

i. $\tilde{\gamma}$ converges to γ_0 almost certainly,
ii. $\sqrt{T_1}(\tilde{\gamma}-\gamma_0) \sim N(0, Q^{-1}HQ'^{-1})$,

where Q and H are as defined in (74) and (73) respectively.

2.5. *Limiting properties and distribution of the ML estimator*

Returning now to eqs. (53) or (54) and (55) we observe that since the initial estimator, say $\tilde{\gamma}$, is strongly consistent, at each step of the iterative procedure we get a (strongly) consistent estimator. Hence, at convergence, the estimator so determined, say $\hat{\gamma}$, is guaranteed to be (strongly) consistent.

The perceptive reader may ask: Why did we not use the apparatus of Section 1.d. instead of going through the intermediate step of obtaining the initial consistent estimator? The answer is, essentially, that Theorem 1 (of Section 1.d.) does not hold in the current context. To see that, recall the (log) LF of our problem and write it as

$$L_T(\gamma) = \frac{1}{T}\sum_{t=1}^{T}\Big\{(1-c_t)\ln\Phi(-\nu_t) - c_t \cdot\left[\frac{1}{2}\ln(2\pi)+\frac{1}{2}\ln\sigma^2+\frac{1}{2\sigma^2}(y_t - x_{t\cdot}\beta)^2\right]\Big\}. \tag{75}$$

Since L_T is at least twice differentiable it is concave if and only if its Hessian is negative (semi)definite over the space of admissible γ-parameters. After some manipulation we can show that

$$\frac{\partial^2 L_T}{\partial\sigma^2\partial\sigma^2} = -\frac{1}{4\sigma^4}\frac{1}{T}\left\{\sum_{t=1}^{T}(1-c_t)\frac{\nu_t\phi(\nu_t)}{\Phi(-\nu_t)}\left(3-\nu_t^2+\frac{\nu_t\phi(\nu_t)}{\Phi(-\nu_t)}\right)\right.$$
$$\left.+c_t\left[4\frac{(y_t-x_{t\cdot}\beta)^2}{\sigma^2}-2\right]\right\}.$$

When $\beta = 0$ the entire first term in brackets is null so that the derivative reduces to

$$-\frac{1}{\sigma^4}\frac{1}{T}\sum_{t=1}^{T}c_t\left(\frac{y_t^2}{\sigma^2}-\frac{1}{2}\right),$$

which could well be positive for some realizations. Hence, we cannot unambiguously assert that over some (large) compact subset of R_{n+2}, the Hessian of the (log) LF is negative semidefinite. Consequently, we have no assurance that, if we attempted to solve

$$\frac{\partial L_T}{\partial\gamma}(\gamma) = 0, \tag{76}$$

beginning with an arbitrary initial point, say $\tilde{\gamma}$, upon convergence we should arrive at the consistent root of (93). On the other hand, from the general theory of ML estimation we know that if the true parameter point is in the interior of the γ-admissible space then (76) has at most one consistent root. Of course, it may have many roots if the function L_T is nonconcave and herein lies the problem. In the previous Section, however, because of Theorem 1 we knew that the (log) likelihood function was concave and hence starting from an arbitrary point we could locate, upon convergence, the global maximizer and hence the ML estimator.

Many of the other results of Section 1.d., however, are available to us in virtue of

Lemma 2

The (log) LF of the problem of this section as exhibited in (75) converges a.c. uniformly in γ. In particular

$$L_T(\gamma) \overset{\text{a.c.}}{\to} \lim_{T\to\infty} E[L_T(\gamma)],$$

uniformly in γ.

Proof

Consider the log LF of (75) and in particular its tth term

$$\xi_t = (1-c_t)\ln\Phi(-\nu_t) - c_t\left[\frac{1}{2}\ln(2\pi) + \frac{1}{2}\ln\sigma^2 + \frac{1}{2\sigma^2}(y_t - x_{t\cdot}\beta)^2\right],$$
$$t = 1,2,\ldots. \quad (77)$$

For any x-realization

$$\{\xi_t\colon t = 1,2,\ldots\},$$

is a sequence of independent random variables with uniformly bounded moments in virtue of assumption (A.2.1) through (A.2.3). Thus, there exists a constant, say k, such that

$$\operatorname{Var}(\xi_t) < k, \qquad \text{for all } t.$$

Consequently, by Kolmogorov's criterion, for all admissible γ,

$$\{L_T(\gamma) - E[L_T(\gamma)]\} \overset{\text{a.c.}}{\to} 0.$$ Q.E.D.

Remark 3

The device of beginning the iterative process for solving (76) with a consistent estimator ensures that for sufficiently large T we will be locating the estimator, say $\hat{\gamma}_T$, satisfying

$$L_T(\hat{\gamma}_T) = \sup_{\gamma} L_T(\gamma).$$

Lemma 2, can be shown to imply that

$$L_T(\hat{\gamma}_T) \overset{\text{a.c.}}{\to} \bar{L}(\bar{\gamma}, \gamma^0), \qquad \bar{L}(\bar{\gamma}, \gamma^0) = \sup_{\gamma} \bar{L}(\gamma, \gamma^0).$$

Moreover, we can also show that

$$\bar{\gamma} = \gamma^0.$$

On the other hand, it is not possible to show routinely that $\hat{\gamma}_T \overset{\text{a.c.}}{\to} \gamma^0$. Essentially, the problem is the term corresponding to σ^2 which contains expressions like

$$c_t \frac{(y_t - x_{t\cdot}\beta)^2}{\sigma^2},$$

which cannot be (absolutely) bounded. This does not prevent us from showing convergence a.c. of $\hat{\gamma}_T$ to γ^0. By the iterative process we have shown that $\hat{\gamma}_T$ converges to γ^0 at least in probability. Convergence a.c. is shown easily once we obtain the limiting distribution of $\hat{\gamma}_T$ – a task to which we now turn.

Thus, as before, consider the expansion

$$\frac{\partial L_T}{\partial \gamma}(\hat{\gamma}_T) = \frac{\partial L_T}{\partial \gamma}(\gamma^0) + \frac{\partial^2 L_T}{\partial \gamma \partial \gamma}(\gamma^*)(\hat{\gamma}_T - \gamma^0), \tag{78}$$

where γ^0 is the true parameter point and

$$|\hat{\gamma}_T - \gamma^0| \leq |\gamma^* - \gamma^0|.$$

We already have an explicit expression in eq. (53) for the derivative $\partial L_T/\partial \gamma$. So let us obtain the Hessian of the LF. We find

$$\begin{aligned}
\frac{\partial^2 L_T}{\partial \beta \partial \beta}(\gamma) &= -\frac{1}{\sigma^2}\frac{1}{T}\sum_{t=1}^{T}\left\{(1-c_t)\psi_t^*(\psi_t^* - \nu_t) + c_t\right\} x_{t\cdot}' x_{t\cdot}, \\
\frac{\partial^2 L_T}{\partial \beta \partial \sigma^2} &= -\frac{1}{2\sigma^3}\frac{1}{T}\sum_{t=1}^{T}\left\{2c_t\left(\frac{y_t - x_{t\cdot}\beta}{\sigma}\right) - (1-c_t)\psi_t^*(1 + \nu_t\psi_t^* - \nu_t^2)\right\} x_{t\cdot}, \\
\frac{\partial^2 L_T}{\partial \sigma^2 \partial \sigma^2} &= -\frac{1}{4\sigma^4}\frac{1}{T}\sum_{t=1}^{T}\left\{c_t^2\left(\frac{y_t - x_{t\cdot}\beta}{\sigma}\right)^2 + (1-c_t)\nu_t\psi_t^*(1 + \nu_t\psi_t^* - \nu_t^2)\right\} \\
&\quad + \frac{1}{2\sigma^4}\frac{1}{T}\sum_{t=1}^{T}\left\{c_t\left[1 - \left(\frac{y_t - x_{t\cdot}\beta}{\sigma}\right)^2\right] - (1-c_t)\nu_t\psi_t^*\right\}.
\end{aligned} \tag{79}$$

We may now define

$$\begin{aligned}
\xi_{1t} &= (1-c_t)\frac{\phi(\nu_t^0)}{\Phi(-\nu_t^0)} - c_t\left(\frac{y_t - x_{t\cdot}\beta^0}{\sigma_0}\right), \\
\xi_{2t} &= c_t\left[1 - \left(\frac{y_t - x_{t\cdot}\beta^0}{\sigma_0}\right)^2\right] - (1-c_t)\frac{\nu_t^0\phi(\nu_t^0)}{\Phi(-\nu_t^0)},
\end{aligned} \tag{80}$$

and

$$\begin{aligned}
\xi_{11t} &= (1-c_t)\psi_t^{*0}(\psi_t^{*0} - \nu_t^0) + c_t, \\
\xi_{12t} &= \xi_{21t} = (1-c_t)\psi_t^{*0}(1 + \nu_t^0 - \nu_t^0\psi_t^{*0}), \\
\xi_{22t} &= c_t^2\left(\frac{y_t - x_{t\cdot}\beta^0}{\sigma_0}\right)^2 + (1-c_t)\nu_t^0\psi_t^{*0}(1 + \nu_t^0\psi_t^{*0} - \nu_t^{02}),
\end{aligned} \tag{81}$$

where, evidently,

$$\psi_t^{*0} = \frac{\phi(\nu_t^0)}{\Phi(-\nu_t^0)}, \qquad \nu_t^0 = \frac{x_{t\cdot}\beta^0}{\sigma_0}, \qquad \psi_t^0 = \frac{\phi(\nu_t^0)}{\Phi(\nu_t^0)}.$$

With the help of the notation in (80) and (81) we find

$$\frac{\partial L_T}{\partial \gamma'}(\gamma^0) = -\frac{1}{\sigma_0}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} x'_{t\cdot} & 0 \\ 0 & \frac{1}{2\sigma_0}\end{bmatrix}\begin{bmatrix}\xi_{1t} \\ \xi_{2t}\end{bmatrix}, \tag{82}$$

and

$$\frac{\partial^2 L_T}{\partial \gamma \partial \gamma}(\gamma^0) = -\frac{1}{\sigma_0^2}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \xi_{11t}x'_{t\cdot}x_{t\cdot} & \frac{1}{2\sigma_0}\xi_{12t}x'_{t\cdot} \\ \frac{1}{2\sigma_0}\xi_{21t}x_{t\cdot} & \frac{1}{4\sigma^2}\xi_{22t}\end{bmatrix} + \Omega_{*T}, \tag{83}$$

where Ω_{*T} is a matrix all of whose elements are zero except the last diagonal element, which is

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{2\sigma^4}\xi_{2t\cdot}.$$

Thus, for every T we have

$$E(\Omega_{*T}) = 0. \tag{84}$$

Consequently, we are now ready to prove

Theorem 3

Consider the model of eq. (43) subject to assumption (A.2.1.) through (A.2.4.); moreover, consider the ML estimator, $\hat{\gamma}_T$, obtained by iteration from an initial consistent estimator as a solution of (76). Then

$$\sqrt{T}(\hat{\gamma}_T - \gamma^0) \sim N(0, \sigma_0^2 C^{-1}),$$

where

$$C = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \omega_{11t}x'_{t\cdot}x_{t\cdot} & \frac{1}{2\sigma_0}\omega_{12t}x'_{t\cdot} \\ \frac{1}{2\sigma_0}\omega_{21t}x_{t\cdot} & \frac{1}{4\sigma_0^2}\omega_{22t}\end{bmatrix},$$

and

$$\omega_{ijt} = E(\xi_{ijt}) \qquad i, j = 1, 2.$$

Proof

From the expansion in (78) and the condition under which the ML estimator is obtained we find

$$\sqrt{T}(\hat{\gamma}_T - \gamma^0) = -\left[\frac{\partial^2 L_T}{\partial\gamma\partial\gamma}(\gamma^*)\right]^{-1} \frac{1}{\sqrt{T}} \frac{\partial L}{\partial\gamma}(\gamma^0).$$

