Semiparametric Estimation and Inference in Multinomial Choice Models

By

Rafic Fahs , Scott Cardell and Ron Mittelhammer[1]

2001 AAEA Annual Conference

Abstract

The purpose of this paper is to incorporate semiparametric alternatives to maximum likelihood estimation and inference in the context of unordered multinomial response data when in practice there is often insufficient information to specify the parametric form of the function linking the observables to the unknown probabilities. We specify the function linking the observables to the unknown probabilities using a very general flexible class of functions belonging to the Pearson system of cumulative distribution equations. In this setting we consider the observations as arising from a multinomial distribution characterized by one of the CDFs in the Pearson system. Given this situation, it is possible to utilize the concept of unbiased estimating functions (EFs), combined with the concept of empirical likelihood (EL) to define an (empirical) likelihood function for the parameter vector based on a nonparametric representation of the sample's PDF. This leads to the concept of maximum empirical likelihood (MEL) estimation and inference, which is analogous to parametric maximum likelihood methods in many respects.

[1] Rafic Fahs is a Ph.D candidate in Agricultural Economics at Washington State University. Scott Cardell is Adjunct Professor in Agricultural Economics at Washington State University and Ron Mittelhammer is a Professor of Ag Economics and Statistics.

1. Introduction

Consider the unordered multinomial response model where outcomes are given in the form of an experiment consisting of N trials on J-dimensional multinomial random variables $(y_{11},...,y_{1J}),...,(y_{N1},...,y_{NJ})$. The variable y_{ij}, for i = 1,2,….N, exhibits a binary outcome, where $y_{ij} = 1$ is observed iff the j^{th} choice among J unordered alternatives j = 1,2,….J, is observed on the i^{th} trial, in which case $y_{ik} = 0 \;\; \forall k \neq j$. It is assumed that the choice situation is such that the J alternatives are mutually exclusive, so that only one of the alternatives can be chosen on trial i, and the J alternatives exhaust the choice possibilities, which then implies that $\sum_{j=1}^{J} y_{ij} = 1, \forall i$.

The probability that $y_{ij} = 1$, denoted by P_{ij}, is then related to a set of K explanatory variables $\mathbf{x}_{i.}$ through a link function

$$P_{ij}(\mathbf{x}_{i.}) = P(y_{ij} = 1 | \mathbf{x}_{i.}, \beta) = G_j(\mathbf{x}_{i.}, \beta) \tag{1.1}$$

for i = 1,2,….N and j = 1,2,….J, where β_j is a $(K\times 1)$ vector of unknown parameters, $\beta = \text{vec}([\beta_1, \beta_2, ...\beta_J])$ is a column-vectorized representation of model parameters of dimension $(KJ\times 1)$, $\mathbf{x}_{i.} = (\mathbf{x}_{i1}, \mathbf{x}_{i2},…., \mathbf{x}_{iK})$ is a (1xK) row vector of explanatory variables values, and $G(\bullet): R \rightarrow [0,1]$ may be either known or unknown, with the additional constraint that

$$\sum_{j=1}^{J} G_j(\mathbf{x}_{i.}, \beta) = 1, \;\forall i \;. \tag{1.2}$$

The explanatory variables, $\mathbf{x}_{i.}$, could be allowed to change by choice alternative, but we focus on a basic case where they do not.

Define the noise term ε_{ij} as

$$\varepsilon_{ij} \equiv y_{ij} - E[y_{ij}|\mathbf{x}_{i.}] = y_{ij} - G_j(\mathbf{x}_{i.}, \beta)$$

where $E[y_{ij}|\mathbf{x}_{i.}] = G_j(\mathbf{x}_{i.}, \beta)$ because of the Bernoulli marginal distribution of the y_{ij} variable. The data sampling process relating to observed choices can then be represented as

$$y_{ij} = P_{ij}(\mathbf{x}_{i.}) + \varepsilon_{ij} = G_j(\mathbf{x}_{i.}, \beta)) + \varepsilon_{ij}. \qquad (1.3)$$

where the ε_{ij}'s are assumed to be independent across observations i = 1,2,….N, the ε_{ij}'s are bounded between [-1,1], and $E[\varepsilon_{ij}\,\mathbf{x}_{i.}] = \mathbf{0}$.

When the parametric functional form of $G_j(\mathbf{x}_{i.}, \beta)$ is known, maximum likelihood (ML) estimation is possible, and a specific functional choice has often been

$$G_j(\mathbf{x}_{i.}, \beta) = \frac{1}{1 + \sum_{k=2}^{J} e^{\mathbf{x}_{i}.\beta_k}} \qquad \text{for } j = 1$$

$$= \frac{e^{x_i.\beta_j}}{1 + \sum_{k=2}^{J} e^{\mathbf{x}_{i}.\beta_k}} \qquad \text{for } j = 2,3,\ldots.J \qquad (1.4)$$

where β_1 has been normalized, without loss of generality, to a zero vector for purposes of parameter identification. The definition in (1.4) satisfies the required properties in (1.2) and defines the *Multinomial Logit* (ML*)* response model applied frequently in practice. Similarly, a *Multinomial Probit* (MNP) model results when a multivariate normal distribution is used in specifying the distribution of the noise component of (1.3).

However, when J is large, the probit model is computationally difficult except when the number of alternative choices is restricted to 3 or less. If the distribution underlying the likelihood specification is in fact the correct parametric family of distributions underlying the sampling process, then the estimator is generally unique, consistent and asymptotically normal. However, the economic theories that motivate these models rarely justify any particular probability distribution for the noise term .

Attempts to introduce flexibility into the specification of $G(\bullet)$ have been problematic in applications. Supposing $G(\bullet)$ is a polynomial of order d, some trigonometric function, or some other flexible functional form, the flexibility added to the estimation problem may introduce unbounded likelihood functions on parameter space boundaries, multiple local maxima, and/or non-concavities that make numerical maximization of the log likelihood function difficult or impossible.

Semi-parametric methods of estimation provide an alternative approach, such as Ichimura (1993) who demonstrates a least squares estimate of β which requires ε_{ij} to be independent of $\mathbf{x}_{i.}$, ruling out endogeneity and/or measurement error. Klien and Spady (1993) developed a quasi-maximum likelihood estimator for the case in which Y_{ij} is binary. These estimators are consistent and asymptotically normal under regularity conditions. But these estimators share the disadvantage of being difficult to compute because they involve nonlinear optimization problems whose objective functions are not necessarily concave or unimodel. Using an information theoretic formulation, Golan, Judge, and Perloff (1996) demonstrate a semiparametric estimator for the traditional multinomial response problem that has asymptotic properties in line with parametric counterparts. But a potential draw back to their method is that the P_{ij}'s have no direct parametric functional link to the $\mathbf{x}_{i.}$'s which makes prediction difficult or impossible.

To cope with the preceding modeling issues one can hedge against misspecification when the form of the link function $G(\bullet)$ is unknown by giving $G(\bullet)$ a flexible form that satisfies (1.2) and that defines a legitimate multinomial response model globally. One might model each dichotomous decision outcome as a Bernoulli process (marginally) and then model the whole vector outcome, $\mathbf{y}_{i.}$, as a multinomial process. One possibility for parameterizing the probabilities in these processes is the set of CDFs belonging to the highly flexible Pearson system of distributions, which themselves satisfy (1.2). The criteria for identifying different members of the Pearson system of functions can be expressed parametrically in terms of a (2x1) ξ vector of unknown parameters. When $G(\bullet)$ is unknown, the sampling process would then be modeled as

$$y_{ij} = P_{ij}(\mathbf{x}_{i.}) + \varepsilon_{ij} = G_j(\mathbf{x}_{i.}, \beta, \xi) + \varepsilon_{ij}. \qquad (1.5)$$

The overall objective of this paper is to seek a semiparametric basis for recovering β in (1.5) by utilizing the concept of unbiased estimating functions (EFs) combined with the concept of empirical likelihood (EL) to define an empirical likelihood function for the parameter vector. This leads to the concept of maximum empirical likelihood estimation of the multinomial choice model. The EL shares the sampling properties of various nonparametric methods based on resampling of the data, such as the bootstrap. However, in contrast to the resampling methods, EL works by optimizing a continuous concave and differentiable function having a unique global maximum, which makes it possible to impose side constraints on the parameters that add information to the data.

The paper is organized as follows. Section 2 reviews the Pearson system of density and cumulative distribution functions. In section 3, we state the extremum estimation problem and investigate the asymptotic properties of the estimator. In section 4, we examine EL estimation in the multinomial choice problem. In section 5, we present some Monte Carlo results relating to the sampling properties of the estimator. Concluding remarks summarizing the major implications of the paper are given in section 6.

2. Pearson's System of Frequency-Curves

The probability distributions contained in the system of curves proposed by Karl Pearson are found as the solutions of the differential equation

$$\frac{1}{\mathrm{y}}\frac{dy}{dx} = \frac{a - x}{b_0 + b_1 x + b_2 x^2} \tag{2.1}$$

where y in this context is the probability density function (PDF) evaluated at x. The motivation by Pearson (1895) of the derivation of the distributions from (2.1) is difficult, as well as difficult to access, and so we provide a brief and more direct overview of the portion of the derivation that is particularly relevant to the objectives of this paper.

2.1 Identifying Probability Distributions in the Family

Multiplying each side of (2.1) by $\mathrm{yx}^{\mathrm{n}}\left(b_0 + b_1 x + b_2 x^2\right)$, and then integrating with respect to x obtains

$$\int \mathrm{x}^{\mathrm{n}}(b_0 + b_1 x + b_2 x^2)\frac{dy}{dx}dx = \int y(a - x)x^n dx\,. \tag{2.2}$$

Integrating the left hand side of (2.2) by parts, treating $\frac{dy}{dx}$ as one part, and representing the right hand integral as the sum of two functions yields

$$x^n(b_0+b_1x+b_2x^2)\,y - \int (nx^{n-1}b_0+(n+1)b_1x^n+(n+2)b_2x^{n+1})\,y\,dx = \int ayx^n\,dx - \int yx^{n+1}\,dx.$$

If at the ends of the range of the curve the expression

$$x^n(b_0+b_1x+b_2x^2)\,y$$

vanishes, that is $x^n(b_0+b_1x+b_2x^2)\Big]_{x=r_1}^{r_2} = 0$ where r_1 and r_2 are the extremes of the range of variation for x, we have

$$-(n+1)b_1\mu'_n-(n+2)b_2\mu'_{n+1} = a\ \mu'_n-\mu'_{n+1} \qquad (2.3)$$

where μ'_n denotes the n[th] moment of x about the origin.

Examining the moment equation (2.3) for n = 0,1,2,3……..q respectively, we get q+1 equations to solve for a,b_0,b_1,b_2 in terms of the moments (μ'_r), r = 0,1,2,3……..q. The solution of these simultaneous equations results in the following representation of (2.1):

$$\frac{1}{y}\frac{dy}{dx} = \frac{\dfrac{\mu'_3(\mu'_4+3\mu'^2_2)}{10\mu'_2\mu'_4-18\mu'^3_2-12\mu'^2_3}-x}{\dfrac{\mu'_2(4\mu'_2\mu'_4-3\mu'^2_3)}{10\mu'_2\mu'_4-18\mu'^3_2-12\mu'^2_3}+\dfrac{\mu'_3(\mu'_4+3\mu'^2_2)}{10\mu'_2\mu'_4-18\mu'^3_2-12\mu'^2_3}x+\dfrac{2\mu'_2\mu'_4-3\mu'^2_3-6\mu'_2}{10\mu'_2\mu'_4-18\mu'^3_2-12\mu'^2_3}x^2}.$$

Defining $\psi_3{}^2=\beta_1=\dfrac{\mu'^2_3}{\mu'^3_2}$, $\psi_4 = \beta_2=\dfrac{\mu'_4}{\mu'^2_2}$ and $\omega=\dfrac{2\psi_4-3\psi_3{}^2-6}{\psi_4+3}$,

we are led to the following expression for the parameters a, b_0, b_1, b_2 contained in (2.1):

$$a=-\frac{\psi_3}{2(1+2\omega)} \qquad b_1=\frac{\psi_3}{2(1+2\omega)}$$

$$b_0=\frac{\omega+2}{2(1+2\omega)} \qquad b_2=\frac{\omega}{2(1+2\omega)}$$

which are valid when $-2<\omega<2$ and $\omega \neq -.5$ (please see appendix for details of the solution).

Turning now to the integration of (2.1) and the various forms of f(x) that arise, it is useful to note that:

1. $f(x) \geq 0$ over the range of values of x when the parameter adhere to the restrictions presented above.

2. $\int_{-\infty}^{+\infty} f(x)dx = 1$ must be satisfied.

3. It is sufficient to consider $\psi_3 \geq 0$ since the curve identified by ψ_3 = -c is a mirror reflection of the curve for ψ_3 = c with respect to the y-axis.

In general, there are three main types of Pearson curves and ten transition types . The various types are designed to handle limited or unlimited supports, as well as skewed or symmetric, bell, U, or J-shaped curves. The criteria identifying each type are expressed in terms of the parameters ψ_3 and ω. The mathematical derivation of the Pearson system of PDFs is given in the book 'Frequency Curves and Correlation' by W.P.Elderton. But the cumulative distributions of the Pearson system are not readily available, and are part of the contribution of this paper. The actual forms of the Pearson curves and their CDF's, restrictions on parameter ranges, and distribution supports are displayed in table (2.1).

Note that in table (2.1) we make use of the variables that are defined below.

$$r_1 = \frac{-\psi_3 + \sqrt{d}}{2\omega},\ r_2 = \frac{-\psi_3 - \sqrt{d}}{2\omega},\ d = \psi_3^2 - 4\omega(\omega+2),$$

$$m_1 = (\frac{(1+\omega)}{\omega})\frac{\psi_3}{\sqrt{d}} - (\frac{1+2\omega}{\omega}),\ m_2 = -(\frac{(1+\omega)}{\omega})\frac{\psi_3}{\sqrt{d}} - (\frac{1+2\omega}{\omega}),$$

$z = x_i \beta_j$, $t = -\frac{\sqrt{d}}{\omega}$, $M = -(\frac{1+2\omega}{\omega})$, $m = \frac{1+2\omega}{\omega}$,

$N = \frac{2}{\psi_3}$, $r = \frac{\psi_3}{2\omega}$, $s_1 = \frac{\sqrt{4\omega(\omega+2)}}{2\omega}$ $s = \frac{\sqrt{-d}}{2\omega}$, $r = \frac{-\psi_3}{2\omega}$, $v = -\frac{1+\omega}{\omega}\frac{\psi_3}{\sqrt{-d}}$.

Derivations of all tabled results are given in the appendix

Table 2.1. Frequency Curves and CDFs

No. of type	PDF	CDF	Criterion	Remarks	Calculations of constants and cdf's see page of appendix
MAIN TYPES I	f(x)= $\dfrac{(x-r_1)^{m_1}(r_2-x)^{m_2}}{\beta(m_1+1,m_2+1)(r_2-r_1)^{m_1+m_2+1}}$ $r_1<x<r_2$	F(z)= p(x $\leq$ z)= $\dfrac{\int_0^{\frac{z-r_1}{r_2-r_1}} v^{m_1}(1-v)^{m_2}\,dv}{\beta(m_1+1,m_2+1)}$ = IncompleteBeta(m_1+1, m_2+1, $\frac{z-r_1}{r_2-r_1}$)	$\psi_{3}\neq 0$, $-1<\omega<0$ [$\omega\neq -.5$, $(2+3\omega)\psi^2_3$ $\neq 4(1+2\omega)^2(2+\omega)$]	The roots r's are real and opposite signs ;limited Range; general form of beta-dist.; skew;when both M's are positive has a bell-shaped, when both M's are negative has a U-shaped, when M's are of opposite signs has J-shaped or twisted J-shaped.	
IV	f(x)= $\dfrac{s^{2m-1}}{\Gamma(2m-2,v)}\left[(x+r)^2+s^2\right]^{-m}\left(\dfrac{x+r-is}{x+r+is}\right)^{\frac{vi}{2}}$		$\psi_{3}\neq 0$, $\omega\succ 0$ $\psi^2_3\prec$ $4(\omega)(2+\omega)$	Unlimited range; Skew; bell-shaped; No common statistical Dist. are of type IV form;	

Continuation Table 2.1.

No. of type	PDF	CDF	Criterion	Remarks	Calculations of constants and cdf's see page of appendix
VI	f(x)= $\dfrac{(x-r_1)^{m_1}(x-r_2)^{m_2}}{\beta(m_1+1,-m_1-m_2-1)(r_1-r_2)^{m_1+m_2+1}}$ $r_1 < x < +\infty$	F(z)= p(x $\leq$ z)= 1 $-\dfrac{\int_0^{\frac{r_1-r_2}{z-r_2}} v^{-m_1-m_2-2}\ (1-v)^{m_2}\,dv}{\beta(m_1+1,-m_1-m_2-1)}$ =1-incomplete $\beta eta(m_1+1,-m_1-m_2-1,$ $\dfrac{r_1-r_2}{z-r_2})$	$\psi_{3\neq 0}$, $0<\omega<2/5$ $\psi_3^{\ 2}>4\omega(\omega+2)$, $(2+3\omega)\psi^2_3\neq$ $4(1+2\omega)^2(2+\omega)]$	The roots are real and same sign; unlimited range in one direction $r_1 < x < +\infty$; skew; bell-shaped if $m_1>0$; J-shaped if $m_1<0$; this is called beta prime dist.	
TRANSITION TYPES Normal	$f(x)=\dfrac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$ $-\infty < x < +\infty$	F(z)= p(x $\leq$ z)= $\int_{-\infty}^{z}\dfrac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}\,dx$	$\psi_3 = \omega = 0$	Unlimited range; symmetrical, bell-shaped.	

Continuation Table 2.1.

No. of type	PDF	CDF	Criterion	Remarks	Calculations of constants and cdf's see page of appendix
II	$f(x) = \dfrac{(t^2 - x^2)^M}{t^{2M+1}\beta(M+1,0.5)}$ $-t < x < t$	for z >= 0; F(z)=p(x $\leq$ z)= $1 - 0.5 \int_0^{1-\frac{z^2}{t^2}} beta(M+1,0.5)$ and for z < 0. F(z)= p(x $\leq$ z)= $0.5 \int_0^{1-\frac{z^2}{t^2}} beta(M+1,0.5)$	ψ_3 = 0, -1< ω < 0, and $\omega \neq$ 0.5	Limited range; a special case of type I; For –1< ω <-0.5 the curve is U-shaped, for –0.5< ω <0 the curve is bell-shaped, symmetrical..	

Continuation Table 2.1.

No. of type	PDF	CDF	Criterion	Remarks	Calculations of constants and cdf's see page of appendix
III	$f(x)=$ $\dfrac{e^{-N^2}N^{N^2}e^{-Nx}(N+x)^{N^2-1}}{\Gamma(N^2)}$ $-N<x<N$	F(z)= p(x $\leq$ z)= $\dfrac{1}{\Gamma(N^2)}\int_0^{N(N+z)} e^{-u}u^{N^2-1}du$ =incomplete gamma(N(N+z),N^2)	$\psi_3 \succ 0$ $\omega=0$	Unlimited range in one direction; for $N^2>1$ is a bell-shaped, for $N^2<1$ is a J-shaped; it is gamma dist.	
V	$f(\mathrm{x})=\dfrac{(2r(m-1))^{2m-1}}{\Gamma(2m-1)}$ $*(\mathrm{x}+r)^{-2m}e^{\frac{-2r(m-1)}{\mathrm{x}+r}}$ $-r<\mathrm{x}<+\infty$	F(z)= p(x $\leq$ z)= $1-\dfrac{\int_0^{\frac{2r(m-1)}{z+r}} w^{2m-2}e^{-w}dw}{\Gamma(2m-1)}$ =1-incomplete gamma(2m-1,2r(m-1)/z+r)	$\psi_{3>0,}$ $0<c<2/5$ $\psi^2{}_3=4\omega(2+\omega)$	Unlimited range in one direction; bell shaped; it is an inverse gaussian	

Continuation Table 2.1.

No. of type	PDF	CDF	Criterion	Remarks	Calculations of constants and cdf's see page of appendix
VII	$f(x)=\dfrac{s_1^{2m-1}(x^2+s_1^2)^{-m}}{\beta(m-0.5,0.5)}$ $-\infty < x < +\infty$	for z > 0; F(z)= p(x $\leq$ z)=1- 0.5incomplete beta($\dfrac{s_1^2}{s_1^2+z^2}$,m-0.5,0.5) and for z < 0; F(z)= p(x $\leq$ z)= 0.5incomplete beta($\dfrac{s_1^2}{s_1^2+z^2}$, m-0.5,0.5)	$\psi_3 = 0$ $\omega \succ 0$	Unlimited range; symmetrical, bell-shaped; special case of type IV; an important distribution belonging to this family is the (central) t-dist.	

Continuation Table 2.1.

No. of type	PDF	CDF	Criterion	Remarks	Calculations of constants and cdf's see page of appendix
VIII	$f(x)=\dfrac{(1-2m)(x-r_1)^{-2m}}{(r_2-r_1)^{1-2m}}$ $r_1<x<r_2$ $1-2m>0$	F(z)= p(x $\leq$ z)= $\dfrac{(z-r_1)^{1-2m}}{(r_2-r_1)^{1-2m}}$	$\psi_3>0,$ $\omega<-0.5,$ $(2+3\omega)\psi_3{}^2=$ $4(1+2\omega)^2(2+\omega)$	Limited range; J-shaped; special case of type I.	
IX	$f(x)=\dfrac{(1-2m)(r_2-x)^{-2m}}{(r_2-r_1)^{1-2m}}$ $r_1<x<r_2$	F(z)= p(x $\leq$ z)= $1-\dfrac{(r_2-z)^{1-2m}}{(r_2-r_1)^{1-2m}}$	$\psi_3>0,$ $-0.5<\omega<0,$ $(2+3\omega)\psi_3{}^2=$ $4(1+2\omega)^2(2+\omega)$	Limited range; J-shaped;	
X	$f(x)=e^{-(x+1)}$ $-1<x<+\infty$	$F(z)=1-e^{-(z+1)}$	$\psi_3{}^2=4,$ $\omega=0$	Limited range in one direction; J-shaped; special case of III.	

Continuation Table 2.1.

No. of type	PDF	CDF	Criterion	Remarks	Calculations of constants and cdf's see page of appendix
XI	$\mathrm{f(x)} = \frac{(2m-1)(x-r_2)^{-2m}}{(r_1-r_2)^{1-2m}}$ $r_1 < x < +\infty$	F(z)= p(x $\leq$ z)= $1 - \frac{(z-r_2)^{1-2m}}{(r_1-r_2)^{1-2m}}$	$\psi_3 > 0,$ $0 < \omega < 2/5,$ $(2+3\omega)\psi_3^{\ 2} =$ $4(1+2\omega)^2(2+\omega)$	Unlimited range in one direction; J-shaped.	
XII	$\mathrm{f(x)} = \frac{(\frac{-x+r_2}{x-r_1})^{m_2}}{\beta(m_2+1,-m_2+1)(-r_1+r_2)}$ $r_1 < x < r_2$		$\psi_3 \geq 0$ $\omega = 0$	Limited range; Twisted J-shaped; Special case of type I.	

2.2. Using Pearson Family Distributions to Model Multinomial Choice

Regarding the use of the Pearson system for specifying models of multinomial choice, first recall that the choice probabilities $P_{ij} = G_j(\mathbf{x}_{i.}, \beta) \in [0,1]$, j = 1,….J, must adhere to the adding up condition (1.2), as $\sum_{j=1}^{J} P_{ij} = \sum_{j=1}^{J} G_j(\mathbf{x}_{i.}, \beta) = 1$. In principle, any nonnegative valued function, $\vartheta_j(\mathbf{x}_i.\beta)$, can be used to define a legitimate specification of $G(\bullet)$ using the specification

$$P_{ij} = G_j(\mathbf{x}_{i.}, \beta) \equiv \frac{\vartheta_j(\mathbf{x}_{i.}, \beta)}{\sum_{j=1}^{J} \vartheta_j(\mathbf{x}_{i.}, \beta)}. \tag{2.4}$$

For example, the special case of the multinomial logit model follows upon setting $\vartheta_j(\mathbf{x}_{i.}, \beta) \equiv e^{\mathbf{x}_i.\beta_j}$. An alternative would be to let $\vartheta_j(\mathbf{x}_{i.}, \beta) = F_j(\mathbf{x}_{i.}, \beta)$, with $F_j(\bullet)$ being a member of the flexible family of curves contained in the Pearson system of CDF's. Letting $\xi = (\psi_3\ ,\ \omega)'$ denote the unknown parameters of the Pearson system, the choice of probability specification would then be

$$P_{ij} = G_j(\mathbf{x}_{i.}, \beta, \xi) \equiv \frac{F_j(\mathbf{x}_{i.}, \beta,\ \xi)}{\sum_{j=1}^{J} F_j(\mathbf{x}_{i.}, \beta,\ \xi)}. \tag{2.5}$$

3. Empirical Likelihood Problem Formulation and Solution

Consider the statistical model defined by (1.1), $y_{ij} = G_j(\mathbf{x}_{i.}, \beta) + \varepsilon_{ij}$. We will later assume that the functional form of $G(\bullet)$ is derived from (2.5) and one of the CDFs in the

Pearson system, and note that in any case $E[\varepsilon_{ij}|\mathbf{x}_{i.}] = 0$. In order to recover information relating to P_{ij} and β, we examine the general concept of empirical likelihood in the case of iid data. In this estimation context, the joint empirical probability mass function is of the form $\prod_{i=1}^{n} \delta_i$, where δ_i, for i = 1,…n, represent empirical probability or sample weights that are associated, respectively, with the n random sample vector outcomes of the multinomial choice process, $(y_{i1},...,y_{iJ})$, $i = 1,...,n.$. To define the value of the empirical likelihood function for θ, where $\theta \equiv \text{vec}([\beta_1, \beta_2, \beta_3,\ldots \beta_J])$ is a column-vectorized representation of model parameters with dimension ((KJ) $\times$ 1), the δ_i's are selected to maximize $\prod_{i=1}^{n} \delta_i$, subject to data-based constraints defined in terms of moment equations

$$E[\mathbf{h}(\mathbf{Y}_{i.},\mathbf{x}_{i.},\theta)] = E[(h_1(Y_{i1},\mathbf{x}_{i.},\theta), h_2(Y_{i2}, \mathbf{x}_{i.},\theta),\ldots h_J(Y_{iJ},\mathbf{x}_{i.},\theta))]' = [\mathbf{0}].$$

In the context of estimating equation parlance, $\mathbf{h}(\mathbf{Y}_{i.}, \mathbf{x}_{i.}, \theta)$ is a vector of unbiased estimating functions relating to the population random vector $\mathbf{Y}$. For now we are considering general forms of the moment equations, but later we will consider different conceptual formulations of the moment equations which are more specific to the multinomial choice model. An empirical representation of the moment condition based on empirical likelihood sample weights is given by $E_\delta[\mathbf{h}(\mathbf{Y},\mathbf{x},\theta] = \sum_{i=1}^{n} \delta_i \mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta) = \mathbf{0}$, where θ is as defined above , $\mathbf{Y}$ is interpreted to have the empirical distribution $\delta_i = P(\mathbf{Y} = \mathbf{y}_{i.})$, i = 1…….n, and $E_\delta[.]$ denotes an expectation taken with respect to the empirical probability distribution $\{\delta_1,\delta_2,.....\delta_n\}$.