But

$$\frac{1}{\sqrt{T}} \frac{\partial L}{\partial\gamma}(\gamma^0) = -\frac{1}{\sigma_0}\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{bmatrix} x'_{t\cdot} & 0 \\ 0 & \frac{1}{2\sigma_0} \end{bmatrix} \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix}. \tag{85}$$

The right member of (85) involves the sum of a sequence of independent random variables with mean zero. Moreover, it is easily verified that such variables have uniformly bounded moments to order at least four. Hence, a Liapounov condition holds. Since the covariance matrix of each term is

$$\frac{1}{\sigma_0^2}\begin{bmatrix} x'_{t\cdot}x_{t\cdot}\omega_{11t} & \frac{1}{2\sigma_0}x'_{t\cdot}\omega_{12t} \\ \frac{1}{2\sigma_0}x_{t\cdot}\omega_{21t} & \frac{1}{4\sigma_0^2}\omega_{22t} \end{bmatrix},$$

with

$$\begin{aligned} \omega_{11t} &= E(\xi_{1t}^2) = \Phi(\nu_t^0) + \psi(\nu_t^0)\left[\psi_t^{*0} - \nu_t^0\right], \\ \omega_{12t} &= \omega_{21t} = E(\xi_{1t}\xi_{2t}) = \phi(\nu_t^0)\left[1 - \nu_t^0\psi_t^{*0} + \nu_t^{02}\right], \\ \omega_{22t} &= E(\xi_{2t}^2) = 2\Phi(\nu_t^0) - \nu_t^0\phi(\nu_t^0)\left[1 - \nu_t^0\psi_t^{*0} + \nu_t^{02}\right]. \end{aligned} \tag{86}$$

Thus we see that

$$\frac{1}{\sqrt{T}} \frac{\partial L}{\partial\gamma}(\gamma^0) \sim N\left(0, \frac{1}{\sigma_0^2}C\right).$$

From (79) we also verify that

$$\frac{\partial^2 L_T}{\partial\gamma\partial\gamma}(\gamma^*) - \frac{\partial^2 L_T}{\partial\gamma\partial\gamma}(\gamma^0),$$

converges in probability to the null matrix, element by element. But the elements of

$$\frac{\partial^2 L_T}{\partial\gamma\partial\gamma}(\gamma^0),$$

are seen to be sums of independent random variables with finite means and bounded variances; hence, they obey a Kolmogorov criterion and thus

$$\frac{\partial^2 L_T}{\partial\gamma\partial\gamma}(\gamma^*) \overset{\text{a.c.}}{\to} \lim_{T\to\infty} E\left[\frac{\partial^2 L_T}{\partial\gamma\partial\gamma}(\gamma^0)\right].$$

We easily verify that

$$\begin{aligned} E(\xi_{11t}) &= \omega_{11t},\ E(\xi_{12t}) = E(\xi_{21t}) = \phi(\nu_t^0)\left[1 - \nu_t^0\psi_t^{*0} + \nu_t^{02}\right] \\ &= \omega_{12t} = \omega_{21t}, \\ E(\xi_{22t}) &= \omega_{22t}. \end{aligned}$$

Hence

$$\begin{aligned} \operatorname*{plim}_{T\to\infty} \frac{\partial^2 L_T}{\partial\gamma\partial\gamma}(\gamma^*) &= \lim_{T\to\infty} -\frac{1}{\sigma_0^2}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \omega_{11t}x_{t\cdot}'x_{t\cdot} & \frac{1}{2\sigma_0}\omega_{12t}x_{t\cdot}' \\ \frac{1}{2\sigma_0}\omega_{21t}x_{t\cdot} & \frac{1}{4\sigma_0^2}\omega_{22t} \end{bmatrix} \\ &= -\frac{1}{\sigma_0^2}C, \end{aligned}$$

and, moreover,

$$\sqrt{T}(\hat{\gamma} - \gamma^0) \sim N(0, \sigma_0^2 C^{-1}) \qquad \text{(Q.E.D.)}$$

Corollary 2

The estimator $\hat{\gamma}_T$ converges a.c. to γ_0.

Proof

From Theorem 3

$$\sqrt{T}\left(\hat{\gamma}_T - \gamma^0\right) \sim \zeta,$$

where ζ is an a.c. finite random variable; hence,

$$\hat{\gamma}_T - \gamma^0 \sim \frac{\zeta}{\sqrt{T}},$$

and thus

$$\hat{\gamma}_T \overset{\text{a.c.}}{\to} \gamma^0.$$

Corollary

The marginal (limiting) distribution of $\hat{\beta}_T$ is given by

$$\sqrt{T}\left(\hat{\beta}_T - \beta\right) \sim N\left(0, \sigma_0^2 P^{-1}\right),$$

where

$$P = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left(\omega_{11t} - \frac{\omega_{21t}^2}{\omega_{22t}} \right) x_{t\cdot}' x_{t\cdot}. \tag{87}$$

Proof

Evident from the definition of C in Theorem 3.

Remark 4

The unknown parameters of the limiting distribution of $\hat{\gamma}_T$ can be estimated by the standard procedure as

$$-\frac{\partial^2 L_T(\hat{\gamma}_T)}{\partial\gamma\partial\gamma}.$$

However, it would be much preferable to estimate C as

$$\hat{C} = \frac{1}{\hat{\sigma}^2}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \hat{\omega}_{11t}x_{t\cdot}'x_{t\cdot} & \frac{1}{2\hat{\sigma}}\hat{\omega}_{12t}x_{t\cdot}' \\ \frac{1}{2\hat{\sigma}}\hat{\omega}_{21t}x_{t\cdot} & \frac{1}{4\hat{\sigma}^2}\hat{\omega}_{22t} \end{bmatrix},$$

with $\hat{\omega}_{ijt}$ given as in (86) evaluated at $\hat{\gamma}_T$.

2.6. *Goodness of fit*

In the context of the truncated dependent variable model the question arises as to what we would want to mean by a "goodness of fit" statistic.

As analyzed in the Section on discrete choice models the usual R^2, in the context of the GLM, serves a multiplicity of purposes; when we complicate the process in which we operate it is not always possible to define a single statistic that would be meaningful in all contexts.

Since the model is

$$\begin{aligned} y_t &= x_{t\cdot}\beta + u_t \qquad && \text{if } x_{t\cdot}\beta + u_t > 0, \\ &= 0 && \text{if } u_t \leq -x_{t\cdot}\beta, \end{aligned}$$

the fitted model may "describe well" the first statement but poorly the second or vice versa. A useful statistic for the former would be the square of the simple correlation coefficient between predicted and actual y_t. Thus, e.g. suppose we follow our earlier convention about the numbering of observations; then for the positive component of the sample we put

$$\hat{y}_t = x_{t\cdot}\hat{\beta} + \hat{\sigma}\hat{\psi}_t, \qquad t = 1,2,\ldots,T_1. \tag{88}$$

An intuitively appealing statistic is

$$r^2 = \frac{\left[\sum_{t=1}^{T_1}(\hat{y}_t - \bar{\hat{y}})(y_t - \bar{y})\right]^2}{\left[\sum_{t=1}^{T_1}(\hat{y}_t - \bar{\hat{y}})^2\right]\left[\sum_{t=1}^{T_1}(y_t - \bar{y})^2\right]}, \tag{89}$$

where

$$\bar{y} = \frac{1}{T_1}\sum_{t=1}^{T_1} y_t, \qquad \bar{\hat{y}} = \frac{1}{T_1}\sum_{t=1}^{T_1}\hat{y}_t. \tag{90}$$

As to how well it discriminates between the zero and positive (dependent variable) observations we may compute $\Phi(-\nu_t)$ for all t; in the perfect discrimination case

$$\Phi(-\hat{\nu}_{t_2}) > \Phi(-\hat{\nu}_{t_1}), \qquad t_1 = 1,2,\dots,T_1, \qquad t_2 = T_1+1,\dots,T. \tag{91}$$

The relative frequency of the reversal of ranks would be another interesting statistic, as would the average probability difference, i.e.

$$\frac{1}{T_2}\sum_{t_2=T_1+1}^{T}\Phi(-\hat{\nu}_{t_2}) - \frac{1}{T_1}\sum_{t_1=1}^{T_1}\Phi(-\hat{\nu}_{t_1}) = d. \tag{92}$$

We have a "right" to expect as a minimum that

$$d > 0. \tag{93}$$

3. Sample selectivity

3.1. Generalities

This is another important class of problems that relate specifically to the issue of how observations on a given economic phenomenon are generated. More particularly, we hypothesize that whether a certain variable, say y_{t1}^*, *is observed or not depends on another variable, say* y_{t2}^*. Thus, the observability of y_{t1}^* depends on the probability structure of the stochastic process that generates y_{t2}^*, as well as on that of the stochastic process that governs the behavior of y_{t1}^*. The variable y_{t2}^* may be inherently unobservable although we assert that we know the variables that enter its "systematic part."

To be precise, consider the model

$$\begin{aligned} y_{t1}^* &= x_{t1\cdot}^*\beta_{\cdot 1}^* + u_{t1}^*, \\ & \qquad\qquad t = 1,2,\dots,T, \\ y_{t2}^* &= x_{t2\cdot}^*\beta_{\cdot 2}^* + u_{t2}^*, \end{aligned} \tag{94}$$

where $x_{t1\cdot}^*$, $x_{t2\cdot}^*$ are r_1, r_2-element row vectors of observable "exogenous" variables which may have elements in common. The vectors

$$u_{t\cdot}^{*} = \left(u_{t1}^{*}, u_{t2}^{*}\right), \qquad t = 1,2,\ldots,$$

form a sequence of i.i.d. random variables with distribution

$$u_{t\cdot}^{*\prime} \sim N(0, \Sigma^{*}), \qquad \Sigma^{*} > 0.$$

The variable y_{t2}^{*} is inherently unobservable, while y_{t1}^{*} is observable if and only if

$$y_{t1}^{*} \geq y_{t2}^{*}.$$

An example of such a model is due to Heckman (1979) where y_{t1}^{*} is an observed wage for the tth worker and y_{t2}^{*} is his reservation wage. Evidently, y_{t1}^{*} is the "market valuation" of his skills and other pertinent attributes, represented by the vector $x_{t1\cdot}^{*}$, while y_{t2}^{*} represents, through the vector $x_{t2\cdot}^{*}$ those personal and other relevant attributes that lead him to seek employment at a certain wage or higher.

Alternatively, in the market for housing y_{t1}^{*} would represent the "market valuation" of a given structure's worth while y_{t2}^{*} would represent the current owner's evaluation.

Evidently a worker accepts a wage for employment or a structure changes hands if and only if

$$y_{t1}^{*} \geq y_{t2}^{*}.$$

If the covariance matrix, Σ^{*}, is diagonal, then there is no correlation between y_{t1}^{*} and y_{t2}^{*} and hence in view of the assumption regarding the error process

$$\{u_{t\cdot}^{*\prime}: t = 1,2,\ldots\},$$

we could treat the sample

$$\{(y_{t1}^{*}, x_{t1\cdot}^{*}): t = 1,2,\ldots,T\},$$

as one of i.i.d. observations; consequently, we can estimate consistently the parameter vector $\beta_{\cdot 1}^{*}$ by OLS given the sample, irrespectively of the second relation in (94).

On the other hand, if the covariance matrix, Σ^{*}, is not diagonal, then the situation is far more complicated, since now there does exist a stochastic link between y_{t1}^{*} and y_{t2}^{*}. The question then becomes: If we apply OLS to the first equation in (94) do we suffer more than just the usual loss in efficiency?