The empirical likelihood maximization step chooses the joint empirical probability distribution $\prod_{i=1}^{n}\delta_i$ for $\mathbf{Y}$ that assigns the maximum possible probability to the sample outcomes $\mathbf{y}$ actually observed, subject to constraints provided by the empirical moment equations. The constraints serve to introduce θ into the estimation problem. The empirical likelihood function for θ can be defined by maximizing the empirical likelihood *conditional* on the value of θ, and then substituting the constrained maximum value of each δ_i, say $\hat{\delta}_i(\boldsymbol{\theta};\mathbf{y})$ into $\prod_{i=1}^{n}\delta_i$, yielding a function of θ as $L_{EL}(\theta;\mathbf{y})$ = $\prod_{i=1}^{n}\hat{\delta}_i(\boldsymbol{\theta};\mathbf{y})$. At this point, the empirical likelihood function operates like an ordinary parametric likelihood function for estimation and inference purposes as long as the estimating functions are unbiased and thus have zero expectations, have a finite variances, and are based on independent or weakly dependent data observations. In particular, maximizing $L_{EL}(\theta;\mathbf{y})$ through choice of θ, defines the maximum empirical likelihood (MEL) estimator of the parameter vector θ.

In the sections ahead we provide more details relating to the EL procedure in the iid case ahead, which will further motivate the general concepts involved and will also serve to define additional notation. In particular, we provide details regarding how one utilizes the EL concept to perform maximum empirical likelihood (MEL) estimation of parameters for the statistical model $y_{ij} = G_j(\mathbf{x}_{i.},\beta) + \varepsilon_{ij}$ in (1.5). We also discuss how to test hypotheses and generate confidence regions and bounds based on the EL function, including the use of the generalized empirical likelihood ratio (GELR) for inference purposes. Finally, we extended the EL principle to the case where the data are independent but not identically distributed. We then specialize the formulation to provide estimates of the parameters of the multinomial choice model in section 4.

3.1. Nonparametric Likelihood Functions

Consider the inverse problem of using a random sample outcome $\mathbf{y}=\left(y_{11},...,y_{1J},...,y_{n1},...,y_{nJ}\right) = (\mathbf{y}_{1.},.....\mathbf{y}_{n.})'$ to recover an estimate of the PDF of $\mathbf{Y}$. In this nonparametric setting, a nonparametric likelihood function can be defined whose arguments are not parameters but entire probability densities or mass functions as

$$\mathrm{L}(\mathrm{f};\mathbf{y}) = \prod_{i=1}^{n} \mathrm{f}(\mathbf{y}_{i.}) \quad j = 1\ldots\ldots J \tag{3.1.1}$$

The nonparametric maximum likelihood (NPML) estimate of $\mathrm{f}(\mathbf{y}_{i.})$ is defined by

$$\hat{\mathrm{f}}(\mathbf{y}) = \arg\max_{\mathrm{f}} [\mathbf{L}(\mathrm{f};\mathbf{y})] = \hat{\mathrm{f}}(\mathbf{y}) = \arg\max_{\mathrm{f}} [\prod_{i=1}^{n} \mathrm{f}(\mathbf{y}_{i.})\,]. \tag{3.1.2}$$

The solution to (3.1.2) defines an empirical probability mass function of the multinomial type that represents discrete probability masses assigned to each of the finite number of observed sample outcomes, where $\delta_i = \mathrm{f}(\mathbf{y}_{i.}) > 0 \ \forall\, \mathrm{i}$.

The preceding maximum likelihood problem (3.1.1) and (3.1.2) can be represented as a nonparametric maximum likelihood problem of finding the optimal choice of δ_i 's in a multinomial-based likelihood function, as

$$\hat{\boldsymbol{\delta}} = \left[\hat{\delta}_1,\hat{\delta}_2,.......,\hat{\delta}_n\right]' = \arg\max_{\delta}[\prod_{i=1}^{n}\delta_i] = \arg\max_{\delta}[\sum_{i=1}^{n}\ln(\delta_i)]\,. \tag{3.1.3}$$

If the δ_i 's are unrestricted in value, (3.1.3) will have no solution since the objective function would be unbounded, and so a normalization condition on the δ_i 's is imposed.

In the case at hand, the constraint $\sum_{i=1}^{n} \delta_i = 1$ is a natural normalization condition on the δ_i's, along with nonnegativity.

3.2 Empirical Likelihood Function for θ

The likelihood (3.1.1) is devoid of the parameter vector θ and so it cannot be used to distinguish likely from unlikely values of a parameter vector θ. Linkage between the data, $\mathbf{y} = (y_{11},...y_{1J},...,y_{n1},...y_{nJ})' = (\mathbf{y}_{1.},.....\mathbf{y}_{n.})'$, the population distribution F(**y**), and the parameter of interest, θ, is accomplished through the use of unbiased estimating functions to define estimating equation constraints on the NPML problem. Information about θ is conveyed by the estimating function in expectation or moment form E[**h**(**Y**, **x**, θ)] = **0**, which defines constraints on the NPML problem that generates the empirical likelihood function. Given that the expectation is unknown because F(**y**) is unknown, an estimated empirical probability distribution is applied to observed sample outcomes of **h**(**Y**, **x**, θ), to define an empirical expectation $\sum_{i=1}^{n} \delta_i \mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta) = \mathbf{0}$ that approximates E[**h**(**Y**, **x**, θ)] = **0** and that can be used in forming an empirical moment equation. The system of m equations $\mathbf{h}_{EL}(\mathbf{Y}, \mathbf{x}, \theta) \equiv \sum_{i=1}^{n} \delta_i \mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta) = \mathbf{0}$, when viewed in the context of estimating equations for θ, is generally underdetermined, just-determined, or over-determine for identifying a ((KJ) × 1) vector θ, depending on whether m < , =, or > KJ, respectively. The choice of the unknown δ_i's is solved by maximizing the empirical likelihood objective function, and in the process, the estimating equations are reconciled to yield a solution for θ (assuming a feasible solution exists).

The log-empirical likelihood function for θ is defined as

$$\ln[L_{EL}(\theta\ ;\mathbf{y})] \equiv \max_{\delta}\left[\sum_{i=1}^{n}\ln(\delta_i) \text{ s.t. } \sum_{i=1}^{n}\delta_i\mathbf{h}(y_{i.},\mathbf{x}_{i.},\theta) = \mathbf{0} \text{ and } \sum_{i=1}^{n}\delta_i = 1\right], \qquad (3.2.1)$$

Imposing both the normalization condition on the δ_i 's and the empirical moment constraints, the solution to the problem of finding the NPML estimate of ln(f(**y**)) is thus defined in terms of the choice of nonnegative δ_i 's that maximize $\sum_{i=1}^{n}\ln(\delta_i)$ subject to the constraints $\sum_{i=1}^{n}\delta_i = 1$ and $\sum_{i=1}^{n}\delta_i\mathbf{h}(y_{i.},\mathbf{x}_{i.},\theta) = \mathbf{0}$. The Lagrange function associated with the constrained optimization problem is given by

$$L(\delta,\eta,\lambda) \equiv \left[\sum_{i=1}^{n}\ln(\delta_i) - \eta\Big(\sum_{i=1}^{n}\delta_i - 1\Big) - \lambda'\sum_{i=1}^{n}\delta_i\mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta)\right]. \qquad (3.2.2)$$

Solving for the optimal δ, η and λ in the Lagrange form of the problem (3.2.2) and then substituting optimal values for δ into the objective function of the maximization problem in (3.2.2), a specific functional form for the EL function in terms of θ can be defined. In particular, first note that the first-order conditions with respect to the δ_i 's are

$$\frac{\partial \ln L(\delta,\eta,\lambda)}{\partial \delta_i} = \frac{1}{n}\frac{1}{\delta_i} - \sum_{j=1}^{J}\lambda_j h_j(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta) - \eta = \mathbf{0}, \ \forall i. \qquad (3.2.3)$$

Also, from the equality $\sum_{i=1}^{n}\delta_i\frac{\partial \ln L(\delta,\eta,\lambda)}{\partial \delta_i} = 0$ and $E_{\boldsymbol{\delta}}[\mathbf{h}(\mathbf{Y}, \mathbf{x}, \theta)] =$

$\sum_{i=1}^{n}\delta_i\mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta) = \mathbf{0}$ it follows that

$$\sum_{i=1}^{n}\delta_i\frac{\partial \ln L(\boldsymbol{\delta},\eta,\boldsymbol{\lambda})}{\partial \delta_i} = \frac{1}{n}n - \eta = 0, \qquad (3.2.4)$$

and thus η = 1. The resulting unique optimal δ_i weights implied by (3.2.3) can be then be expressed as the following function of θ and λ,

$$\delta_i(\theta,\lambda) = \left[n\left(\sum_{j=1}^{J} \lambda_j h_j(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) + 1 \right) \right]^{-1} . \tag{3.2.5}$$

Substituting (3.2.5) into the empirical moment equations $\sum_{i=1}^{n} \delta_i \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) = \mathbf{0}$ produces a system of equations that λ must satisfy as follows:

$$\sum_{i=1}^{n} \delta_i \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) = \sum_{i=1}^{n} n^{-1} \left[\left(\sum_{j=1}^{J} \lambda_j h_j(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) + 1 \right) \right]^{-1} \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) = \mathbf{0} . \tag{3.2.6}$$

Under regularity conditions, Qin and Lawless (1994,pp.304-5) show that a well-defined solution for λ in (3.2.6) exists. However, the solution $\lambda(\theta)$ is only an implicit function of θ, which we denote in general by

$$\lambda(\theta) = \arg_{\delta} \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{1+\lambda' \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta)} \right) \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) = \mathbf{0} \right]. \tag{3.2.7}$$

The solution $\lambda(\theta)$ is continuous and differentiable in θ under regularity conditions.

Substituting the optimal Lagrangian multiplier values $\lambda(\theta)$ into (3.2.5) allows the empirical probabilities to be represented in terms of θ as $\delta_i(\theta) \equiv \delta_i[\theta, \lambda(\theta)]$ $= \left[n\left(\sum_{j=1}^{J} \lambda_j(\theta) h_j(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) + 1 \right) \right]^{-1}$. Then, substitution of the optimal $\delta(\theta)$ values into the (unscaled) objective function $\sum_{i=1}^{n} \ln(\delta_i)$ in (3.2.1) yields the expression for the log-empirical likelihood function evaluated at θ given by

$$\mathrm{Ln}[L_{EL}(\theta; \mathbf{y})] = -\sum_{i=1}^{n} \ln(n[1+\lambda(\theta)' \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta)]) . \tag{3.2.8}$$

3.3 Maximum Empirical Likelihood Estimator

We can define a maximum empirical likelihood (MEL) estimator for θ by choosing the value of θ that maximizes the empirical likelihood function (3.2.1), or equivalently maximizes the logarithm of the EL function as follows:

$$\hat{\theta} = \arg\max_{\theta}\left[\ln(L_{EL}(\theta, \mathbf{Y}, \mathbf{x}))\right]. \tag{3.3.1}$$

The MEL estimator, $\hat{\theta}_{EL}$, is an extremum estimator whose solution is not generally obtainable in closed form because the $\lambda(\theta)$ of the EL function (recall (3.2.7)) is not a closed-form function of θ, and thus numerical optimization techniques are most often required to obtain outcomes of the MEL estimator. We could also obtain the MEL estimate of θ as the solution $\hat{\theta}_{EL}$ to the system of equations

$$\mathbf{h}_{EL}(\mathbf{y}, \mathbf{x}, \theta) = E_{\delta}\mathbf{h}(\mathbf{y}, \mathbf{x}, \theta) = \sum_{i=1}^{n} \hat{\delta}_i \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) = \mathbf{0}$$

where

$$\hat{\delta}_i(\hat{\theta}_{EL}) \equiv \delta_i[\hat{\theta}_{EL}, \lambda(\hat{\theta}_{EL})] = \left[n\left(\sum_{j=1}^{J} \lambda_j(\hat{\theta}_{EL}) h_j(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \hat{\theta}_{EL}) + 1 \right) \right]^{-1}, \tag{3.3.2}$$

for i = 1,…,n. Therefore, the MEL method of estimation can be viewed as a procedure for combining the set of estimating functions $\mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta)$, i = 1,…….n, into a vector-estimating equation $\mathbf{h}_{EL}(\mathbf{y}, \mathbf{x}, \theta)$ that can be solved for an estimate of θ.