3.2. Inconsistency of least squares procedures

In the current context, it would be convenient to state the problem in canonical form before we attempt further analysis. Thus, define

$$\begin{aligned} &y_{t1}=y_{t1}^{*}, \qquad y_{t2}=y_{t1}^{*}-y_{t2}^{*}, \qquad x_{t1\cdot}=x_{t1\cdot}^{*}, \qquad x_{t2\cdot}=\left(x_{t1\cdot}^{*},x_{t2\cdot}^{*}\right),\\ &\beta_{\cdot1}=\beta_{\cdot1}^{*}, \qquad \beta_{\cdot2}=\begin{pmatrix}\beta_{\cdot1}^{*}\\ -\beta_{\cdot2}^{*}\end{pmatrix}, \qquad u_{t1}=u_{t1}^{*}, \qquad u_{t2}=u_{t1}^{*}-u_{t2}^{*}, \end{aligned} \tag{95}$$

with the understanding that if $x_{t1\cdot}^{*}$ and $x_{t2\cdot}^{*}$ have elements in common, say,

$$x_{t1\cdot}^{*}=\left(z_{t1},z_{t1}^{*}\right), \qquad x_{t2\cdot}^{*}=\left(z_{t1\cdot},z_{t2\cdot}^{*}\right),$$

then

$$x_{t2\cdot}=\left(z_{t1\cdot},z_{t1\cdot}^{*},z_{t2\cdot}^{*}\right), \qquad \beta_{\cdot2}=\begin{pmatrix}\beta_{\cdot11}^{*}-\beta_{\cdot12}^{*}\\ \beta_{\cdot21}^{*}\\ -\beta_{\cdot22}^{*}\end{pmatrix}, \tag{96}$$

where $\beta_{\cdot11}^{*}, \beta_{\cdot12}^{*}$ are the coefficients of $z_{t1\cdot}$ in $x_{t1\cdot}^{*}$ and $x_{t2\cdot}^{*}$ respectively, β_{21}^{*} is the coefficient of $z_{t1\cdot}^{*}$ and $\beta_{\cdot22}^{*}$ is the coefficient of $z_{t2\cdot}^{*}$.

Hence, the model in (94) can be stated in the canonical form

$$\begin{cases} y_{t1}=x_{t1\cdot}\beta_{\cdot1}+u_{t1},\\ y_{t2}=x_{t2\cdot}\beta_{\cdot2}+u_{t2}, \end{cases} \tag{97}$$

such that $x_{t2\cdot}$ contains at least as many elements as $x_{t1\cdot}$,

$$\left\{u_{t\cdot}'=\left(u_{t1},u_{t2}\right)':\ t=1,2,\dots\right\},$$

is a sequence of i.i.d. random variables with distribution

$$u_{t\cdot}'\sim N(0,\Sigma), \qquad \Sigma>0,$$

and subject to the condition that y_{t1} is observable (observed) if and only if $y_{t2}\geq 0$.

If we applied OLS methods to the first equation in (97) do we obtain, at least, consistent estimators for its parameters? The answer hinges on whether that question obeys the standard assumptions of the GLM.

Clearly, and solely in terms of the system in (97),

$$\{u_{t1}:\ t=1,2,\dots\}, \tag{98}$$

is a sequence of i.i.d. random variables and if in (94) we are prepared to assert that the standard conditions of the typical GLM hold, nothing in the subsequent discussion suggests a correlation between $x_{t1\cdot}$ and u_{t1}; hence, if any problem should arise it ought to be related to the probability structure of the sequence in (98) insofar as it is associated with observable y_{t1} – a problem to which we now turn. We note that the conditions hypothesized by the model imply that (potential) realizations of the process in (98) are conditioned on[4]

$$u_{t2} \geq -x_{t2\cdot}\beta_{\cdot 2}. \tag{99}$$

Or, perhaps more precisely, we should state that (implicit) realizations of the process in (98) associated with *observable realizations*

$$\{y_{t1}: t = 1, 2, \dots\},$$

are conditional on (99). Therefore, in dealing with the error terms of (potential) samples the marginal distribution properties of (98) are not relevant; what *are relevant* are its conditional properties – as conditioned by (99).

We have

Lemma 3

The distribution of realizations of the process in (98) as conditioned by (99) has the following properties:

i. The elements $\{u_{t1}, u_{t2}\}$ are mutually independent for $t \neq t'$.

ii. The density of u_{t1}, given that the corresponding y_{t1} is observable (observed) is

$$f(u_{t1} | u_{t2} > -x_{t2\cdot}\beta_{\cdot 2}) = \frac{\Phi(\pi_t)}{\Phi(\nu_{t2})} \frac{1}{\sqrt{2\pi\sigma_{11}}} \exp - \frac{1}{2\sigma_{11}} u_{t1}^2, \tag{100}$$

where

$$\begin{aligned} \nu_{t2} &= \frac{x_{t2\cdot}\beta_{\cdot 2}}{\sigma_{22}^{1/2}}, \qquad \pi_t = \frac{1}{\alpha^{1/2}} \left(\nu_{t2} + \frac{\rho_{12}}{\sigma_{11}^{1/2}} u_{t1} \right), \\ \rho_{12}^2 &= \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} \qquad \alpha = 1 - \rho_{12}^2, \end{aligned} \tag{101}$$

and $\Phi(\cdot)$ is the c.d.f. of a $N(0,1)$.

[4] Note that in terms of the original variables (99) reads

$$u_{t1}^* \geq u_{2t}^* + x_{t2\cdot}^* \beta_{\cdot 2}^* - x_{t1\cdot}^* \beta_{\cdot 1}^*.$$

We shall not use this fact in subsequent discussion, however.

Proof

i. is quite evidently valid since by the standard assumptions of the GLM we assert that $(x^*_{t1\cdot}, x^*_{t2\cdot})$ and $u^*_{t\cdot} = (u^*_{t1}, u^*_{t2})$ are mutually independent and that

$$\{u^{*\prime}_{t\cdot}: t = 1,2,\dots\},$$

is a sequence of i.i.d. random variables.

As for part ii. we begin by noting that since the conditional density of u_{t1} given u_{t2} is given by

$$u_{t1}|u_{t2} \sim N\left(\frac{\sigma_{12}}{\sigma_{22}}u_{t2}, \sigma_{11}\alpha\right),$$

and since the restriction in (99) restricts us to the space

$$u_{t2} \geq -x_{t2\cdot}\beta_{\cdot 2},$$

the required density can be found as

$$f(u_{t1}|u_{t2} \geq -x_{t2\cdot}\beta_{\cdot 2}) = \frac{1}{\Phi(\nu_{t2})}\frac{1}{\sqrt{2\pi\alpha\sigma_{11}}}\frac{1}{\sqrt{2\pi\sigma_{22}}}$$

$$\cdot\int_{-x_{t2\cdot}\beta_{\cdot 2}}^{\infty}\exp -\frac{1}{2\sigma_{11}}\left(u_{t1} - \frac{\sigma_{12}}{\sigma_{22}}\xi\right)^2 \exp -\frac{1}{2\sigma_{22}}\xi^2\,\mathrm{d}\xi.$$

Completing the square (in ξ) and making the change in variable

$$\zeta = \left(\xi - \frac{\sigma_{12}}{\sigma_{11}}u_{t1}\right)/(\sigma_{22}\alpha)^{1/2},$$

we find

$$f(u_{t1}|u_{t2} \geq -x_{t2\cdot}\beta_{\cdot 2}) = \frac{\Phi(\pi_t)}{\Phi(\nu_{t2})}\frac{1}{\sqrt{2\pi\sigma_{11}}}\exp -\frac{1}{2\sigma_{11}}u_{t1}^2.$$

To verify that this is, indeed, a density function we note that it is everywhere nonnegative and

$$\int_{-\infty}^{\infty} f(\xi_1|u_{t2} \geq -x_{t2\cdot}\beta_{\cdot 2})\,\mathrm{d}\xi_1$$

$$= \frac{1}{\Phi(\nu_{t2})}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_{11}}}\left[\int_{-\infty}^{\pi_t}\frac{1}{\sqrt{2\pi}}\exp -\frac{1}{2}\xi_2^2\,\mathrm{d}\xi_2\right]\cdot\exp -\frac{1}{2\sigma_{11}}\xi_1^2\,\mathrm{d}\xi_1.$$

Making the transformation

$$\zeta_1 = \frac{1}{\sigma_{11}^{1/2}}\xi_1, \qquad \zeta_2 = \alpha^{1/2}\xi_2 - \rho_{12}\zeta_1,$$

the integral is reduced to

$$\frac{1}{\Phi(\nu_{t2})}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\alpha}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\nu_{t2}}\exp-\frac{1}{2\alpha}(\zeta_2+\rho_{12}\zeta_1)^2\exp-\tfrac{1}{2}\zeta_1^2\,\mathrm{d}\zeta_2\,\mathrm{d}\zeta_1$$
$$=\frac{1}{\Phi(\nu_{t2})}\int_{-\infty}^{\nu_{t2}}\frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\alpha}}\exp-\frac{1}{2\alpha}(\zeta_1+\rho_{12}\zeta_2)^2\,\mathrm{d}\zeta_1\right]$$
$$\cdot\exp-\tfrac{1}{2}\zeta_2^2\,\mathrm{d}\zeta_2=1.$$

Q.E.D.

Lemma 4

The kth moment of realizations of the process in (98) corresponding to observable realizations $\{y_{t1}:\ t=1,2,\dots\}$ is given, for k even ($k=2,4,6,\dots$), by

$$I_{k,t}=\sigma_{11}(k-1)I_{k-2,t}-\sigma_{11}^{k/2}\alpha^{(k-2)/2}\rho_{12}^2\nu_{t2}\psi(\nu_{t2})\sum_{r=0}^{(k-2)/2}\binom{k-1}{2r+1}\left(\frac{\rho_{12}^2\nu_{t2}^2}{\alpha}\right)^r \cdot\frac{\left[2\left(\frac{k-2}{2}-r\right)\right]!}{2^{\frac{k-2}{2}-r}\left(\frac{k-2}{2}-r\right)!},\tag{102}$$

while for k odd ($k=3,5,7,\dots$) it is given by

$$I_{k,t}=\sigma_{11}(k-1)I_{k-2,t}+\sigma_{11}^{k/2}\alpha^{(k-1)/2}\rho_{12}\psi(\nu_{t2})\sum_{r=0}^{(k-1)/2}\binom{k-1}{2r}\left(\frac{\rho_{12}^2\nu_{t2}^2}{\alpha}\right)^r \cdot\frac{\left[2\left(\frac{k-1}{2}-r\right)\right]!}{2^{\frac{k-1}{2}-r}\left(\frac{k-1}{2}-r\right)!},\tag{103}$$

where

$$\psi(\nu_{t2})=\frac{\phi(\nu_{t2})}{\Phi(\nu_{t2})},\qquad I_{0,t}=1,\qquad I_{1,t}=\sigma_{11}^{1/2}\rho_{12}\psi(\nu_{t2}).\tag{104}$$

Remark 5

It is evident, from the preceding discussion, that the moments of the error process corresponding to observable y_{t1} are uniformly bounded in $\beta_{\cdot 1}$, $\beta_{\cdot 2}$, σ_{11}, σ_{12}, σ_{22}, $x_{t1\cdot}$ and $x_{t2\cdot}$ – provided the parameter space is compact and the elements of $x_{t1\cdot}$, $x_{t2\cdot}$ are bounded.

Remark 6

It is also evident from the preceding that for (potential) observations from the model

$$y_{t1} = x_{t1\cdot}\beta_{\cdot 1} + u_{t1},$$

we have that

$$E(y_{t1}|x_{t1\cdot}) = x_{t1\cdot}\beta_{\cdot 1} + \sigma_{11}^{1/2}\rho_{12}\psi(\nu_{t2}). \tag{105}$$

We are now in a position to answer the question, raised earlier, whether OLS methods applied to the first equation in (97) will yield at least consistent estimators. In this connection we observe that the error terms of *observations* on the first equation of (97) obey

$$E(u_{t1}|u_{t2} \geq -x_{t2\cdot}\beta_{\cdot 2}) = I_{1t} = \sigma_{11}^{1/2}\rho_{12}\psi(\nu_{t2}),$$

$$\begin{aligned}\operatorname{Var}(u_{t1}|u_{t2} \geq -x_{t2\cdot}\beta_{\cdot 2}) = I_{2t} - I_{1t}^2 &= \sigma_{11} - \sigma_{11}\rho_{12}^2\nu_{t2}\psi(\nu_{t2}) \\ &\quad - \sigma_{11}\rho_{12}^2\psi^2(\nu_{t2}) \\ &= \sigma_{11} - \sigma_{11}\rho_{12}^2\psi(\nu_{t2})[\nu_{t2} + \psi(\nu_{t2})].\end{aligned}$$

As is well known, the second equation shows the errors to be *heteroskedastic* – whence we conclude that OLS estimators *cannot be efficient*. The first equation above shows the errors to have a nonzero mean. As shown in Dhrymes (1978a) a nonconstant (nonzero) mean implies misspecification due to left out variables and *hence inconsistency*.

Thus, OLS estimators are inconsistent; hence, we must look to other methods for obtaining suitable estimators for $\beta_{\cdot 1}$, σ_{11}, etc. On the other hand, *if*, *in* (105), $\rho_{12} = 0$, *then OLS estimators would be* consistent but inefficient.

3.3. *The LF and ML estimation*

We shall assume that in our sample we have entities for which y_{t1} is observed and entities for which it is not observed; if y_{t1} is not observable, then we know that

$y_{t2} < 0$, hence that

$$u_{t2} < -x_{t2\cdot}\beta_{\cdot 2}.$$

Consequently, the probability attached to that event is

$$\Phi(-\nu_{t2}).$$

Evidently, the probability of observing y_{t1} is $\Phi(\nu_{t2})$ and given that y_{t1} is observed the probability it will assume a value in some internal Δ is

$$\frac{1}{\Phi(\nu_{t2})}\frac{1}{\sqrt{2\pi\sigma_{11}}}\int_{\Delta}\Phi(\pi_t)\exp-\frac{1}{2\sigma_{11}}\xi^2\,\mathrm{d}\xi.$$

Hence, the unconditional probability that y_{t1} will assume a value in the interval Δ is

$$\frac{1}{\sqrt{2\pi\sigma_{11}}}\int_{\Delta}\Phi(\pi_t)\exp-\frac{1}{2\sigma_{11}}\xi^2\,\mathrm{d}\xi.$$

Define

$$\begin{aligned} c_t &= 1 \quad \text{if } y_{t1} \text{ is observed,} \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

and note that the LF is given by

$$L^* = \prod_{t=1}^{T}\left[\Phi(\nu_{t2})f(y_{t1}-x_{t1\cdot}\beta_{\cdot 1}|u_{t2}\geq -x_{t2\cdot}\beta_{\cdot 2})\right]^{c_t}\left[\Phi(-\nu_{t2})\right]^{1-c_t}. \tag{106}$$

Thus, e.g. if for a given sample we have no observations on y_{t1} the LF becomes

$$\prod_{t=1}^{T}\Phi(-\nu_{t2}),$$

while, if all sample observations involve observable y_{t1}'s the LF becomes

$$\prod_{t=1}^{T}\left\{\frac{1}{\sqrt{2\pi\sigma_{11}}}\Phi\left[\frac{1}{\alpha^{1/2}}\left(\nu_{t2}+\rho_{12}\left(\frac{y_{t1}-x_{t1\cdot}\beta_{\cdot 1}}{\sigma_{11}^{1/2}}\right)\right)\right]\exp-\frac{1}{2\sigma_{11}}(y_{t1}-x_{t1\cdot}\beta_{\cdot 1})^2\right\}.$$

Finally, if the sample contains entities for which y_{t1} is observed as well as entities for which it is not observed, then we have the situation in (106). We shall examine the estimation problems posed by (106) in its general form.

Remark 7

It is evident that we can parametrize the problem in terms of $\beta_{\cdot 1}, \beta_{\cdot 2}, \sigma_{11}, \sigma_{22}, \sigma_{12}$; it is further evident that $\beta_{\cdot 2}$ and σ_{22} appear only in the form $(\beta_{\cdot 2}/\sigma_{22}^{1/2})$ – hence, that σ_{22} cannot be, separately, identified. We shall, thus, adopt the convention

$$\sigma_{22} = 1. \tag{107}$$

A consequence of (107) is that (105) reduces to

$$E(y_{t1}|x_{t1\cdot}, u_{t2} \geq -x_{t2\cdot}\beta_{\cdot 2}) = x_{t1\cdot}\beta_{\cdot 1} + \sigma_{12}\psi(\nu_{t2}). \tag{108}$$

The logarithm of the LF is given by

$$L = \sum_{t=1}^{T} \Bigg\{ (1-c_t)\ln\Phi(-\nu_{t2}) + c_t\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma_{11} - \frac{1}{2\sigma_{11}}(y_{t1} - x_{t1\cdot}\beta_{\cdot 1})^2\right] + \ln\Phi\left[\frac{1}{\alpha^{1/2}}\left(\nu_{t2} + \rho_{12}\left(\frac{y_{t1} - x_{t1\cdot}\beta_{\cdot 1}}{\sigma_{11}^{1/2}}\right)\right)\right]\Bigg\}. \tag{109}$$

Remark 8

We shall proceed to maximize (109) treating $\beta_{\cdot 1}, \beta_{\cdot 2}$ as free parameters. As pointed out in the discussion following eq. (95) the two vectors will, generally, have elements in common. While we shall ignore this aspect here, for simplicity of exposition, we can easily take account of it by considering as the vector of unknown parameters γ whose elements are the distinct elements of $\beta_{\cdot 1}, \beta_{\cdot 2}$ and σ_{11}, ρ_{12}.

The first order conditions yield

$$\frac{\partial L}{\partial \beta_{\cdot 1}} = \frac{1}{\sigma_{11}^{1/2}} \sum_{t=1}^{T} c_t \left[\frac{y_{t1} - x_{t1\cdot}\beta_{\cdot 1}}{\sigma_{11}^{1/2}} - \frac{\rho_{12}}{\alpha^{1/2}}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\right] x_{t1\cdot}, \tag{110}$$

$$\frac{\partial L}{\partial \beta_{\cdot 2}} = \sum_{t=1}^{T} \left[c_t \frac{\phi(\pi_t)}{\Phi(\pi_t)}\frac{1}{\alpha^{1/2}} - (1-c_t)\frac{\phi(\nu_{t2})}{\Phi(-\nu_{t2})}\right] x_{t2\cdot}, \tag{111}$$

$$\frac{\partial L}{\partial \sigma_{11}} = \frac{1}{2\sigma_{11}} \sum_{t=1}^{T} c_t \left[-1 + \left(\frac{y_{t1} - x_{t1\cdot}\beta_{\cdot 1}}{\sigma_{11}^{1/2}}\right)^2 - \frac{\phi(\pi_t)}{\Phi(\pi_t)}\frac{\rho_{12}}{\alpha^{1/2}}\left(\frac{y_{t1} - x_{t1\cdot}\beta_{\cdot 1}}{\sigma_{11}^{1/2}}\right)\right], \tag{112}$$

$$\frac{\partial L}{\partial \rho_{12}} = \frac{1}{\alpha^{3/2}} \sum_{t=1}^{T} c_t \frac{\phi(\pi_t)}{\Phi(\pi_t)}\left[\rho_{12}\nu_{t2} + \left(\frac{y_{t1} - x_{t1\cdot}\beta_{\cdot 1}}{\sigma_{11}^{1/2}}\right)\right]. \tag{113}$$

Putting

$$\gamma = \left(\beta_{\cdot 1}', \beta_{\cdot 2}', \sigma_{11}, \sigma_{12}\right)', \tag{114}$$

we see that the ML estimator, say $\hat{\gamma}$, is defined by the condition

$$\frac{\partial L}{\partial \gamma}(\hat{\gamma}) = 0. \tag{115}$$

Evidently, this is a highly nonlinear set of relationships which can be solved only by iteration, from an initial consistent estimator, say $\tilde{\gamma}$.

3.4. *An initial consistent estimator*

To obtain an initial consistent estimator we look at the sample solely from the point of view of whether information is available on y_{t1}, i.e. whether y_{t1} is observed with respect to the economic entity in question. It is clear that this, at best, will identify only $\beta_{\cdot 2}$, since absent any information on y_{t1} we cannot possibly hope to estimate $\beta_{\cdot 1}$. Having estimated $\beta_{\cdot 2}$ by this procedure we proceed to construct the variable

$$\tilde{\psi}_t = \tilde{\psi}_t(\tilde{\nu}_{t2}) = \frac{\phi(x_{t2\cdot}\tilde{\beta}_{\cdot 2})}{\Phi(x_{t2\cdot}\tilde{\beta}_{\cdot 2})}, \qquad t = 1, 2, \ldots, T. \tag{116}$$

Then, turning our attention to that part of the sample which contains observations on y_{t1}, we simply regress y_{t1} on $(x_{t1\cdot}, \tilde{\psi}_t)$. In this fashion we obtain estimators of

$$\delta = \left(\beta_{\cdot 1}', \sigma_{12}\right)' \tag{117}$$

as well as of σ_{11}.

Examining the sample from the point of view first set forth at the beginning of this section we have the log likelihood function

$$L_1 = \sum_{t=1}^{T} \left[c_t \ln \Phi(\nu_{t2}) + (1 - c_t) \ln \Phi(-\nu_{t2})\right], \tag{118}$$

which is to be maximized with respect to the unknown vector $\beta_{\cdot 2}$. In Section 1.d. we noted that L_1 is strictly concave with respect to $\beta_{\cdot 2}$; moreover, the matrix of its second order derivatives is given by

$$\frac{\partial^2 L_1}{\partial \beta_{\cdot 2} \partial \beta_{\cdot 2}} = -\sum_{t=1}^{T} \phi(x_{t2\cdot}\beta_{\cdot 2}) \left[c_t \frac{S_1(x_{t2\cdot}\beta_{\cdot 2})}{\Phi^2(x_{t2\cdot}\beta_{\cdot 2})} + (1-c_t) \frac{S_2(x_{t2\cdot}\beta_{\cdot 2})}{\Phi^2(-x_{t2\cdot}\beta_{\cdot 2})} \right] \cdot x'_{t2\cdot} x_{t2\cdot} \tag{119}$$

where

$$S_1(x_{t2\cdot}\beta_{\cdot 2}) = \phi(x_{t2\cdot}\beta_{\cdot 2}) + (x_{t2\cdot}\beta_{\cdot 2})\Phi(x_{t2\cdot}\beta_{\cdot 2}), \tag{120}$$

$$S_2(x_{t2\cdot}\beta_{\cdot 2}) = \phi(x_{t2\cdot}\beta_{\cdot 2}) - (x_{t2\cdot}\beta_{\cdot 2})\Phi(-x_{t2\cdot}\beta_{\cdot 2}). \tag{121}$$

It is also shown in the discussion of Section 1.d. that asymptotically

$$\sqrt{T}(\hat{\beta}_{\cdot 2} - \beta_2^0) \sim N\left(0, -\lim_{T\to\infty} \left\{ \frac{1}{T} E\left[\frac{\partial^2 L}{\partial \beta_{\cdot 2} \partial \beta_{\cdot 2}} (\beta_{\cdot 2}^0) \right] \right\}^{-1} \right), \tag{122}$$

where $\hat{\beta}_{\cdot 2}$ is the ML estimator, i.e. it solves

$$\frac{\partial L_1}{\partial \beta_{\cdot 2}}(\hat{\beta}_{\cdot 2}) = 0, \tag{123}$$

and $\beta_{\cdot 2}^0$ is the true parameter point. It is evident from (122) that $\hat{\beta}_{\cdot 2}$ converges a.c. to $\beta_{\cdot 2}^0$. Define now

$$\tilde{\psi}_t = \frac{\phi(x_{t2\cdot}\hat{\beta}_{\cdot 2})}{\Phi(x_{t2\cdot}\hat{\beta}_{\cdot 2})}, \qquad t = 1, 2, \ldots, T, \tag{124}$$

and consider the estimator

$$\tilde{\delta} = (X_1^{*\prime} X_1^*)^{-1} X_1^{*\prime} y_{\cdot 1}, \qquad \delta = (\beta'_{\cdot 1}, \sigma_{12})', \tag{125}$$

where we have written

$$y_{t1} = x_{t1\cdot}\beta_1 + \sigma_{12}\psi_t + v_{t1}, \qquad v_{t1} = u_{t1} - \sigma_{12}\psi_t, \tag{126}$$

$$X_1^* = (X_1, \tilde{\psi}), \qquad X_1 = (x_{t1\cdot}), \qquad \tilde{\psi} = (\tilde{\psi}_t), \qquad t = 1, 2, \ldots, T. \tag{127}$$

We observe that

$$(\tilde{\delta} - \delta^0) = (X_1^{*\prime} X_1^*)^{-1} X_1^{*\prime} [v_{\cdot 1} - \sigma_{12}(\tilde{\psi} - \psi)]. \tag{128}$$

It is our contention that the estimator in (125) is consistent for $\beta_{\cdot 1}$ and σ_{12}; moreover that it naturally implies a consistent estimator for σ_{11}, thus yielding the initial consistent estimator, say

$$\tilde{\gamma} = \left(\tilde{\beta}'_{\cdot 1}, \hat{\beta}'_{\cdot 2}, \tilde{\sigma}_{11}, \tilde{\rho}_{12}\right)' \tag{129}$$

which we require for obtaining the LM estimator.