Qin and Lawless (1994) show that the usual consistency and asymptotic normality properties of extremum estimators hold for the MEL estimator under regularity conditions related to the twice continuous differentiability of $\mathbf{h}(\mathbf{y}, \mathbf{x}, \theta)$ with respect to θ

and the boundedness of **h** and its first and second derivatives, all in a neighborhood of the true parameter value θ_0 . They also assume that the row rank of

$E\left[\left.\frac{\partial \mathbf{h}(\mathbf{y},\mathbf{x},\theta)}{\partial\theta}\right|_{\theta_0}\right]$ equals the number of parameters in the vector θ (Qin and Lawless,1994,p.305-6). These conditions lead to the MEL estimator's being consistent and asymptotically normal with limiting distribution

$$n^{\frac{1}{2}}\left(\hat{\theta}_{EL} - \theta_0\right) \xrightarrow{d} N(\mathbf{0},\Sigma) \tag{3.3.3}$$

where

$$\Sigma = \left[E\left[\left.\frac{\partial \mathbf{h}(\mathbf{Y},\mathbf{x},\theta)}{\partial\theta}\right|_{\theta_0}\right]\left[E\left[\mathbf{h}(\mathbf{Y},\mathbf{x},\theta)\,\mathbf{h}(\mathbf{Y},\mathbf{x},\theta)'\Big|_{\theta_0}\right]\right]^{-1} E\left[\left.\frac{\partial \mathbf{h}(\mathbf{Y},\mathbf{x},\theta)}{\partial\theta'}\right|_{\theta_0}\right]\right]^{-1}. \tag{3.3.4}$$

The covariance matrix Σ of the limiting normal distribution can be consistently estimated by

$$\hat{\Sigma} = \left[\left[\sum_{i=1}^{n}\hat{\delta}_i \left.\frac{\partial \mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta)}{\partial\theta}\right|_{\hat{\theta}_{EL}}\right]\left[\sum_{i=1}^{n}\hat{\delta}_i \mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\hat{\theta}_{EL})\mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\hat{\theta}_{EL})'\right]^{-1} \times \left[\sum_{i=1}^{n}\hat{\delta}_i \left.\frac{\partial \mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta)}{\partial\theta}\right|_{\hat{\theta}_{EL}}\right]'\right]^{-1}, \tag{3.3.5}$$

where the $\hat{\delta}_i$'s are the same as defined via (3.3.2). By substituting n^{-1} for $\hat{\delta}_i$'s in (3.3.5), an alternative consistent estimate is defined, which amounts to applying probability weights based on the empirical distribution function instead of the empirical probability weights generated by the empirical likelihood. The $\hat{\delta}_i$ probability weight estimates obtained from the EL procedure would be generally more efficient in finite samples if the estimating function information is unbiased. The normal limiting distribution of $\hat{\theta}_{EL}$ allows asymptotic hypothesis tests and confidence regions to be constructed.

3.4 Optimal Estimating Functions

An optimal estimating function is an unbiased estimating function having the smallest covariance matrix. Godambe (1960) was the first to suggest that the vector estimating function be standardized as

$$\mathbf{h}_s(\mathbf{Y}, \mathbf{x}, \theta) = \left[E_\theta \left[\frac{\partial \mathbf{h}(\mathbf{Y}, \mathbf{x}, \theta)}{\partial \theta} \right] \right]^{-1} \mathbf{h}(\mathbf{Y}, \mathbf{x}, \theta) \tag{3.4.1}$$

so that the multivariate optimal estimating function, or OptEF, is then the unbiased estimating function that minimizes , in the sense of symmetric positive definite matrix comparisons, the covariance matrix

$$\begin{aligned} \operatorname{cov}\left[\mathbf{h}_s(\mathbf{Y}, \mathbf{x}, \theta)\right] = & \left[E_\theta \left[\frac{\partial \mathbf{h}(\mathbf{Y}, \mathbf{x}, \theta)}{\partial \theta} \right] \right]^{-1} E_\theta \left[\mathbf{h}(\mathbf{Y}, \mathbf{x}, \theta)\, \mathbf{h}(\mathbf{Y}, \mathbf{x}, \theta)' \right] \\ & \times \left[E_\theta \left[\frac{\partial \mathbf{h}(\mathbf{Y}, \mathbf{x}, \theta)}{\partial \theta'} \right] \right]^{-1} . \end{aligned} \tag{3.4.2}$$

In the special case in which $\mathbf{h}(\mathbf{y}, \mathbf{x}, \theta)$ is actually proportional to, or a scaled version of the log of the score or gradient vector function corresponding to a genuine likelihood function, it follows under the standard regularity conditions applied to maximum likelihood estimation that

$$-E_\theta \left[\frac{\partial \mathbf{h}(\mathbf{Y}, \mathbf{x}, \theta)}{\partial \theta} \right] \propto E_\theta \left[-\frac{\partial^2 \ln L(\mathbf{Y}, \mathbf{x}, \theta)}{\partial \theta \partial \theta'} \right] \tag{3.4.3}$$

and

$$E_\theta\left[\mathbf{h}(\mathbf{Y},\mathbf{x},\theta)\,\mathbf{h}(\mathbf{Y},\mathbf{x},\theta)'\right] \propto E_\theta\left[\frac{\partial\ \ln L(\mathbf{Y},\mathbf{x},\theta)}{\partial\theta}\frac{\partial\ \ln L(\mathbf{Y},\mathbf{x},\theta)}{\partial\theta'}\right], \tag{3.4.4}$$

where the expectations on the right-hand sides of (3.4.3) and (3.4.4) are equal. In this case (3.4.2) becomes

$$\text{cov}(\mathbf{h}_s(\mathbf{Y},\mathbf{x},\theta)) = \left[-E_\theta\left[\frac{\partial^2 \ln L(\mathbf{Y},\mathbf{x},\theta)}{\partial\theta\partial\theta'}\right]\right]^{-1}, \tag{3.4.5}$$

which is recognized as the usual ML covariance matrix and the CRLB for estimating the parameter vector θ . This provides an OptEF finite sample justification for ML estimation in the case of estimating a vector of parameter θ and is analogous to the Gauss-Markov theorem justification for LS estimation.

The EL empirical moment constraints defined in terms of the conditional-on- θ optimum empirical probability weights are given by

$$\mathbf{h}_{EL}(\mathbf{Y},\mathbf{x},\theta) \equiv \hat{E}_\delta\left[\mathbf{h}(\mathbf{Y},\mathbf{x},\theta)\right] \equiv \sum_{i=1}^{n}\hat{\delta}_i(\theta,\mathbf{Y},\mathbf{x})\mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta) = \mathbf{0}$$

and these empirical moment constraints can be interpreted as vector estimating equations. The EL provides a method for forming a convex combination of n (m $\times$ 1) estimating functions, $\mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta)$, for i = 1,…,n. Thus, we want to investigate whether a particular combination of the n estimating functions used in the MEL approach is in some sense the best combination. Consider the class of estimation procedures that can be defined by a combination of the estimating equation information as

$$\mathbf{h}_\tau(\mathbf{Y},\mathbf{x},\theta) \equiv \sum_{i=1}^{n}\tau(\mathbf{x},\theta)\mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta) = \mathbf{0} \tag{3.4.6}$$

where $\boldsymbol{\tau}(\mathbf{x},\theta)$ is a ((KJ) $\times$ m) real-valued function such that the ((KJ) $\times$ 1) vector equation $\mathbf{h}_\tau(\mathbf{Y},\mathbf{x},\theta) = \mathbf{0}$ can be solved for the (K$\times$ 1) vector θ as $\hat{\theta}_\tau(\mathbf{y})$.

McCullagh and Nelder (1989,p.341) show that the optimal choice of τ, in the sense of defining a consistent estimator with minimum asymptotic covariance matrix in the class of estimators for θ defined as solutions to (3.4.6), is given by

$$\tau(\mathbf{x},\theta) = \mathrm{E}\left[\frac{\partial \mathbf{h}(\mathbf{y},\mathbf{x},\theta)}{\partial\theta}\right]\left[\mathrm{cov}(\mathbf{h}(\mathbf{y},\mathbf{x},\theta))\right]^{-1},$$

(3.4.7)

where **Y** denotes the random variable whose probability distribution is the common population distribution of the $\mathbf{Y}_i$'s. In case the **Y**'s are independent, but not identically distributed, we have

$$\tau_i(\mathbf{x}_{i.},\theta) = \mathrm{E}\left[\frac{\partial \mathbf{h}(y_{i.},\mathbf{x}_{i.},\theta)}{\partial\theta}\right]\left[\mathrm{cov}(\mathbf{h}(y_{i.},\mathbf{x}_{i.},\theta))\right]^{-1}$$

(3.4.8)

and $\mathbf{h}_\tau(\mathbf{Y},\mathbf{x},\theta) \equiv \sum_{i=1}^{n} \tau_i(\mathbf{x}_{i.},\theta)\mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta) = \mathbf{0}$. Using the optimal definition of τ_i in (3.4.8) defines an estimator for θ that has precisely the same asymptotic covariance matrix as the MEL estimator (3.3.4) because, given the unbiased nature of the estimating equations, $\mathrm{cov}\left[\mathbf{h}(y_{i.},\mathbf{x}_{i.},\theta)\right] = \mathrm{E}_\theta\left[\mathbf{h}(y_{i.},\mathbf{x}_{i.},\theta)\,\mathbf{h}(y_{i.},\mathbf{x}_{i.},\theta)'\right]$ (McCullagh and Nelder,1989, p.341).

4. EL Estimation in the Multinomial Choice Problem

In this section we examine an extended illustrative example demonstrating the setup of the MEL approach to estimating the parameters of a multinomial choice model. In this application, the form of the unbiased estimating functions for θ is given by

$$\mathbf{h}_{\tau_i}(\mathbf{Y}, \mathbf{x}, \theta) \equiv \sum_{i=1}^{n} \boldsymbol{\tau}_i(\mathbf{x}_{i.}, \theta)\mathbf{h}(y_{i.}, \mathbf{x}_{i.}, \theta) \tag{4.1.1}$$

where

$$\mathbf{h}(y_{i.}, \mathbf{x}_{i.}, \theta) = \begin{bmatrix} \begin{pmatrix} y_{i1} - G_1(\mathbf{x}_{i.}, \beta) \\ y_{i2} - G_2(\mathbf{x}_{i.}, \beta) \\ y_{i3} - G_3(\mathbf{x}_{i.}, \beta) \\ . \\ . \\ y_{iJ} - G_J(\mathbf{x}_{i.}, \beta) \end{pmatrix} \odot \mathbf{x}_{i.}' \\ y_{i1}{}^2 - 2y_{i1}G_1(\mathbf{x}_{i.}, \beta) + 2G_1{}^2(\mathbf{x}_{i.}, \beta) - G_1(\mathbf{x}_{i.}, \beta) \\ y_{i2}{}^2 - 2y_{i2}G_2(\mathbf{x}_{i.}, \beta) + 2G_2{}^2(\mathbf{x}_{i.}, \beta) - G_2(\mathbf{x}_{i.}, \beta) \\ y_{i3}{}^2 - 2y_{i3}G_3(\mathbf{x}_{i.}, \beta) + 2G_3{}^2(\mathbf{x}_{i.}, \beta) - G_3(\mathbf{x}_{i.}, \beta) \\ . \\ . \\ y_{iJ}{}^2 - 2y_{iJ}G_J(\mathbf{x}_{i.}, \beta) + 2G_J{}^2(\mathbf{x}_{i.}, \beta) - G_J(\mathbf{x}_{i.}, \beta) \end{bmatrix} \tag{4.1.2}$$

and $\mathbf{h}(\mathrm{y}_{i.}, \mathbf{x}_{i.}, \theta)$ is a vector of dimension ((KJ+J) × 1), and recall that $\odot$ denotes the Hadamard (elementwise) product. Taking into consideration the adding up condition in (1.2), which implies that there are redundant moment equations among the (KJ+J) equations, we reformulate (4.1.2) and represent $\mathbf{h}(\mathrm{y}_{i.}, \mathbf{x}_{i.}, \theta)$ with dimension ((K(J-1)+(J-1)) × 1) as follow