Formally, we will establish that

$$\sqrt{T}(\tilde{\delta} - \delta^0) = \left(\frac{X_1^{*\prime} X_1^*}{T}\right) \frac{1}{\sqrt{T}} X_1^{*\prime}\left[v_{\cdot 1} - \sigma_{12}(\tilde{\psi} - \psi)\right] \sim N(0, F), \tag{130}$$

for suitable matrix F, thus showing that $\tilde{\delta}$ converges to δ^0 with probability one (almost surely).

In order that we may accomplish this task it is imperative that we must specify more precisely the conditions under which we are to consider the model[5] in (94), as expressed in (97). We have:

(A.3.1.) The basic error process

$$\left\{u'_{t\cdot}: t = 1, 2, \dots\right\}, \qquad u_{t\cdot} = (u_{t1}, u_{t2}),$$

is one of i.i.d. random variables with

$$u'_{t\cdot} \sim N(0, \Sigma), \qquad \Sigma > 0, \qquad \sigma_{22} = 1,$$

and is independent of the process generating the exogenous variables $x_{t1\cdot}, x_{t2\cdot}$.

(A.3.2.) The admissible parameter space, say $H \subset R_{n+3}$, is closed and bounded and contains an open neighborhood of the true parameter point

$$\gamma^0 = \left(\beta^0_{\cdot 1}, \beta^{0\prime}_{\cdot 2}, \sigma^0_{11}, \rho^0_{12}\right)'.$$

(A.3.3.) The exogenous variables are nonstochastic and are bounded, i.e.

$$|x_{t2i}| < k_i, \qquad i = 0, 1, 2, \dots n$$

for all t.[6]

[5]As pointed out earlier, it may be more natural to state conditions in terms of the basic variables $x^*_{t1\cdot}$, $x^*_{t2\cdot}$, u^*_{t1} and u^*_{t2}; doing so, however, will disrupt the continuity of our discussion; for this reason we state conditions in terms of $x_{t1\cdot}$, $x_{t2\cdot}$, $u_{t1\cdot}$ and $u_{t2\cdot}$.

(A.3.4.) The matrix

$$X_2 = (x_{t2\cdot}) \qquad t = 1, 2, \ldots, T,$$

is of rank $n+1$ and moreover

$$\lim_{T} \frac{1}{T} X_2' X_2 = P, \qquad P > 0.$$

Remark 9

It is a consequence of the assumptions above that, for any $x_{t2\cdot}$ and admissible $\beta_{\cdot 2}$, there exists k such that

$$-r \leq x_{t2\cdot}\beta_{\cdot 2} \leq r, \qquad 0 < r < k, \qquad k < \infty,$$

so that, for example,

$$\begin{aligned}
&\phi(x_{t2\cdot}\beta_{\cdot 2}) > \phi(-k) > 0,\\
&\Phi(x_{t2\cdot}\beta_{\cdot 2}) < \Phi(k) < 1, \qquad\qquad (131)\\
&\Phi(x_{t2\cdot}\beta_{\cdot 2}) > \Phi(-k) > 0.
\end{aligned}$$

Consequently,

$$\psi(\nu_t) = \frac{\phi(x_{t2\cdot}\beta_{\cdot 2})}{\Phi(x_{t2\cdot}\beta_{\cdot 2})}, \qquad \psi^*(\nu_t) = \frac{\phi(x_{t2\cdot}\beta_{\cdot 2})}{\Phi(-x_{t2\cdot}\beta_{\cdot 2})},$$

are both bounded continuous functions of their argument.

To show the validity of (130) we proceed by a sequence of Lemmata.

Lemma 5

The probability limit of the matrix to be inverted in is given by

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} X_1^{*\prime} X_1^* = \lim_{T\to\infty} \frac{1}{T} X_1^{0\prime} X_1^0 = Q_0, \qquad Q_0 > 0,$$

[6] We remind the reader that in the canonical representation of (97), the vector $x_{t1\cdot}$ is a subvector of $x_{t2\cdot}$; hence the boundedness assumptions on $x_{t2\cdot}$ imply similar boundedness conditions on $x_{t1\cdot}$.
Incidentally, note that $\beta_{\cdot 1}^0$ is not necessarily a subvector of $\beta_{\cdot 2}^0$, since the latter would contain $\beta_{\cdot 11}^* - \beta_{\cdot 12}^{*0}$ and in addition $\beta_{\cdot 21}^{*0}$, $-\beta_{\cdot 22}^{*0}$, while the former will contain $\beta_{\cdot 11}^{*0}$, $\beta_{\cdot 21}^*$.

where

$$X_1^0 = (X_1, \psi^0), \qquad \psi^0 = (\psi_t^0), \qquad \psi_t^0 = \frac{\phi(x_{t2\cdot}\beta_{\cdot 2}^0)}{\Phi(x_{t2\cdot}\beta_{\cdot 2}^0)}.$$

Proof

We examine

$$S_T = \frac{1}{T}\left[X_1^{*\prime}X_1^* - X_1^{0\prime}X_1^0\right] = \frac{1}{T}\begin{bmatrix} 0 & X_1'(\tilde{\psi} - \psi^0) \\ (\tilde{\psi} - \psi^0)'X_1 & (\tilde{\psi} + \psi^0)(\tilde{\psi} - \psi^0) \end{bmatrix}, \tag{132}$$

and the problem is reduced to considering

$$\tilde{\psi}_t - \psi_t^0 = \alpha_t^0 x_{t2\cdot}(\hat{\beta}_{\cdot 2} - \beta_{\cdot 2}^0) + s_t^*(\hat{\beta}_{\cdot 2} - \beta_{\cdot 2}^0)'x_{t2\cdot}'x_{t2\cdot}(\hat{\beta}_{\cdot 2} - \beta_{\cdot 2}^0), \tag{133}$$

where

$$\alpha_t^0 = \frac{\partial \psi(\nu_{t2})}{\partial \nu_{t2}} \qquad \text{evaluated at } \beta_{\cdot 2} = \beta_{\cdot 2}^0,$$

$$s_t^* = \frac{\partial^2 \psi(\nu_{t2})}{\partial \nu_{t2}^2} \qquad \text{evaluated at } \beta_{\cdot 2} = \beta_{\cdot 2}^*,$$

$$|\beta_{\cdot 2}^* - \beta_{\cdot 2}^0| < |\beta_{\cdot 2} - \beta_{\cdot 2}^0|.$$

It is evident that, when the expansion in (133) is incorporated in (132) quadratic terms in $(\hat{\beta}_{\cdot 2} - \beta_{\cdot 2}^0)$ will vanish with T.

Hence we need be concerned only with the terms of the form

$$\frac{1}{T}\sum_{t=1}^{T} x_{t1\cdot}'(\tilde{\psi}_t - \psi_t^0) \sim \frac{1}{T^{3/2}}\sum_{t=1}^{T}\left[\alpha_t^0 x_{t1\cdot}'x_{t2}\right]\sqrt{T}(\hat{\beta}_{\cdot 2} - \beta_{\cdot 2}^0),$$

or of the form

$$\frac{1}{T}\sum_{t=1}^{T}(\tilde{\psi}_t + \psi_t^0)(\tilde{\psi}_t - \psi_t^0) \sim \frac{1}{T^{3/2}}\sum_{t=1}^{T}\left[\alpha_t^0(\tilde{\psi}_t + \psi_t^0)x_{t2\cdot}\right]\sqrt{T}(\hat{\beta}_{\cdot 2} - \beta_{\cdot 2}^0).$$

In either case we note that by assumption (A.3.4.) and Remark 9

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \alpha_t^0 x'_{t1\cdot} x_{t2\cdot},$$

has bounded elements; similarly, for

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \alpha_t^0 (\tilde{\psi}_t + \psi_t^0) x_{t2\cdot}.$$

Consequently, in view of (122) and (132) we conclude

$$\operatorname*{plim}_{T\to\infty} S_T = 0,$$

which implies

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} X_1^{*\prime} X_1^* = \lim_{T\to\infty} \frac{1}{T} X_1^{0\prime} X_1^0 = Q_0. \tag{134}$$

Corollary 4

The limiting distribution of the left member of (130) is obtainable through

$$\sqrt{T}(\tilde{\delta} - \delta^0) \sim Q_0^{-1} X_1^{*\prime}\left[v_{\cdot 1} - \sigma_{12}(\tilde{\psi} - \psi^0)\right], \; v_{\cdot 1} = (v_{11}, v_{21} \cdots v_{T1})'.$$

Indeed, by standard argumentation we may establish

Theorem 4

Under assumption (A.3.1) through (A.3.4) the initial (consistent) estimator of this section has the limiting distribution

$$\sqrt{T}(\tilde{\delta} - \delta^0) \sim N(0, F), \qquad F = Q_0^{-1} P Q_0^{-1},$$

where

$$P = \sigma_{11} \lim_{T\to\infty} \frac{1}{T} \begin{bmatrix} \sum_{t=1}^{T} \omega_{11t} x'_{t1\cdot} x_{t1\cdot} & \sum_{t=1}^{T} \omega_{11t} \psi_t^0 x'_{t1\cdot} \\ \sum_{t=1}^{T} \omega_{11t} \psi_t^0 x_{t1\cdot} & \sum_{t=1}^{T} \omega_{11t} \psi_t^{0^2} \end{bmatrix}$$

Q_0 is defined in (134) and

$$E(v_{t1}^2) = \sigma_{11}\omega_{11t} = \sigma_{11}\left[1 - \rho_{12}^{0^2}\nu_{t2}^0\psi_t^0 - \rho_{12}^{0^2}\psi_t^{0^2}\right].$$

Corollary 5

The initial estimator above is strongly consistent.

Proof

From the theorem above

$$\sqrt{T}(\tilde{\delta} - \delta^0) \sim \zeta,$$

where ζ is an a.c. finite random vector.

Thus

$$\tilde{\delta} - \delta^0 \sim \frac{1}{\sqrt{T}}\zeta,$$

$\tilde{\delta}$ converges to δ^0 a.c.

Evidently, the parameter σ_{11} can be estimated (at least consistently) by

$$\tilde{\sigma}_{11} = \frac{1}{T}\left[\tilde{v}_{t1}^2 + \tilde{\sigma}_{12}\tilde{\psi}_t\tilde{v}_t 2 + \tilde{\sigma}_{12}\tilde{\psi}_t^2\right].$$

3.5. Limiting distribution of the ML estimator

In the previous section we outlined a procedure for obtaining an initial estimator, say

$$\tilde{\gamma} = (\tilde{\beta}_1, \tilde{\beta}_2', \tilde{\sigma}_{11}, \tilde{\sigma}_{12})',$$

and have shown that it converges to the true parameter point, say γ^0, with probability one (a.c.).

We now investigate the properties of the ML estimator, say $\hat{\gamma}$, obtained by solving

$$\frac{\partial L}{\partial \gamma}(\hat{\gamma}) = 0$$

through iteration, beginning with $\tilde{\gamma}$. The limiting distribution of $\hat{\gamma}$ may be found by examining

$$\sqrt{T}(\hat{\gamma}-\gamma^0)=\left[-\frac{1}{T}\frac{\partial^2 L}{\partial\gamma\partial\gamma}(\gamma^*)\right]^{-1}\frac{1}{\sqrt{T}}\frac{\partial L}{\partial\gamma}(\gamma^0), \tag{135}$$

where γ^0 is the true parameter point, γ^* obeys

$$|\gamma^*-\gamma^0|\leq|\hat{\gamma}-\gamma^0|,$$

and

$$\frac{\partial L}{\partial\gamma'}(\gamma^0)=\sum_{t=1}^{T}A_t\xi_{\cdot t}, \tag{136}$$

where

$$A_t=\begin{bmatrix}\frac{1}{\sigma_{11}^{1/2}}x'_{t1\cdot} & 0 & 0 & 0\\ 0 & x'_{t2\cdot} & 0 & 0\\ 0 & 0 & \frac{1}{2\sigma_{11}} & 0\\ 0 & 0 & 0 & \frac{1}{\alpha^{3/2}}\end{bmatrix} \tag{137}$$

$$\begin{aligned}\xi_{1t}&=c_t\left[\frac{u_{t1}}{\sigma_{11}^{1/2}}-\frac{\rho_{12}}{\alpha^{1/2}}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\right],\\ \xi_{2t}&=c_t\left[\frac{1}{\alpha^{1/2}}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\right]-(1-c_t)\frac{\phi(\nu_{t2})}{\Phi(\nu_{t2})},\\ \xi_{3t}&=c_t\left[\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)^2-\frac{\rho_{12}}{\alpha^{1/2}}\frac{\phi(\pi_t)}{\phi(\pi_t)}\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)-1\right],\\ \xi_{4t}&=c_t\frac{\phi(\pi_t)}{\Phi(\pi_t)}\left[\rho_{12}\nu_{t2}+\frac{u_{t1}}{\sigma_{11}^{1/2}}\right],\\ \xi_{\cdot t}&=(\xi_{1t},\xi_{2t},\xi_{3t},\xi_{4t})'.\end{aligned} \tag{138}$$

In order to complete the problem we also require an expression for the Hessian of the likelihood function in addition to the expressions in (110) through (113).

To this effect define

$$
\begin{aligned}
\xi_{11t} &= c_t\left[1+\frac{\rho_{12}^2}{\alpha}\pi_t\frac{\phi(\pi_t)}{\Phi(\pi_t)}+\frac{\rho_{12}^2}{\alpha}\frac{\phi^2(\pi_t)}{\Phi^2(\pi_t)}\right],\\
\xi_{21t} &= -c_t\left[\frac{\rho_{12}}{\alpha}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\pi_t+\frac{\rho_{12}}{\alpha}\frac{\phi^2(\pi_t)}{\Phi^2(\pi_t)}\right],\\
\xi_{31t} &= c_t\left[2\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)-\frac{\rho_{12}}{\alpha}\frac{\phi(\pi_t)}{\Phi(\pi_t)}+\frac{\rho_{12}^2}{\alpha}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\pi_t\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)\right.\\
&\quad\left.+\frac{\rho_{12}^2}{\alpha}\frac{\phi^2(\pi_t)}{\Phi^2(\pi_t)}\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)\right],\\
\xi_{41t} &= c_t\left[\frac{\phi(\pi_t)}{\Phi(\pi_t)}-\frac{\rho_{12}}{\alpha^{1/2}}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\pi_t\left(\rho_{12}\nu_{t2}+\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)\right.\\
&\quad\left.-\frac{\rho_{12}}{\alpha^{1/2}}\frac{\phi^2(\pi_t)}{\Phi^2(\pi_t)}\left(\rho_{12}\nu_{t2}+\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)\right],\\
\xi_{22t} &= \frac{1}{\alpha}c_t\left[\frac{\phi(\pi_t)}{\Phi(\pi_t)}\pi_t+\frac{\phi^2(\pi_t)}{\Phi^2(\pi_t)}\right]+(1-c_t)\psi^*(\nu_{t2})\left[\psi^*(\nu_{t2})-\nu_{t2}\right],\\
\xi_{32t} &= -c_t\left[\frac{\rho_{12}}{\alpha}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\pi_t\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)+\frac{\rho_{12}}{\alpha}\frac{\phi^2(\pi_t)}{\Phi^2(\pi_t)}\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)\right],\\
\xi_{42t} &= -c_t\frac{1}{\alpha^{1/2}}\left[\frac{\phi^2(\pi_t)}{\Phi^2(\pi_t)}\left(\rho_{12}\nu_{t2}+\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)\right],\\
\xi_{42t}^* &= c_t\left[\rho_{12}\frac{\phi(\pi_t)}{\Phi(\pi_t)}-\frac{1}{\alpha^{1/2}}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\pi_t\left(\rho_{12}\nu_{t2}+\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)\right],\\
\xi_{33t} &= c_t\left[2\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)^2-\frac{\rho_{12}}{\alpha^{1/2}}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\frac{u_{t1}}{\sigma_{11}^{1/2}}+\frac{\rho_{12}^2}{\alpha}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\pi_t\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)^2\right.\\
&\quad\left.+\frac{\rho_{12}^2}{\alpha}\frac{\phi^2(\pi_t)}{\Phi^2(\pi_t)}\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right)^2\right],
\end{aligned}
\tag{139}
$$

$$\xi^*_{33t} = \xi_{3t},$$

$$\xi_{43t} = c_t\left[\frac{\phi(\pi_t)}{\Phi(\pi_t)}\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right) - \frac{\rho_{12}}{\alpha^{1/2}}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\pi_t\left(\rho_{12}\nu_{t2}\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right) + \frac{u_{t1}^2}{\sigma_{11}}\right)\right.$$

$$\left. - \frac{\rho_{12}}{\alpha^{1/2}}\frac{\phi^2(\pi_t)}{\Phi^2(\pi_t)}\left(\rho_{12}\nu_{t2}\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right) + \frac{u_{t1}^2}{\sigma_{11}}\right)\right],$$

$$\xi_{44t} = \xi_{4t}^2,$$

$$\xi^*_{44t} = \frac{3\rho_{12}}{\alpha}\xi_{4t} + c_t\left[\frac{\phi(\pi_t)}{\Phi(\pi_t)}\nu_{t2} - \frac{1}{\alpha^{3/2}}\frac{\phi(\pi_t)}{\Phi(\pi_t)}\pi_t\left(\rho_{12}\nu_{t2} + \frac{u_{t1}}{\sigma_{11}^{1/2}}\right)^2\right].$$

In the expressions of (138) and (139) we have replaced, for reasons of notational economy only,

$$\left(\frac{y_{t1} - x_{t\cdot}\beta_{\cdot 1}}{\sigma_{11}^{1/2}}\right),$$

by

$$\left(\frac{u_{t1}}{\sigma_{11}^{1/2}}\right).$$

Remark 10

The starred symbols, for example, ξ^*_{42t}, ξ^*_{33t}, ξ^*_{44t}, all correspond to components of the Hessian of the log LF *having mean zero.* Hence, such components can be ignored both in determining the limiting distribution of the ML estimator and in its numerical derivation, given a sample. We can, then, represent the Hessian of the log of the LF as

$$\frac{\partial^2 L}{\partial\gamma\partial\gamma} = \sum_{t=1}^{T}\Omega_t + \sum_{t=1}^{T}\Omega_t^*,$$

where Ω_t^* contains only zeros or elements having mean zero. It is also relatively straightforward to verify that

$$A_t\mathrm{Cov}(\xi_{\cdot t})A_t' = E(\Omega_t),$$

where the elements of A_t, $\xi_{\cdot t}$ and Ω_t have been evaluated at the true parameter point γ^0.

To determine the limiting distribution of the ML estimator (i.e. the converging iterate beginning with an initial consistent estimator) we need

Lemma 6

Let A_t, $\xi_{\cdot t}$ be as defined in (139) and (138); then,

$$\frac{1}{\sqrt{T}}\frac{\partial L}{\partial \gamma'}(\gamma^0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} A_t \xi_{\cdot t} \sim N(0, C_*),$$

where

$$C_* = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} A_t \mathrm{Cov}(\xi_{\cdot t}) A_t' = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E(\Omega_t). \tag{141}$$

Proof

The sequence

$$\{A_t\xi_{\cdot t}: t = 1,2,\ldots\},$$

is one of independent nonidentically distributed random vectors with mean zero and uniformly bounded moments to any finite order; moreover, the sequence obeys a Liapounov condition. Consequently

$$\frac{1}{\sqrt{T}}\frac{\partial L}{\partial \gamma}(\gamma^0) \sim N(0, C_*). \qquad \text{(Q.E.D.)}$$

An explicit representation of Ω_t or C_* is omitted here because of its notational complexity. To complete the argument concerning the limiting distribution of the ML estimator we obtain the limit of

$$\frac{1}{T}\frac{\partial^2 L}{\partial \gamma \partial \gamma}(\gamma), \qquad \gamma \in H.$$

Again for the sake of brevity of exposition we shall only state the result without proof

Lemma 7

Under assumptions (A.3.1) through (A.3.4)

$$\frac{1}{T}\frac{\partial^2 L}{\partial \gamma \partial \gamma}(\gamma) \overset{\text{a.c.}}{\to} \lim_{T\to\infty}\frac{1}{T}E\left[\frac{\partial^2 L}{\partial \gamma \partial \gamma}(\gamma)\right],$$

uniformly in γ.

Remark 11

We note that for $\gamma = \gamma^0$

$$\lim_{T \to \infty} \frac{1}{T} E\left[\frac{\partial^2 L}{\partial \gamma \partial \gamma}(\gamma)\right] = -C_*.$$

We finally have

Theorem 5

(Asymptotic Normality): Consider the model of (97) subject to the condition in (107) and assumptions (A.3.1.) through (A.3.4.). Let $\hat{\gamma}$ be the ML estimator of γ^0 – the true parameter point. Then

$$\sqrt{T}(\hat{\gamma} - \gamma^0) \sim N(0, C),$$

where

$$C = C_*^{-1},$$

and C_* is as is defined in (141).

Proof

From the expansion in (135)

$$\sqrt{T}(\hat{\gamma} - \gamma^0) \sim C_*^{-1} \frac{1}{\sqrt{T}} \frac{\partial L}{\partial \gamma'}(\gamma^0).$$

From Lemma 6 we then conclude

$$\sqrt{T}(\hat{\gamma} - \gamma^0) \sim N(0, C). \qquad \text{Q.E.D.}$$

Corollary 6

The ML estimator $\hat{\gamma}$ obeys

$$\hat{\gamma} \overset{\text{a.c.}}{\to} \gamma^0.$$

Proof

By Theorem 5

$$\sqrt{T}(\hat{\gamma} - \gamma^0) \sim \zeta,$$

where ζ is a well defined a.c. finite random variable. Hence,

$$\tilde{\gamma}-\gamma^0 \sim \frac{\zeta}{\sqrt{T}} \overset{\text{a.c.}}{\to} 0.$$

Corollary 7

The matrix in the expansion of (135) obeys

$$\frac{1}{T}\frac{\partial^2 L}{\partial\gamma\partial\gamma}(\gamma^*) \overset{\text{a.c.}}{\to} \lim_{T\to\infty}\frac{1}{T}E\left[\frac{\partial^2 L}{\partial\gamma\partial\gamma}(\gamma^0)\right].$$

Proof

Lemma 7 and Corollary 6.