$$\mathbf{h}(\mathrm{y}_{i.}, \mathbf{x}_{i.}, \theta) = \begin{bmatrix} \begin{pmatrix} \mathrm{y}_{i2} - G_2(\mathbf{x}_{i.}, \beta) \\ \mathrm{y}_{i3} - G_3(\mathbf{x}_{i.}, \beta) \\ \mathrm{y}_{i4} - G_4(\mathbf{x}_{i.}, \beta) \\ . \\ . \\ \mathrm{y}_{iJ} - G_J(\mathbf{x}_{i.}, \beta) \end{pmatrix} \odot \mathbf{x}_{i.}' \\ \mathrm{y}_{i2}^2 - 2\mathrm{y}_{i2} G_2(\mathbf{x}_{i.}, \beta) + 2G_2^2(\mathbf{x}_{i.}, \beta) - G_2(\mathbf{x}_{i.}, \beta) \\ \mathrm{y}_{i3}^2 - 2\mathrm{y}_{i3} G_3(\mathbf{x}_{i.}, \beta) + 2G_3^2(\mathbf{x}_{i.}, \beta) - G_3(\mathbf{x}_{i.}, \beta) \\ \mathrm{y}_{i4}^2 - 2\mathrm{y}_{i4} G_4(\mathbf{x}_{i.}, \beta) + 2G_4^2(\mathbf{x}_{i.}, \beta) - G_4(\mathbf{x}_{i.}, \beta) \\ . \\ . \\ \mathrm{y}_{iJ}^2 - 2\mathrm{y}_{iJ} G_J(\mathbf{x}_{i.}, \beta) + 2G_J^2(\mathbf{x}_{i.}, \beta) - G_J(\mathbf{x}_{i.}, \beta) \end{bmatrix} \tag{4.1.3}$$

and $G_j(\mathbf{x}_{i.}, \beta)$ denotes the conditional expectation of $\mathbf{Y}_{ij}$ given $\mathbf{x}_{i.}$, $G_j(\mathbf{x}_{i.}, \beta) = \mathrm{E}(\mathrm{y}_{ij} | \mathbf{x}_{i.})$. In the context of multinomial choice problem, $G_j(\mathbf{x}_{i.}, \beta)$ denotes the conditional-on- $\mathbf{x}_{i.}$ probability of choosing alternative j for observation i. For the sake of expositional clarity, we henceforth consider the special case of the multinomial logit model upon setting

$$G_j = G_j(\mathbf{x}_{i.},\beta) = \frac{1}{1+\sum_{k=2}^{J} e^{\mathbf{x}_{i.}\beta_k}} \qquad \text{for } j = 1$$

$$= \frac{e^{\mathbf{x}_{i.}\beta_j}}{1+\sum_{k=2}^{J} e^{\mathbf{x}_{i.}\beta_k}} \qquad \text{for } j = 2,3,\ldots..J \qquad (4.1.4)$$

We emphasize that $G_j(\mathbf{x}_{i.},\beta)$ could be any link function of flexible form that satisfies (1.2) and that defines a legitimate multinomial response model globally. Later we will consider $G_j(\mathbf{x}_{i.},\beta)$ as being formed from CDFs in the Pearson system, which themselves satisfy (1.2).

The OptEF estimator is in the general class of estimating equations based on the estimating functions of the form (4.1.1) characterized by the solution to

$$\mathbf{h}_{\mathbf{Opt}}(\mathbf{y}_{i.}, \mathbf{x}_{i.},\theta) = \sum_{i=1}^{N}\left[E\left(\frac{\partial\, \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.},\theta)}{\partial\theta}\right)(\Phi_i)^{-1}\, \mathbf{h}((\mathbf{y}_{i.}, \mathbf{x}_{i.},\theta))\right] = \mathbf{0} \qquad (4.1.5)$$

where $E\left(\frac{\partial\, \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.},\theta)}{\partial\theta}\right)$ is a matrix of dimension ((K(J-1))$\times$ (K(J-1)+(J-1))) as

$$E\left(\frac{\partial\, \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.},\theta)}{\partial\theta}\right) =$$

$$\begin{bmatrix} \frac{-\partial[G_2\mathbf{x}'_{i.}]}{\partial\beta_2} & \frac{-\partial[G_3\mathbf{x}'_{i.}]}{\partial\beta_2} & \frac{-\partial[G_4\mathbf{x}'_{i.}]}{\partial\beta_2} & \frac{\partial G_2}{\partial\beta_2}[2G_2-1] & \frac{\partial G_3}{\partial\beta_2}[2G_3-1] & \frac{\partial G_4}{\partial\beta_2}[2G_4-1] \\ \frac{-\partial[G_2\mathbf{x}'_{i.}]}{\partial\beta_3} & \frac{-\partial[G_3\mathbf{x}'_{i.}]}{\partial\beta_3} & \frac{-\partial[G_4\mathbf{x}'_{i.}]}{\partial\beta_3} & \frac{\partial G_2}{\partial\beta_3}[2G_2-1] & \frac{\partial G_3}{\partial\beta_3}[2G_3-1] & \frac{\partial G_4}{\partial\beta_3}[2G_4-1] \\ \frac{-\partial[G_2\mathbf{x}'_{i.}]}{\partial\beta_4} & \frac{-\partial[G_3\mathbf{x}'_{i.}]}{\partial\beta_4} & \frac{-\partial[G_4\mathbf{x}'_{i.}]}{\partial\beta_4} & \frac{\partial G_2}{\partial\beta_4}[2G_2-1] & \frac{\partial G_3}{\partial\beta_4}[2G_3-1] & \frac{\partial G_4}{\partial\beta_4}[2G_4-1] \end{bmatrix} \qquad (4.1.6)$$

where

$$\frac{\partial G_j}{\partial \beta_j} = \frac{\partial G_j(\mathbf{x}_{i.},\beta)}{\partial \beta_j} = \mathbf{x}_{i.}' \left(\frac{e^{\mathbf{x}_{i.}\beta_j}}{1+\sum_{k=2}^{J} e^{\mathbf{x}_{i.}\beta_k}} \right)\left(1 - \frac{e^{\mathbf{x}_{i.}\beta_j}}{1+\sum_{k=2}^{J} e^{\mathbf{x}_{i.}\beta_k}} \right) =$$

$$= \mathbf{x}_{i.}' G_j(1-G_j) \;\forall\; j \geq 2 \tag{4.1.7}$$

$$\frac{\partial G_j}{\partial \beta_k} = \frac{\partial G_j(\mathbf{x}_{i.},\beta)}{\partial \beta_k} = -\mathbf{x}_{i.}' G_j G_k \quad \text{for } k \neq j \text{ and } k>1$$

$$\frac{\partial \left[G_j \mathbf{x}_{i.}' \right]}{\partial \beta_j} = \mathbf{x}_{i.}\mathbf{x}_{i.}' G_j(1-G_j) \;\forall\; j \geq 2$$

$$= \quad 0 \qquad\qquad \text{for } j=1 \tag{4.1.8}$$

$$\frac{\partial \left[G_j \mathbf{x}_{i.}' \right]}{\partial \beta_k} = -\mathbf{x}_{i.}\mathbf{x}_{i.}' G_j G_k \qquad \text{for } k \neq j \text{ and } k > 1$$

and $\Phi_i = \mathbf{cov}\left(\mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta)\right) = E_\theta\left[\mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta)\, \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta)'\right]$ is the covariance matrix of ($Y_{ij}|\mathbf{x}_{i.}$) having dimension ((K(J-1)+(J-1) × (K(J-1)+(J-1)). Given the 0-1 dichotomous outcomes of the Y_{ij}'s, note that

$E(Y_{ij}^n) = G_j(\mathbf{x}_{i.},\beta)$ for every positive integer n.

Also, given that the Y_{ij}'s must sum to 1, it follows that

$E(Y_{ij}^n\, Y_{ik}^m) = 0$ for every m and n positive integers greater than one

$= -G_j(\mathbf{x}_{i.},\beta)\, G_k(\mathbf{x}_{i.},\beta)$ for m=n=1.

In order to define the OptEF, note that the covariance matrix of any [Y_{i1},.. Y_{iJ}, $(Y_{i1}-G_1)^2$, … $(Y_{iJ}-G_J)^2$] vector in this multinomial choice problem is given by

$$\mathbf{\Phi}_{\mathrm{i}} =$$

$$\begin{bmatrix}
G_2(1-G_2)\mathbf{x}_{i.}\mathbf{x}'_{i.} & -2G_2G_3\mathbf{x}_{i.}\mathbf{x}'_{i.} & -2G_2G_4\mathbf{x}_{i.}\mathbf{x}'_{i.} & G_2(1-G_2)(1-2G_2)\mathbf{x}'_{i.} & -G_2G_3(1\text{-}4G_3)\mathbf{x}'_{i.} & -G_2G_4(1\text{-}4G_4)\mathbf{x}'_{i.} \\
-2G_2G_3\mathbf{x}_{i.}\mathbf{x}'_{i.} & G_3(1-G_3)\mathbf{x}_{i.}\mathbf{x}'_{i.} & -2G_3G_4\mathbf{x}_{i.}\mathbf{x}'_{i.} & -G_2G_3(1\text{-}4G_2)\mathbf{x}'_{i.} & G_3(1-G_3)(1-2G_3)\mathbf{x}'_{i.} & -G_3G_4(1\text{-}4G_4)\mathbf{x}'_{i.} \\
-2G_2G_4\mathbf{x}_{i.}\mathbf{x}'_{i.} & -2G_3G_4\mathbf{x}_{i.}\mathbf{x}'_{i.} & G_4(1-G_4)\mathbf{x}_{i.}\mathbf{x}'_{i.} & -G_2G_4(1\text{-}4G_2)\mathbf{x}'_{i.} & -G_3G_4(1\text{-}4G_3)\mathbf{x}'_{i.} & G_4(1-G_4)(1-2G_4)\mathbf{x}'_{i.} \\
G_2(1-G_2)(1-2G_2)\mathbf{x}'_{i.} & -G_2G_3(1\text{-}4G_2)\mathbf{x}'_{i.} & -G_2G_4(1\text{-}4G_2)\mathbf{x}'_{i.} & G_2(1-G_2)(2G_2-1)^2 & G_2G_3[2G_2(1-4G_3)-(1-2G_3)] & G_2G_4[2G_4(1-4G_2)-(1-2G_2)] \\
-G_2G_3(1\text{-}4G_3)\mathbf{x}'_{i.} & G_3(1-G_3)(1-2G_3)\mathbf{x}'_{i.} & -G_2G_4(1\text{-}4G_2)\mathbf{x}'_{i.} & G_2G_3[2G_2(1-4G_3)-(1-2G_3)] & G_3(1-G_3)(2G_3-1)^2 & G_3G_4[2G_4(1-4G_3)-(1-2G_3)] \\
-G_2G_4(1\text{-}4G_4)\mathbf{x}'_{i.} & -G_3G_4(1\text{-}4G_4)\mathbf{x}'_{i.} & G_4(1-G_4)(1-2G_4)\mathbf{x}'_{i.} & G_2G_4[2G_4(1-4G_2)-(1-2G_2)] & G_3G_4[2G_4(1-4G_3)-(1-2G_3)] & G_4(1-G_4)(2G_4-1)^2
\end{bmatrix} \quad (4.1.7).$$

By substituting (4.1.3) through (4.1.9) in (4.1.5) we have constructed an optimal estimating function of the form

$$\mathbf{h}_{\mathbf{Opt}}(\mathbf{y}\ ,\mathbf{x},\theta) = \sum_{i=1}^{N}\left[\mathrm{E}\left(\frac{\partial\ \mathbf{h}(\mathbf{y}_{i.}\ ,\mathbf{x}_{i.},\theta)}{\partial\theta}\right)\left(\mathrm{cov}\left(\mathbf{h}(\mathbf{y}_{i.}\ ,\mathbf{x}_{i.},\theta)\right)\right)^{-1}\mathbf{h}(\mathbf{y}_{i.}\ ,\mathbf{x}_{i.},\theta)\right] =$$

$$= \sum_{i=1}^{n}\left[\begin{array}{l} \begin{bmatrix} -\mathbf{x}_{i.}\mathbf{x}_{i.}'G_2(1-G_2) & \mathbf{x}_{i.}\mathbf{x}_{i.}'G_3G_2 & \mathbf{x}_{i.}\mathbf{x}_{i.}'G_4G_2 & \mathbf{x}_{i.}'G_2(1-G_2)\left[2G_2-1\right] & -\mathbf{x}_{i.}'G_3G_2\left[2G_3-1\right] & -\mathbf{x}_{i.}'G_2G_4\left[2G_4-1\right] \\ \mathbf{x}_{i.}\mathbf{x}_{i.}'G_2G_3 & -\mathbf{x}_{i.}\mathbf{x}_{i.}'G_3(1-G_3) & \mathbf{x}_{i.}\mathbf{x}_{i.}'G_4G_3 & -\mathbf{x}_{i.}'G_2G_3\left[2G_2-1\right] & \mathbf{x}_{i.}'G_3(1-G_3)\left[2G_3-1\right] & -\mathbf{x}_{i.}'G_4G_3\left[2G_4-1\right] \\ \mathbf{x}_{i.}\mathbf{x}_{i.}'G_4G_2 & \mathbf{x}_{i.}\mathbf{x}_{i.}'G_4G_3 & -\mathbf{x}_{i.}\mathbf{x}_{i.}'G_4(1-G_4) & -\mathbf{x}_{i.}'G_4G_2\left[2G_2-1\right] & -\mathbf{x}_{i.}'G_4G_3\left[2G_3-1\right] & \mathbf{x}_{i.}'G_4(1-G_4)\left[2G_4-1\right] \end{bmatrix}\mathbf{\Phi}_i^{-1} \\ \times\begin{bmatrix} \begin{pmatrix} y_{i2}-G_2(\mathbf{x}_{i.},\beta) \\ y_{i3}-G_3(\mathbf{x}_{i.},\beta) \\ y_{i4}-G_4(\mathbf{x}_{i.},\beta) \\ . \\ . \\ y_{iJ}-G_J(\mathbf{x}_{i.},\beta) \end{pmatrix}\odot\mathbf{x}_{i.}' \\ y_{i2}^2-\ 2y_{i2}G_2(\mathbf{x}_{i.},\beta)+2G_2^2(\mathbf{x}_{i.},\beta)-G_2(\mathbf{x}_{i.},\beta) \\ y_{i3}^2-\ 2y_{i3}G_3(\mathbf{x}_{i.},\beta)+2G_3^2(\mathbf{x}_{i.},\beta)-G_3(\mathbf{x}_{i.},\beta) \\ y_{i4}^2-\ 2y_{i4}G_4(\mathbf{x}_{i.},\beta)+2G_4^2(\mathbf{x}_{i.},\beta)-G_4(\mathbf{x}_{i.},\beta) \\ . \\ . \\ y_{iJ}^2-\ 2y_{iJ}G_J(\mathbf{x}_{i.},\beta)+2G_J^2(\mathbf{x}_{i.},\beta)-G_J(\mathbf{x}_{i.},\beta) \end{bmatrix}\end{array}\right] = \mathbf{0} \quad (4.1.9)$$

that can be used in the constrained optimization problem of (3.2.2).