3.6. *A test for selectivity bias*

A test for selectivity bias is formally equivalent to the test of

$$H_0\colon \rho_{12}=0 \qquad \text{or} \qquad \gamma=\left(\beta'_{\cdot 1},\beta'_{\cdot 2},\sigma_{11},0\right)'$$

as against the alternative

H_1: γ unrestricted (except for the obvious conditions, $\sigma_{11}>0$, $\rho_{12}^2\varepsilon[0,1]$). From the likelihood function in eq. (109) the (log) LF under H_0 becomes

$$L(\gamma|H_0)=\sum_{t=1}^{T}\Bigg\{(1-c_t)\ln\Phi(-\nu_{t2})+c_t\ln\Phi(\nu_{t2}) - \frac{1}{2}c_t\left[\ln(2\pi)+\ln\sigma_{11}+\frac{1}{\sigma_{11}}(y_{t1}-x_{t1\cdot}\beta_{\cdot 1})^2\right]\Bigg\}. \tag{142}$$

We note that (142) is separable in the parameters $(\beta'_{\cdot 1},\sigma_{11})'$ and $\beta_{\cdot 2}$. Indeed, the ML estimator of $\beta_{\cdot 2}$ is the "probit" estimator, $\hat{\beta}_{\cdot 2}$, obtained in connection with eq. (118) in Section 3.d.; the ML estimator of $(\beta'_{\cdot 1},\sigma_{11})'$ is the usual one obtained by least squares except that σ_{11} is estimated with bias – as all maximum likelihood procedures imply in the normal case. Denote the estimator of γ obtained under H_0, by $\tilde{\gamma}$. Denote by $\hat{\gamma}$ the ML estimator whose limiting distribution was obtained in the preceding section.

Thus

$$\lambda = L(\tilde{\gamma}|H_1) - L(\hat{\gamma}|H_0). \tag{143}$$

is the usual likelihood rationtest statistic. It may be shown that

$$-2\lambda \sim \chi_1^2. \tag{144}$$

We have thus proved

Theorem 6

In the context of the model of this section a test for the absence of selectivity bias can be carried out by the likelihood ratio (LR) principle. The test statistic is

$$-2\lambda \sim \chi_1^2,$$

where

$$\lambda = \sup_{H_0} L(\gamma) - \sup_{H_1} L(\gamma).$$

References

Aitchison, J. and J. Bennett (1970) "Polychotomous Quantal Response by Maximum Indicant", *Biometrika*, 57, 253–262.

Aitchison, J. and S. Silvey (1957) "The Generalization of Probit Analysis to the Case of Multiple Responses", *Biometrika*, 37, 131–140.

Amemiya, T. (1973) "Regression Analysis When the Dependent Variable Is Truncated Normal", *Econometrica*, 41, 997–1016.

Amemiya, T. (1974) "Bivariate Probit Analysis: Minimum Chi-Square Methods", *Journal of the American Statistical Association*, 69, 940–944.

Amemiya, T. (1974) "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal", *Econometrica*, 42, 999–1012.

Amemiya, T. (1974) "A Note on the Fair and Jaffee Model", *Econometrica*, 42, 759–762.

Amemiya, T. (1975) "Qualitative Response Models", *Annals of Economic and Social Measurement*, 4, 363–372.

Amemiya, T. (1976) "The Maximum Likelihood, the Minimum Chi-Square, and the Non-linear Weighted Least Squares Estimator in the General Qualitative Response Model", *JASA*, 71.

Ameniya, T. (1978) "The Estimation of a Simultaneous Equation Generalized Probit Model", *Econometrica*, 46, 1193–1205.

Amemiya, T. (1978) "On a Two-Step Estimation of a Multivariate Logit Model", *Journal of Econometrics*, 8, 13–21.

Amemiya, T. and F. Nold (1975) "A Modified Logit Model", *Review of Economics and Statistics*, 57, 255–257.

Anscombe, E. J. (1956) "On Estimating Binomial Response Relations", *Biometrika*, 43, 461–464.

Ashford, J. R. and R. R. Sowden (1970) "Multivariate Probit Analysis", *Biometrics*, 26, 535–546.

Ashton, W. (1972) *The Logit Transformation*. New York: Hafner.

Bartlett, M. S. (1935) "Contingent Table Interactions", *Supplement to the Journal of the Royal Statistical Society*, 2, 248–252.
Berkson, J. (1949) "Application of the Logistic Function to Bioassay", *Journal of the American Statistical Association*, 39, 357–365.
Berkson, J. (1951) "Why I Prefer Logits to Probits", *Biometrika*, 7, 327–339.
Berkson, J. (1953) "A Statistically Precise and Relatively Simple Method of Estimating the Bio-Assay with Quantal Response, Based on the Logistic Function", *Journal of the American Statistical Association*, 48, 565–599.
Berkson, J. (1955) "Estimate of the Integrated Normal Curve by Minimum Normit Chi-Square with Particular Reference to Bio-Assay", *Journal of the American Statistical Association*, 50, 529–549.
Berkson, J. (1955) "Maximum Likelihood and Minimum Chi-Square Estimations of the Logistic Function", *Journal of the American Statistical Association*, 50, 130–161.
Bishop, T., S. Feiberg and P. Hollan (1975) *Discrete Multivariate Analysis*. Cambridge: MIT Press.
Block, H. and J. Marschak (1960) "Random Orderings and Stochastic Theories of Response", in: I. Olkin, ed., *Contributions to Probability and Statistics*. Stanford: Stanford University Press.
Bock, R. D. (1968) "Estimating Multinomial Response Relations", in: *Contributions to Statistics and Probability: Essays in Memory of S. N. Roy*. Chapel Hill: University of North Carolina Press.
Bock, R. D. (1968) *The Measurement and Prediction of Judgment and Choice*. San Francisco: Holden-Day.
Boskin, M. (1974) "A Conditional Logit Model of Occupational Choice", *Journal of Political Economy*, 82, 389–398.
Boskin, M. (1975) "A Markov Model of Turnover in Aid to Families with Dependent Children", *Journal of Human Resources*, 10, 467–481.
Chambers, E. A. and D. R. Cox (1967) "Discrimination between Alternative Binary Response Models", *Biometrika*, 54, 573–578.
Cosslett, S. (1980) "Efficient Estimators of Discrete Choice Models", in: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data*. Cambridge: MIT Press.
Cox, D. (1970) *Analysis of Binary Data*. London: Methuen.
Cox, D. (1972) "The Analysis of Multivariate Binary Data", *Applied Statistics*, 21, 113–120.
Cox, D. (1958) "The Regression Analysis of Binary Sequences", *Journal of the Royal Statistical Society*, Series B, 20, 215–242.
Cox, D. (1966) "Some Procedures Connected with the Logistic Response Curve", in: F. David, ed., *Research Papers in Statistics*. New York: Wiley.
Cox, D. and E. Snell (1968) "A General Definition of Residuals", *Journal of the Royal Statistical Society*, Series B, 30, 248–265.
Cragg, J. G. (1971) "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods", *Econometrica*, 39, 829–844.
Cragg, J. and R. Uhler (1970) "The Demand for Automobiles", *Canadian Journal of Economics*, 3, 386–406.
Cripps, T. F. and R. J. Tarling (1974) "An Analysis of the Duration of Male Unemployment in Great Britain 1932–1973", *The Economic Journal*, 84, 289–316.
Daganzo, C. (1980) *Multinomial Probit*. New York: Academic Press.
Dagenais, M. G. (1975) "Application of a Threshold Regression Model to Household Purchases of Automobiles", *The Review of Economics and Statistics*, 57, 275–285.
Debreu, G. (1960) "Review of R. D. Luce Individual Choice Behavior", *American Economic Review*, 50, 186–188.
Dhrymes, P. J. (1970) *Econometrics: Statistical Foundations and Applications*. Harper & Row, 1974, New York: Springer-Verlag.
Dhrymes, P. J. (1978a) *Introductory Econometrics*. New York: Springer-Verlag.
Dhrymes, P. J. (1978b) *Mathematics for Econometrics*. New York: Springer-Verlag.
Domencich, T. and D. McFadden (1975) *Urban Travel Demand: A Behavioral Analysis*. Amsterdam: North-Holland.
Efron, B. (1975) "The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis", *Journal of the American Statistical Association*, 70, 892–898.
Fair, R. C. and D. M. Jaffee (1972) "Methods of Estimation for Markets in Disequilibrium", *Econometrica*, 40, 497–514.

Finney, D. (1964) *Statistical Method in Bio-Assay*. London: Griffin.
Finney, D. (1971) *Probit Analysis*. New York: Cambridge University Press.
Gart, J. and J. Zweifel (1967) "On the Bias of Various Estimators of the Logit and Its Variance", *Biometrika*, 54, 181–187.
Gillen, D. W. (1977) "Estimation and Specification of the Effects of Parking Costs on Urban Transport Mode Choice", *Journal of Urban Economics*, 4, 186–199.
Goldberger, A. S. (1971) "Econometrics and Psychometrics: A Survey of Communalities", *Psychometrika*, 36, 83–107.
Goldberger, A. S. (1973) "Correlations Between Binary Outcomes and Probabilistic Predictions", *Journal of American Statistical Association*, 68, 84.
Goldfeld, S. M. and R. E. Quandt (1972) *Nonlinear Methods on Econometrics*. Amsterdam: North-Holland.
Goldfeld, S. M. and R. E. Quandt (1973) "The Estimation of Structural Shifts by Switching Regressions", *Annals of Economic and Social Measurement*, 2, 475–485.
Goldfeld, S. M. and R. E. Quandt (1976) "Techniques for Estimating Switching Regressions", in: S. Goldfeld and R. Quandt, eds., *Studies in Non-Linear Estimation*. Cambridge: Ballinger.
Goodman, I. and W. H. Kruskal (1954) "Measures of Association for Cross Classifications", *Journal of the American Statistical Association*, 49, 732–764.
Goodman, I. and W. H. Kruskal (1954) "Measures of Association for Cross Classification II, Further Discussion and References", *Journal of the American Statistical Association*, 54, 123–163.
Goodman, L. A. (1970) "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications", *Journal of the American Statistical Association*, 65, 226–256.
Goodman, L. A. (1971) "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications", *Technometrics*, 13, 33–61.
Goodman, L. A. (1972) "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables", *American Sociological Review*, 37, 28–46.
Goodman, L. A. (1972) "A General Model for the Analysis of Surveys", *American Journal of Sociology*, 77, 1035–1086.
Goodman, L. A. (1973) "Causal Analysis of Panel Study Data and Other Kinds of Survey Data", *American Journal of Sociology*, 78, 1135–1191.
Griliches, Z., B. H. Hall and J. A. Hausman (1978) "Missing Data and Self-Selection in Large Panels", *Annals de l'Insee*, 30–31, 137–176.
Grizzle, J. (1962) "Asymptotic Power of Tests of Linear Hypotheses Using the Probit and Logit Transformations", *Journal of the American Statistical Association*, 57, 877–894.
Grizzle, J. (1971) "Multivariate Logit Analysis", *Biometrics*, 27, 1057–1062.
Gronau, R. (1973) "The Effect of Children on the Housewife's Value of Time", *Journal of Political Economy*, 81, 168–199.
Gronau, R. (1974) "Wage Comparisons: A Selectivity Bias", *Journal of Political Economy*, 82, 1119–1143.
Gurland, J., I. Lee and P. Dahm (1960) "Polychotomous Quantal Response in Biological Assay", *Biometrics*, 16, 382–398.
Haberman, S. (1974) *The Analysis of Frequency Data*. Chicago: University of Chicago Press.
Haldane, J. (1955) "The Estimation and Significance of the Logarithm of a Ratio of Frequencies", *Annals of Human Genetics*, 20, 309–311.
Harter, J. and A. Moore (1967) "Maximum Likelihood Estimation, from Censored Samples, of the Parameters of a Logistic Distribution", *Journal of the American Statistical Association*, 62, 675–683.
Hausman, J. (1979) "Individual Discount Rates and the Purchase and Utilization of Energy Using Durables", *Bell Journal of Economics*, 10, 33–54.
Hausman, J. A. and D. A. Wise (1976) "The Evaluation of Results from Truncated Samples: The New Jersey Negative Income Tax Experiment", *Annals of Economic and Social Measurement*, 5, 421–445.
Hausman, J. A. and D. A. Wise (1977) "Social Experimentation, Truncated Distributions and Efficient Estimation", *Econometrica*, 45, 319–339.
Hausman, J. A. and D. A. Wise (1978) "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences", *Econometrica*, 46, 403–426.