If the $\varepsilon_{i.}$*'s* are iid, each with extreme value distribution, then the special case of the multinomial logit model is defined as in (4.1.4) and the log-likelihood function of the multinomial logit model can be specified as

$$\ln(L(\beta;\mathbf{y})) = \sum_{i=1}^{n}\left[\sum_{j=2}^{J} y_{ij}[\mathbf{x}_{i.}\beta_j] - \ln(1+\sum_{k=2}^{J} e^{\mathbf{x}_{i.}\beta_j})\right] \quad (4.1.10)$$

where $G_j(\mathbf{x}_{i.}\beta) = \dfrac{e^{\mathbf{x}_{i.}\beta_j}}{1+\sum_{k=2}^{J} e^{\mathbf{x}_{i.}\beta_k}}$ and $\beta_1 \equiv [\mathbf{0}]$.

Solving the first-order conditions of (4.1.10) obtains

$$\frac{\partial \ln(L(\beta;\mathbf{y}))}{\partial \beta_j} = \sum_{i=1}^{n} \mathbf{x}'_{i.}(y_{ij} - G_j(\mathbf{x}_{i.}\beta)) = \mathbf{0}\text{, for } j = 2,\ldots..J \quad . \quad (4.1.11)$$

Note that the $\mathbf{h}_{\mathbf{Opt}}(\mathbf{Y}, \mathbf{x}, \theta)$ in (4.1.5) can be specified to represent the first-order conditions (4.1.11) and therefore

$$\mathbf{h}_{\mathbf{Opt}}(\mathbf{Y}, \mathbf{x}, \theta) = \sum_{i=1}^{n}\begin{bmatrix} y_{i2} - G_2(x_{i.},\beta) \\ y_{i3} - G_3(x_{i.},\beta) \\ . \\ . \\ y_{iJ} - G_J(x_{i.},\beta) \end{bmatrix} \odot \mathbf{x}'_{i.} \quad (4.1.12).$$

The solution for β obtained from (4.1.11) or (4.1.12) is the optimal estimating function (OptEF) estimator for β. The estimating function given by (4.1.12) is asymptotically optimal in the sense that it solves the problem of seeking the unbiased estimating function that produces the consistent estimating equation (EE) estimator of β with the smallest asymptotic covariance matrix. Furthermore, the ML estimator has the finite sample optimality property of representing the estimating function (4.1.12) with the smallest standardized covariance matrix. We emphasize that these optimality results are predicted on the assumption that the logistic-extreme value distribution assumption underlying the likelihood specification is in fact the correct parametric family of distributions underlying the data sampling process. It is also useful to note that (4.1.3) subsumes (4.1.12) and the asymptotic covariance matrix of the MEL estimator generally becomes smaller (by a positive definite matrix) as the number of estimating equations on which it is based increases (Qin and Lawless,1994, Corollary 1).

4.1 Adding Flexibility to the EL Formulation

In this section we introduce flexibility into the specification of $G(\bullet)$ by adding parameters to index members of the class of Pearson Family distributions. While we focus on the Pearson class here, we emphasize that any other class of distributions could be used. The criteria for identifying different members of the system of Pearson distributions can be expressed parametrically in terms of a (2×1) vector, ξ ,of parameters, so that $\mathrm{G}_j(\mathbf{x}_{i.},\theta)=\mathrm{G}_j(\mathbf{x}_{i.},\beta,\xi)$ where $\theta\equiv \mathrm{vec}([\beta_1,\beta_2,.......\beta_J,\psi_3,\omega])$ is a column-vectorized representation of model parameters now of dimension $((KJ+2)\times 1)$. Hence, the alternative formulation of unbiased estimating functions for θ is of the form

$$\mathbf{h}_{\boldsymbol{\tau}}(\mathbf{Y},\mathbf{x},\theta)\equiv\sum_{i=1}^{n}\tau_i(\mathbf{x}_{i.},\theta)\mathbf{h}(\mathbf{y}_{i.},\mathbf{x}_{i.},\theta) \tag{4.2.1}$$

where

$$\mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) = \begin{bmatrix} \begin{pmatrix} y_{i2} - G_2(\mathbf{x}_{i.}, \beta, \xi) \\ y_{i3} - G_3(\mathbf{x}_{i.}, \beta, \xi) \\ y_{i4} - G_4(\mathbf{x}_{i.}, \beta, \xi) \\ . \\ . \\ y_{iJ} - G_J(\mathbf{x}_{i.}, \beta, \xi) \end{pmatrix} \odot \mathbf{x}_{i.}' \\ y_{i2}^2 - 2y_{i2}G_2(\mathbf{x}_{i.}, \beta, \xi) + 2G_2^2(\mathbf{x}_{i.}, \beta, \xi) - G_2(\mathbf{x}_{i.}, \beta, \xi) \\ y_{i3}^2 - 2y_{i3}G_3(\mathbf{x}_{i.}, \beta, \xi) + 2G_3^2(\mathbf{x}_{i.}, \beta, \xi) - G_3(\mathbf{x}_{i.}, \beta, \xi) \\ y_{i4}^2 - 2y_{i4}G_4(\mathbf{x}_{i.}, \beta, \xi) + 2G_4^2(\mathbf{x}_{i.}, \beta, \xi) - G_4(\mathbf{x}_{i.}, \beta, \xi) \\ . \\ . \\ y_{iJ}^2 - 2y_{iJ}G_J(\mathbf{x}_{i.}, \beta, \xi) + 2G_J^2(\mathbf{x}_{i.}, \beta, \xi) - G_J(\mathbf{x}_{i.}, \beta, \xi) \end{bmatrix} \qquad (4.2.2)$$

and $G_j(\mathbf{x}_{i.}, \beta, \xi)$ again denotes the conditional expectation of $\mathbf{Y}_{ij}$ given $\mathbf{x}_{i.}$ In the context of multinomial choice problem, $G_j(\mathbf{x}_{i.}, \beta, \xi)$ denotes the conditional-on- $\mathbf{x}_{i.}$ choice probability. The OptEF estimator is in the general class of estimating equations based on the estimating functions of the form (4.2.1) characterized by the solution to

$$\mathbf{h}_{\mathbf{Opt}}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) = \sum_{i=1}^{N} \left[E\left(\frac{\partial\, \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta)}{\partial\theta} \right) \left(\mathrm{cov}\left(\mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) \right) \right)^{-1} \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta) \right] = \mathbf{0} \qquad (4.2.3)$$

where $E\left(\frac{\partial\, \mathbf{h}(y_{ij}, \mathbf{x}_{i.}, \theta)}{\partial\theta} \right)$ is now a matrix of dimension ((K(J-1)+2)$\times$(K(J-1)+(J-1))),

where , for example in the case of J = 4 ,

$$E\left(\frac{\partial\, \mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta)}{\partial\theta}\right) =$$

$$\begin{bmatrix} \frac{-\partial[G_2\mathbf{x}'_{i.}]}{\partial\beta_2} & \frac{-\partial[G_3\mathbf{x}'_{i.}]}{\partial\beta_2} & \frac{-\partial[G_4\mathbf{x}'_{i.}]}{\partial\beta_2} & \frac{\partial G_2}{\partial\beta_2}[2G_2-1] & \frac{\partial G_3}{\partial\beta_2}[2G_3-1] & \frac{\partial G_4}{\partial\beta_2}[2G_4-1] \\ \frac{-\partial[G_2\mathbf{x}'_{i.}]}{\partial\beta_3} & \frac{-\partial[G_3\mathbf{x}'_{i.}]}{\partial\beta_3} & \frac{-\partial[G_4\mathbf{x}'_{i.}]}{\partial\beta_3} & \frac{\partial G_2}{\partial\beta_3}[2G_2-1] & \frac{\partial G_3}{\partial\beta_3}[2G_3-1] & \frac{\partial G_4}{\partial\beta_3}[2G_4-1] \\ \frac{-\partial[G_2\mathbf{x}'_{i.}]}{\partial\beta_4} & \frac{-\partial[G_3\mathbf{x}'_{i.}]}{\partial\beta_4} & \frac{-\partial[G_4\mathbf{x}'_{i.}]}{\partial\beta_4} & \frac{\partial G_2}{\partial\beta_4}[2G_2-1] & \frac{\partial G_3}{\partial\beta_4}[2G_3-1] & \frac{\partial G_4}{\partial\beta_4}[2G_4-1] \\ \frac{-\partial[G_2\mathbf{x}'_{i.}]}{\partial\psi_3} & \frac{-\partial[G_3\mathbf{x}'_{i.}]}{\partial\psi_3} & \frac{-\partial[G_4\mathbf{x}'_{i.}]}{\partial\psi_3} & \frac{\partial G_2}{\partial\psi_3}[2G_2-1] & \frac{\partial G_3}{\partial\psi_3}[2G_3-1] & \frac{\partial G_4}{\partial\psi_3}[2G_4-1] \\ \frac{-\partial[G_2\mathbf{x}'_{i.}]}{\partial\omega} & \frac{-\partial[G_3\mathbf{x}'_{i.}]}{\partial\omega} & \frac{-\partial[G_4\mathbf{x}'_{i.}]}{\partial\omega} & \frac{\partial G_2}{\partial\omega}[2G_2-1] & \frac{\partial G_3}{\partial\omega}[2G_3-1] & \frac{\partial G_4}{\partial\omega}[2G_4-1] \end{bmatrix} \qquad (4.2.4)$$

and $\Phi_i = \operatorname{cov}\left(\mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta)\right) = E_\theta\left[\mathbf{h}(\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta)\, \mathbf{h}((\mathbf{y}_{i.}, \mathbf{x}_{i.}, \theta))'\right]$ is the

((K(J-1)+(J-1) $\times$ (K(J-1)+(J-1)) covariance matrix of ($Y_{ij}|\mathbf{x}_{i.}$).

In order to define the OptEF, note that the covariance matrix of any [Y_{i1},.. Y_{iJ}, $(Y_{i1}-G_1)^2$, … $(Y_{iJ}-G_J)^2$] vector in this multinomial choice problem is given by (again for an illustrative case where J = 4)

$$\Phi_i = \qquad (4.2.5)$$

$$\begin{bmatrix} G_2(1-G_2)\mathbf{x}_{i.}\mathbf{x}'_{i.} & -2G_2G_3\mathbf{x}_{i.}\mathbf{x}'_{i.} & -2G_2G_4\mathbf{x}_{i.}\mathbf{x}'_{i.} & G_2(1-G_2)(1-2G_2)\mathbf{x}'_{i.} & -G_2G_3(1\text{-}4G_3)\mathbf{x}'_{i.} & -G_2G_4(1\text{-}4G_4)\mathbf{x}'_{i.} \\ -2G_2G_3\mathbf{x}_{i.}\mathbf{x}'_{i.} & G_3(1-G_3)\mathbf{x}_{i.}\mathbf{x}'_{i.} & -2G_3G_4\mathbf{x}_{i.}\mathbf{x}'_{i.} & -G_2G_3(1\text{-}4G_2)\mathbf{x}'_{i.} & G_3(1-G_3)(1-2G_3)\mathbf{x}'_{i.} & -G_3G_4(1\text{-}4G_4)\mathbf{x}'_{i.} \\ -2G_2G_4\mathbf{x}_{i.}\mathbf{x}'_{i.} & -2G_3G_4\mathbf{x}_{i.}\mathbf{x}'_{i.} & G_4(1-G_4)\mathbf{x}_{i.}\mathbf{x}'_{i.} & -G_2G_4(1\text{-}4G_2)\mathbf{x}'_{i.} & -G_3G_4(1\text{-}4G_3)\mathbf{x}'_{i.} & G_4(1-G_4)(1-2G_4)\mathbf{x}'_{i.} \\ G_2(1-G_2)(1-2G_2)\mathbf{x}'_{i.} & -G_2G_3(1\text{-}4G_2)\mathbf{x}'_{i.} & -G_2G_4(1\text{-}4G_2)\mathbf{x}'_{i.} & G_2(1-G_2)(2G_2-1)^2 & G_2G_3[2G_2(1-4G_3)-(1-2G_3)] & G_2G_4[2G_4(1-4G_2)-(1-2G_2)] \\ -G_2G_3(1\text{-}4G_3)\mathbf{x}'_{i.} & G_3(1-G_3)(1-2G_3)\mathbf{x}'_{i.} & -G_2G_4(1\text{-}4G_2)\mathbf{x}'_{i.} & G_2G_3[2G_2(1-4G_3)-(1-2G_3)] & G_3(1-G_3)(2G_3-1)^2 & G_3G_4[2G_4(1-4G_3)-(1-2G_3)] \\ -G_2G_4(1\text{-}4G_4)\mathbf{x}'_{i.} & -G_3G_4(1\text{-}4G_4)\mathbf{x}'_{i.} & G_4(1-G_4)(1-2G_4)\mathbf{x}'_{i.} & G_2G_4[2G_4(1-4G_2)-(1-2G_2)] & G_3G_4[2G_4(1-4G_3)-(1-2G_3)] & G_4(1-G_4)(2G_4-1)^2 \end{bmatrix}$$

By substituting (4.2.4) and (4.2.5) in (4.2.3) we have constructed an optimal estimating function that can be used in the constrained optimization problem of (3.2.2).

5. Sampling Experiments

We performed a Monte Carlo experiment to estimate a *Multinomial Logit* (ML*)* response model where β_1 has been normalized, without loss of generality, to a zero vector for purposes of parameter identification and there existed four choice alternatives. The **x** data were all generated iid from the uniform distribution having support on the interval $(-5,5)$. The logistic distribution was used to generate the choice probabilities underlying the data sampling process. The link function used to model the multinomial choice problem was Pearson X. The parameters of the latent variable equations underlying the Multinomial Logit model are given by

$$\beta_2 = \begin{bmatrix}\beta_{21}\\ \beta_{22}\\ \beta_{23}\end{bmatrix} = \begin{bmatrix}0.1\\ 0.2\\ 0.3\end{bmatrix}, \beta_3 = \begin{bmatrix}\beta_{31}\\ \beta_{32}\\ \beta_{33}\end{bmatrix} = \begin{bmatrix}0.4\\ 0.5\\ 0.6\end{bmatrix}, \text{ and } \beta_4 = \begin{bmatrix}\beta_{41}\\ \beta_{42}\\ \beta_{43}\end{bmatrix} = \begin{bmatrix}0.7\\ 0.8\\ 0.9\end{bmatrix}.$$

The results of the Monte Carlo experiment, for 200 repetitions of the sampling experiment, are displayed in Table 1. The results suggest that the EL estimation procedure produces reasonably accurate estimates of the model parameters. As the sample size increases, the mean square error decreases, indicative of the consistency of the EL estimator. The means of the estimates for the 200 Monte Carlo replications are very close to the true values of the model parameters for sample sizes ≥ 500, suggesting that for all practical purposes, the EL estimators are producing near-unbiased estimates of the parameters. For smaller sample sizes, there is some indication that the parameter estimates are biased to some degree, although the degree of bias is relatively small. Overall, the estimates were quite accurate across all sample sizes, and accurate for large sample sizes, and would appear to be useful from an empirical application application perspective.

Table 1. Monte Carlo Results: Multinomial Choice Model with Four Alternatives[1]

		Sample Sizes							
Parameter	**TrueValue**	**50**	**100**	**200**	**250**	**300**	**500**	**600**	**700**
β_{21}	0.1	0.115	0.134	0.121	0.127	0.096	0.094	0.106	0.093
β_{22}	0.2	0.212	0.216	0.209	0.209	0.208	0.198	0.199	0.197
β_{23}	0.3	0.307	0.306	0.319	0.311	0.298	0.296	0.301	0.301
β_{31}	0.4	0.436	0.439	0.457	0.425	0.412	0.410	0.399	0.421
β_{32}	0.5	0.523	0.533	0.523	0.530	0.521	0.509	0.506	0.504
β_{33}	0.6	0.629	0.618	0.634	0.628	0.627	0.610	0.611	0.608
β_{41}	0.7	0.739	0.720	0.751	0.747	0.698	0.708	0.707	0.719
β_{42}	0.8	0.856	0.851	0.852	0.834	0.832	0.814	0.815	0.804
β_{43}	0.9	0.956	0.945	0.963	0.941	0.935	0.917	0.908	0.910
[1]MSE(β)		**0.531**	**0.519**	**0.459**	**0.292**	**0.283**	**0.149**	**0.130**	**0.114**

1) Values below the sample size indicators are the sample means of the estimates for 200 MC repetitions of the experiment.

We also note that the computation of the estimates for this 4-dimensional choice model was relatively quick with effectively no numerical difficulties when finding solutions. We also note that the discrepancy in some of the parameters may be due to the fact that we have used numerical gradients instead of analytical gradients in solving the EL optimization problem. Analytical gradients could serve to speed convergence further, and would also allow solutions to higher levels of tolerance, potentially further increasing the accuracy of the parameter estimates.

6. Concluding Remarks

This paper has presented a flexible semiparametric methodology for estimating multinomial choice models. The parameter estimates from the Monte Carlo results appear quite reasonable and demonstrate the potential usefulness of the proposed approach. The estimates obtained by this procedure are consistent and asymptotically normal. However, our consistent estimator will generally not be fully efficient. Nonetheless, because of the computational difficulties associated with more efficient estimators, the empirical tractability of the method for estimating a system of multinomial choice models for large data sets and for relatively large dimensional choice sets is very attractive in empirical practice. Moreover, in practice, there is often insufficient information to specify the parametric form of the function linking the observable data to the unknown choice probabilities, in which case a fully efficient method of estimating the model parameters will generally remain unknown in any case. In such cases, the flexible Pearson family of parametric distributions may be useful as a basis for a flexible specification of a link function.

7. References

Ahn, H., Ichimura, H. and Powell, J. L., (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism, *Journal of Econometrics* **58**: 3-29.

Armstrong, B.G., (1985). The generalized linear model, *Communications in Statistics* **14**(B): 529-544.

Baggerly, K.A. Empirical Likelihood as a Goodness of Fit Measure. Biometrika, 85: 535-547, 1998.

Blundell, R. and Powell, J.L., (1999). Endogeneity in single index models, *Working Paper, Department of Economics, UC Berkeley*.

Caroll, R.J., D. Ruppert, and L.A. Stefansky, (1995), Measurement Error in Nonlinear Models, London; Chapman and Hall.

Cressie, N. and Read, T. Multinomial Goodness of Fit Tests. Journal of Royal Statistical Society of Series B 46:440-464, 1984.

Elderton, W.P., Frequency Curves and Correlation, Haren Press, 1953.

Elderton, W.P. and Johnson, N.L. Systems of Frequency Curves, Cambridge University Press, 1969.

Golan, A., Judge, G.G. and Perloff, J., (1996). A Maximum Entropy Approach to Recovering Information from Multinomial Response Data, *Journal of the American Statistical Association* **91**: 841-853.

Ichimura, H., (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics* **58**: 71-120.

Klein, R., Spady, R.H., (1993). An efficient semiparametric estimator for binary response models, *Econometrica* **61**: 387-421.

Kullback, S. Information Theory and Statistics. New York: John Wiley and Sons, 1959.

Kullback, S. and Leibler, R.A. On information and sufficiency, The Annals of Mathematical Statistics 22: 79-86, 1951.

Lewbel, A., (1997). Semiparametric estimation of location and other discrete choice moments, *Econometric Theory* **13**: 32-51.

Lewbel, A., (1998). Semiparametric latent variable model estimation with endogenous or mismeasured regressors, *Econometrica* **66**: 105-121.

Lewbel, A., (2000). Semiparametric qualitative response model equation with unknown heteroscedasticity or instrumental variables, *Journal of Econometrics* **97**: 145-177.

Maddala, G.S., (1983). Limited Dependent and Qualitative Variables in Econometrics, In: Econometric Society Monograph No. 3. Cambridge University Press, Cambridge.

McFadden, D.L., (1984). Econometric analysis of qualitative response models, In: Griliches, Z., Intriligator, M.D. (Eds), Handbook of Econometrics, Vol.2 Elsevier, Amsterdam, pp.1395-1457.

Mittelhammer, R.C., Judge, G.G. and Miller, D.J., Econometric Foundations, Cambridge University Press, 2000.

Newey, W.K., (1985). Semiparametric estimation of limited dependent variable models with endogenous explanatory variables. Annales de l'INSEE 59-60, 219-237.

Owen, A. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75: 237-249.

Owen, A. Empirical likelihood for linear models. The Annals of Statistics 19 (4): 1725-1747, 1991.

Owen, A. Empirical likelihood. New York: Chapman and Hall, 2000.

Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variations in homogeneous material, *Philosophical Transactions of the Royal Society of London, Series A*, **186**, 343-414.

Pearson, K. (1924). On the mean error of frequency distributions, *Biometrika*, **16**, 198-200.

Pearson, E.S. and Hartley, H.O., Biometricka Tables for Statisticians, Cambridge University Press, Vol. 1, 1956.

Powell, J.L., Stock, J.H., Stoker, T.M., (1989). Semiparametric estimation of index coefficients, *Econometrica* **57**: 1403-1430.

Qin, J. and Lawless, J. Empirical likelihood and general estimating equations. The Annals of Statistics 22 (1): 300-325, 1994.

Ruud, P.A. (1986). Consistent estimation of limited dependent variable models despite misspecification of distribution, *Journal of Econometrics* **32**: 157-187.

Appendix: Derivation of CDFs of the Pearson Family of Distributions

Pearson type I

$$f(x) = \frac{(x-r_1)^{m_1}(r_2-x)^{m_2}}{\beta(m_1+1, m_2+1)(r_2-r_1)^{m_1+m_2+1}}$$

$$k\int_{\mathbf{r}1}^{\mathbf{z}}(\mathbf{x}-\mathbf{r}1)^{\mathbf{m}1}(\mathbf{r}2-\mathbf{x})^{\mathbf{m}2}\mathbf{dx} = 1$$

let u = x-r1 then dx = du then $\int_0^{\mathbf{r}2-\mathbf{r}1}(\mathrm{u})^{\mathbf{m}1}(1-\frac{\mathrm{u}}{\mathrm{r2-r1}})^{\mathbf{m}2}(\mathbf{r}2-\mathbf{r}1)^{\mathbf{m}2}\mathbf{du}$

let $\mathbf{w} = \frac{\mathbf{u}}{\mathbf{r}2-\mathbf{r}1} \Rightarrow \mathbf{dw}(\mathbf{r}2-\mathbf{r}1) = \mathbf{du} = \mathbf{k}\int_0^1(\mathbf{r}2-\mathbf{r}1)^{\mathbf{m}1}\mathbf{w}^{\mathbf{m}1}(1-\mathbf{w})^{\mathbf{m}2}(\mathbf{r}2-\mathbf{r}1)^{\mathbf{m}2+1}\mathbf{dw} = 1$

which implies

$$k = \frac{1}{(r2-r1)^{m1+m2+1}Beta(m1+1, m2+1)}.$$

F(z) = $\mathbf{p}(\mathbf{x} \le \mathbf{z}) = \mathbf{k}\int_{\mathbf{r}1}^{\mathbf{z}}(\mathbf{x}-\mathbf{r}1)^{\mathbf{m}1}(\mathbf{r}2-\mathbf{x})^{\mathbf{m}2}\mathbf{dx}$ let u=x then

$$\mathbf{p}(\mathbf{x} \le \mathbf{z}) = \mathbf{k}\int_0^{\mathbf{z}-\mathbf{r}1}(\mathbf{u})^{\mathbf{m}1}(\mathbf{r}2-\mathbf{r}1-\mathbf{u})^{\mathbf{m}2}\mathbf{du} = \mathbf{k}(\mathbf{r}2-\mathbf{r}1)^{\mathbf{m}2}\int_0^{\mathbf{z}-\mathbf{r}1}\mathbf{u}^{\mathbf{m}1}(1-\frac{\mathbf{u}}{\mathbf{r}2-\mathbf{r}1})^{\mathbf{m}2}\mathbf{du}$$

let $\mathbf{w} = \frac{\mathbf{u}}{\mathbf{r}2-\mathbf{r}1} \Rightarrow (\mathbf{r}2-\mathbf{r}1)\mathbf{dw} = \mathbf{du}$ then

$$\mathbf{p}(\mathbf{x} \le \mathbf{z}) = \mathbf{k}(\mathbf{r}2-\mathbf{r}1)^{\mathbf{m}2+\mathbf{m}1+1}\int_0^{\frac{\mathbf{z}-\mathbf{r}1}{\mathbf{r}2-\mathbf{r}1}}(\mathbf{w})^{\mathbf{m}1}(1-\mathbf{w})^{\mathbf{m}2}\mathbf{dw} = \frac{1}{\mathbf{beta}(\mathbf{m}1+1, \mathbf{m}2+1)}\int_0^{\frac{\mathbf{z}-\mathbf{r}1}{\mathbf{r}2-\mathbf{r}1}}(\mathbf{w})^{\mathbf{m}1}(1-\mathbf{w})^{\mathbf{m}2}\mathbf{dw}$$

= incomplete beta($\frac{\mathbf{z}-\mathbf{r}1}{\mathbf{r}2-\mathbf{r}1}$,m1+1,m2+1).