Hausman, J. A. and D. A. Wise (1980) "Stratification on Endogenous Variables and Estimation: The Gary Experiment", in: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data*. Cambridge: MIT Press.
Heckman, J. (1974) "Shadow Prices, Market Wages, and Labor Supply", *Econometrica*, 42, 679–694.
Heckman, J. (1976) "Simultaneous Equations Model with Continuous and Discrete Endogenous Variables and Structural Shifts", in: S. M. Goldfeld and E. M. Quandt, eds., *Studies in Non-Linear Estimation*. Cambridge: Ballinger.
Heckman, J. (1976) "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimation for Such Models", *Annals of Economic and Social Measurement*, 5, 475–492.
Heckman, J. (1978) "Dummy Exogenous Variables in a Simultaneous Equation System", *Econometrica*, 46, 931–959.
Heckman, J. (1978) "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence Against the Hypothesis of Spurious State Dependence", *Annals de l'Insee*, 30–31, 227–270.
Heckman, J. (1979) "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153–163.
Heckman, J. (1980) "Statistical Models for the Analysis of Discrete Panel Data", in: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data*. Cambridge: MIT Press.
Heckman, J. (1980) "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Stochastic Process and Some Monte Carlo Evidence on Their Practical Importance", in: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data*. Cambridge: MIT Press.
Heckman, J. and R. Willis (1975) "Estimation of a Stochastic Model of Reproduction: An Econometric Approach", in: N. Terleckyj, ed., *Household Production and Consumption*. New York: National Bureau of Economic Research.
Heckman, J. and R. Willis (1977) "A Beta Logistic Model for the Analysis of Sequential Labor Force Participation of Married Women", *Journal of Political Economy*, 85, 27–58.
Joreskog, K. and A. S. Goldberger (1975) "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable Model", *Journal of the American Statistical Association*, 70, 631–639.
Kiefer, N. (1978) "Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model", *Econometrica*, 46, 427–434.
Kiefer, N. (1979) "On the Value of Sample Separation Information", *Econometrica*, 47, 997–1003.
Kiefer, N. and G. Neumann (1979) "An Empirical Job Search Model with a Test of the Constant Reservation Wage Hypothesis", *Journal of Political Economy*, 87, 89–107.
Kohn, M., C. Manski and D. Mundel (1976), "An Empirical Investigation of Factors Influencing College Going Behavior", *Annals of Economic and Social Measurement*, 5, 391–419.
Ladd, G. (1966) "Linear Probability Functions and Discriminant Functions", *Econometrica*, 34, 873–885.
Lee, L. F. (1978) "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables", *International Economic Review*, 19, 415–433.
Lee, L. F. (1979) "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables", *Econometrica*, 47, 977–996.
Lee, L. F. (1980) "Simultaneous Equations Models with Discrete and Censored Variables", in: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data*. Cambridge: MIT Press.
Lee, L. F. and R. P. Trost (1978) "Estimation of Some Limited Dependent Variable Models with Applications to Housing Demand", *Journal of Econometrics*, 8, 357–382.
Lerman, S. and C. Manski (1980) "On the Use of Simulated Frequencies to Approximate Choice Probabilities", in: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data*. Cambridge: MIT Press.
Li, M. (1977) "A Logit Model of Home Ownership", *Econometrica*, 45, 1081–1097.
Little, R. E. (1968) "A Note on Estimation for Quantal Response Data", *Biometrika*, 55, 578–579.
Luce, R. D. (1959) *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
Luce, R. D. (1977) "The Choice Axiom After Twenty Years", *Journal of Mathematical Psychology*, 15, 215–233.
Luce, R. D. and P. Suppes (1965) "Preference, Utility, and Subjective Probability", in: R. Luce, R.

Bush and E. Galanter, eds., *Handbook of Mathematical Psychology III*. New York: Wiley.
Maddala, G. S. (1977) "Self-Selectivity Problem in Econometric Models", in: P. Krishniah, ed., *Applications of Statistics*. Amsterdam: North-Holland.
Maddala, G. S. (1977) "Identification and Estimation Problems in Limited Dependent Variable Models", in: A. S. Blinder and P. Friedman, eds., *Natural Resources, Uncertainty and General Equilibrium Systems: Essays in Memory of Rafael Lusky*. New York: Academic Press.
Maddala, G. S. (1978) "Selectivity Problems in Longitudinal Data", *Annals de l'INSEE*, 30–31, 423–450.
Maddala, G. S. and L. F. Lee (1976) "Recursive Models with Qualitative Endogenous Variables", *Annals of Economic and Social Measurement*, 5.
Maddala, G. and F. Nelson (1974) "Maximum Likelihood Methods for Markets in Disequilibrium", *Econometrica*, 42, 1013–1030.
Maddala, G. S. and R. Trost (1978) "Estimation of Some Limited Dependent Variable Models with Application to Housing Demand", *Journal of Econometrics*, 8, 357–382.
Maddala, G. S. and R. Trost (1980) "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity", *Econometrica*, 48, 491–503.
Manski, C. (1975) "Maximum Score Estimation of the Stochastic Utility Model of Choice", *Journal of Econometrics*, 3, 205–228.
Manski, C. (1977) "The Structure of Random Utility Models", *Theory and Decision*, 8, 229–254.
Manski, C. and S. Lerman (1977) "The Estimation of Choice Probabilities from Choice-Based Samples", *Econometrica*, 45, 1977–1988.
Manski, C. and D. McFadden (1980) "Alternative Estimates and Sample Designs for Discrete Choice Analysis", in: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data*. Cambridge: MIT Press.
Marshak, J. "Binary-Choice Constraints and Random Utility Indicators", in: K. Arrow, S. Karlin and P. Suppes, eds., *Mathematical Methods in the Social Sciences*. Stanford University Press.
McFadden, D. "Conditional Logit Analysis of Qualitative Choice Behavior", in: P. Zarembka, ed., *Frontiers in Econometrics*. New York: Academic Press.
McFadden, D. (1976) "A Comment on Discriminant Analysis 'Versus' Logit Analysis", *Annals of Economics and Social Measurement*, 5, 511–523.
McFadden, D. (1976) "Quantal Choice Analysis: A Survey", *Annals of Economic and Social Measurement*, 5, 363–390.
McFadden, D. (1976) "The Revealed Preferences of a Public Bureaucracy", *Bell Journal*, 7, 55–72.
Miller, L. and R. Radner (1970) "Demand and Supply in U.S. Higher Education", *American Economic Review*, 60, 326–334.
Moore, D. H. (1973) "Evaluation of Five Discrimination Procedures for Binary Variables", *Journal of American Statistical Association*, 68, 399–404.
Nelson, F. (1977) "Censored Regression Models with Unobserved Stochastic Censoring Thresholds", *Journal of Econometrics*, 6, 309–327.
Nelson, F. S. and L. Olsen (1978) "Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables", *International Economic Review*, 19, 695–710.
Nerlove, M. (1978) "Econometric Analysis of Longitudinal Data: Approaches, Problems and Prospects", *Annales de l'Insee*, 30–31, 7–22.
Nerlove, M. and J. Press (1973) "Univariable and Multivariable Log-Linear and Logistic Models", RAND Report No. R-1306-EDA/NIH.
Oliveira, J. T. de (1958) "Extremal Distributions", *Revista de Faculdada du Ciencia, Lisboa, Serie A*, 7, 215–227.
Olsen, R. J. (1978) "Comment on 'The Effect of Unions on Earnings and Earnings on Unions: A Mixed Logit Approach'", *International Economic Review*, 259–261.
Plackett, R. L. (1974) *The Analysis of Categorical Data*. London: Charles Griffin.
Poirier, D. J. (1976) "The Determinants of Home Buying in the New Jersey Graduated Work Incentive Experiment", in: H. W. Watts and A. Rees, eds., *Impact of Experimental Payments on Expenditure, Health and Social Behavior, and Studies on the Quality of the Evidence*. New York: Academic Press.
Poirier, D. J. (1980) "A Switching Simultaneous Equation Model of Physician Behavior in Ontario",

in: D. McFadden and C. Manski, eds., *Structural Analysis of Discrete Data: With Econometric Applications*. Cambridge: MIT Press.

Pollakowski, H. (1980) *Residential Location and Urban Housing Markets*. Lexington: Heath.

Quandt, R. (1956) "Probabilistic Theory of Consumer Behavior", *Quarterly Journal of Economics*, 70, 507–536.

Quandt, R. (1970) *The Demand for Travel*. London: Heath.

Quandt, R. (1972) "A New Approach to Estimating Switching Regressions", *Journal of the American Statistical Association*, 67, 306–310.

Quandt, R. (1978) "Tests of the Equilibrium vs. Disequilibrium Hypothesis", *International Economic Review*, 19, 435–452.

Quandt, R. and W. Baumol (1966) "The Demand for Abstract Travel Modes: Theory and Measurement", *Journal of Regional Science*, 6, 13–26.

Quandt, R. E. and J. B. Ramsey (1978) "Estimating Mixtures of Normal Distributions and Switching Regressions", *Journal of the American Statistical Association*, 71, 730–752.

Quigley, J. M. (1976) "Housing Demand in the Short-Run: An Analysis of Polytomous Choice", *Explorations in Economic Research*, 3, 76–102.

Radner, R. and L. Miller (1975) *Demand and Supply in U.S. Higher Education*. New York: McGraw-Hill.

Sattath, S. and A. Tversky (1977) "Additive Similarity Trees", *Psychometrika*, 42, 319–345.

Shakotko, Robert A. and M. Grossman (1981) "Physical Disability and Post-Secondary Educational Choices", in: V. R. Fuchs, ed., *Economic Aspects of Health*. National Bureau of Economic Research, Chicago: University of Chicago Press.

Sickles, R. C. and P. Schmidt (1978) "Simultaneous Equation Models with Truncated Dependent Variables: A Simultaneous Tobit Model", *Journal of Economics and Business*, 31, 11–21.

Theil, H. (1969) "A Multinomial Extension of the Linear Logit Model", *International Economic Review*, 10, 251–259.

Theil, H. (1970) "On the Estimation of Relationships Involving Qualitative Variables", *American Journal of Sociology*, 76, 103–154.

Thurstone, L. (1927) "A Law of Comparative Judgement", *Psychological Review*, 34, 273–286.

Tobin, J. (1958) "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, 24–36.

Tversky, A. (1972) "Choice by Elimination", *Journal of Mathematical Psychology*. 9, 341–367.

Tversky, A. (1972) "Elimination by Aspects: A Theory of Choice", *Psychological Review*, 79, 281–299.

Walker, S. and D. Duncan (1967) "Estimation of the Probability of an Event as a Function of Several Independent Variables", *Biometrika*, 54, 167–179.

Westin, R. (1974) "Predictions from Binary Choice Models", *Journal of Econometrics*, 2, 1–16.

Westin, R. B. and D. W. Gillen (1978) "Parking Location and Transit Demand: A Case Study of Endogenous Attributes in Disaggregate Mode Choice Functions", *Journal of Econometrics*, 8, 75–101.

Willis, R. and S. Rosen (1979) "Education and Self-Selection", *Journal of Political Economy*, 87, 507–536.

Yellot, J. (1977) "The Relationship Between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution", *Journal of Mathematical Psychology*, 15, 109–144.

Zellner, A. and T. Lee (1965) "Joint Estimation of Relationships Involving Discrete Random Variables", *Econometrica*, 33, 382–394.