Pearson type IV

Is not available.

Pearson type VI

f(x) = $$\frac{(\mathbf{x}-\mathbf{r}_1)^{\mathbf{m}_1}(\mathbf{x}-\mathbf{r}_2)^{\mathbf{m}_2}}{\beta(\mathbf{m}_1+1,-\mathbf{m}_1-\mathbf{m}_2-1)(\mathbf{r}_1-\mathbf{r}_2)^{\mathbf{m}_1+\mathbf{m}_2+1}}$$

$$\mathbf{r}_1 < \mathbf{x} < +\infty$$

$\mathrm{k}\int_{\mathbf{r}1}^{\infty}(\mathbf{x}-\mathbf{r}1)^{\mathbf{m}1}(-\mathbf{r}2+\mathbf{x})^{\mathbf{m}2}\mathbf{dx}=1$ let x-r2=u then $\mathbf{k}\int_{\mathbf{r}1-\mathbf{r}2}^{\infty}(\mathbf{u})^{\mathbf{m}2}(\mathbf{r}2-\mathbf{r}1+\mathbf{u})^{\mathbf{m}1}\mathbf{du}=1$ let

a = r1-r2 which implies $\mathbf{k}\int_{\mathbf{r}1-\mathbf{r}2}^{\infty}(\mathbf{u})^{\mathbf{m}2}(\mathbf{u}-\mathbf{a})^{\mathbf{m}1}\mathbf{du}=1=\int_{\mathbf{a}}^{\infty}\left(\frac{\mathbf{u}}{\mathbf{a}}-1\right)^{\mathbf{m}1}\mathbf{a}^{\mathbf{m}1}\left(\frac{\mathbf{u}}{\mathbf{a}}\right)^{\mathbf{m}2}\mathbf{du}$

let

$$\frac{1}{\mathbf{z}}=\frac{\mathbf{u}}{\mathbf{a}}\Rightarrow -\mathbf{k}\mathbf{a}^{\mathbf{m}1+\mathbf{m}2+1}\int_{1}^{0}\left(\frac{1}{\mathbf{z}}-1\right)^{\mathbf{m}1}\left(\frac{1}{\mathbf{z}}\right)^{\mathbf{m}2}\mathbf{z}^{-2}\mathbf{dz}\Rightarrow\left[\mathbf{k}=\frac{1}{\mathbf{a}^{\mathbf{m}1+\mathbf{m}2+1}\mathbf{beta}(\mathbf{m}1+1,-\mathbf{m}1-\mathbf{m}2-1)}\right].$$

$\mathbf{p}(\mathbf{x}\le\mathbf{z})=\mathbf{k}\int_{\mathbf{r}1}^{\mathbf{z}}(\mathbf{x}-\mathbf{r}1)^{\mathbf{m}1}(-\mathbf{r}2+\mathbf{x})^{\mathbf{m}2}\mathbf{dx}$ let u= x-r2 then

$\mathbf{p}(\mathbf{x}\le\mathbf{z})=\mathbf{k}\int_{\mathbf{r}1-\mathbf{r}2}^{\mathbf{z}-\mathbf{r}2}(\mathbf{u}+\mathbf{r}2-\mathbf{r}1)^{\mathbf{m}1}(\mathbf{u})^{\mathbf{m}2}\mathbf{du}=\mathbf{k}\mathbf{a}^{\mathbf{m}1+\mathbf{m}2}\int_{\mathbf{a}}^{\mathbf{z}-\mathbf{r}2}(\frac{\mathbf{u}}{\mathbf{a}}-1)^{\mathbf{m}1}(\frac{\mathbf{u}}{\mathbf{a}})^{\mathbf{m}2}\mathbf{du}$ let wu = a then

du=aw^{-2}dw then $\mathbf{p}(\mathbf{x}\le\mathbf{z})=\mathbf{k}\mathbf{a}^{\mathbf{m}1+\mathbf{m}2+1}\int_{\frac{\mathbf{a}}{\mathbf{z}-\mathbf{r}2}}^{1}(1-\mathbf{w})^{\mathbf{m}1}(\mathbf{w})^{-\mathbf{m}1-2-\mathbf{m}2}\mathbf{dx}=$ incomplete beta.

Transition Pearson type II

$$f(x)=\frac{(t^2-x^2)^M}{t^{2M+1}\beta(M+1,0.5)}$$

$$-t<x<t$$

find k

$$\mathbf{k}\int_{-\mathbf{s}}^{\mathbf{s}}(\mathbf{s}^2-\mathbf{x}^2)^{\mathbf{m}}\,\mathbf{dx}=1=2\mathbf{k}\int_0^{\mathbf{s}}(\mathbf{s}^2-\mathbf{x}^2)^{\mathbf{m}}\,\mathbf{dx}=2\mathbf{ks}^{2\mathbf{m}}\int_0^{\mathbf{s}}(1-\frac{\mathbf{x}^2}{\mathbf{s}^2})^{\mathbf{m}}\,\mathbf{dx}=1$$

let u = $(1-\frac{\mathbf{x}^2}{\mathbf{s}^2})$ then $2\mathbf{ks}^{2\mathbf{m}}\int_1^0\mathbf{u}^{\mathbf{m}}(\frac{\mathbf{s}^2}{2\mathbf{x}})\ \mathbf{du}=\mathbf{ks}^{2\mathbf{m}}\int_0^1\mathbf{u}^{\mathbf{m}}(\frac{\mathbf{s}^2}{\mathbf{x}})\ \mathbf{du}=\mathbf{ks}^{2\mathbf{m}+1}\int_0^1\mathbf{u}^{\mathbf{m}}(1-\mathbf{u})^{-\frac{1}{2}}\mathbf{du}=1$

$$\mathbf{k}=\frac{1}{\mathbf{s}^{2\mathbf{m}+1}\mathbf{beta}(\mathbf{m}+1,.5)}$$

F(z) = p(x$\leq$z) = $\mathbf{k}\int_{-\mathbf{s}}^{\mathbf{s}}(\mathbf{s}^2-\mathbf{x}^2)^{\mathbf{m}}\,\mathbf{dx}=\mathbf{k}\left[\int_{-\mathbf{s}}^{0}(\mathbf{s}^2-\mathbf{x}^2)^{\mathbf{m}}\,\mathbf{dx}+\int_0^{\mathbf{z}}(\mathbf{s}^2-\mathbf{x}^2)^{\mathbf{m}}\,\mathbf{dx}\right]$

$1-\frac{\mathbf{x}^2}{\mathbf{s}^2}=\mathbf{u}$ then

F(z) = p(x$\leq$z) =

$$-\frac{1}{2}\mathbf{s}^{2\mathbf{m}}\int_0^1(\mathbf{u})^{\mathbf{m}}\left(\frac{\mathbf{s}}{\mathbf{x}}\right)\mathbf{sdu}-\frac{1}{2}\mathbf{s}^{2\mathbf{m}+1}\int_1^{1-\frac{\mathbf{z}^2}{\mathbf{s}^2}}(\mathbf{u})^{\mathbf{m}}(1-\mathbf{u})^{\frac{-1}{2}}\mathbf{sdu}=\frac{1}{2}+\frac{1}{2}\frac{1}{\mathbf{beta}(\mathbf{m}+1,.5)}\int_{1-\frac{\mathbf{z}^2}{\mathbf{s}^2}}^{1}(\mathbf{u})^{\mathbf{m}}(1-\mathbf{u})^{\frac{-1}{2}}\mathbf{du}=$$

.5 incomplete beta($1-\frac{\mathbf{z}^2}{\mathbf{s}^2}$,m+1,.5) for z > 0 and for z<0

F(z) = p(x$\leq$z)=1-.5 incomplete beta($1-\frac{\mathbf{z}^2}{\mathbf{s}^2}$,m+1,.5)

Pearson type III

$$f(x)=$$
$$\frac{e^{-N^2}N^{N^2}e^{-Nx}(N+x)^{N^2-1}}{\Gamma(N^2)}$$ given k= $\frac{\mathbf{e}^{-\mathbf{n}^2}\mathbf{n}^{\mathbf{n}^2}}{\mathbf{gamma}(\mathbf{n}^2)}$
$$-N<x<N$$

Then F(z) = p(x ≤ z) = $\mathbf{k}\int_{-\mathbf{n}}^{\mathbf{z}}\mathbf{e}^{-\mathbf{nx}}(\mathbf{n}+\mathbf{x})^{\mathbf{n}^2-1}\mathbf{dx}$

Let u = n(n+x) which implies F(z) = p(x ≤ z) =

$$\mathbf{k}\int_{0}^{\mathbf{n}(\mathbf{n}+\mathbf{z})}\mathbf{e}^{-\mathbf{u}+\mathbf{n}^2}(\mathbf{u})^{\mathbf{n}^2-1}\mathbf{n}^{1-\mathbf{n}^2}\mathbf{n}^{-1}\mathbf{du}=\frac{1}{\mathbf{gamma}(\mathbf{n}^2)}\int_{0}^{\mathbf{n}(\mathbf{n}+\mathbf{z})}\mathbf{e}^{-\mathbf{u}}(\mathbf{u})^{\mathbf{n}^2-1}\mathbf{du} =$$

= incomplete gamma(n(n+z),n^2).

Pearson type V

Given k = $\frac{(2\mathbf{r}(\mathbf{m}-1))^{2\mathbf{m}-1}}{\mathbf{gamma}(2\mathbf{m}-1)}$ then F(z) = p(x ≤ z) = $\mathbf{k}\int_{-\mathbf{n}}^{\mathbf{z}}\mathbf{e}^{-\frac{2\mathbf{r}(\mathbf{m}-1)}{\mathbf{x}+\mathbf{r}}}(\mathbf{r}+\mathbf{x})^{-2\mathbf{m}}\mathbf{dx}$

X+ r = $\frac{1}{\mathbf{u}}$ which implies F(z) = p(x ≤ z) =

$$\mathbf{k}\int_{\frac{1}{\mathbf{z}+\mathbf{r}}}^{\infty}\mathbf{e}^{-2\mathbf{r}(\mathbf{m}-1)}(\mathbf{u})^{2\mathbf{m}-2}\mathbf{du}=\frac{(2\mathbf{r}(\mathbf{m}-1))^{2\mathbf{m}-1}}{\mathbf{gamma}(2\mathbf{m}-1)}\int_{\frac{1}{\mathbf{z}+\mathbf{r}}}^{\infty}\mathbf{e}^{-2\mathbf{r}(\mathbf{m}-1)}(\mathbf{u})^{2\mathbf{m}-2}\mathbf{du} = \text{incomplete gamma.}$$

Pearson type VII

$f(x)=\frac{s_1^{\,2m-1}(x^2+s_1^{\,2})^{-m}}{\beta(m-0.5,0.5)}$ given k= $\frac{\mathbf{s}^{2\mathbf{m}-1}}{\mathbf{beta}(\mathbf{m}-.5,.5)}$ then F(z) = p(x ≤ z) =

$-\infty < x < +\infty$

$$\mathbf{k}\int_{-\infty}^{\mathbf{z}}(\mathbf{s}^2+\mathbf{x}^2)^{-\mathbf{m}}\mathbf{dx}=\mathbf{k}\left[\int_{-\infty}^{0}(\mathbf{s}^2+\mathbf{x}^2)^{-\mathbf{m}}\mathbf{dx}+\int_{0}^{\mathbf{z}}(\mathbf{s}^2+\mathbf{x}^2)^{-\mathbf{m}}\mathbf{dx}\right]$$

let $\frac{1}{\mathbf{u}}=1+\frac{\mathbf{x}^2}{\mathbf{s}^2}$ which implies then F(z) = p(x ≤ z) =

$$\frac{\mathbf{s}^{1-2\mathbf{m}}}{2}\int_0^1 \mathbf{u}^{\mathbf{m}-\frac{3}{2}}(1-\mathbf{u})^{-\frac{1}{2}}\mathbf{du}+\frac{\mathbf{s}^{1-2\mathbf{m}}}{2}\int_{\frac{\mathbf{s}^2}{\mathbf{s}^2+\mathbf{z}^2}}^1 \mathbf{u}^{\mathbf{m}-\frac{3}{2}}(1-\mathbf{u})^{-\frac{1}{2}}\mathbf{du}$$

for z > 0

F(z) = p(x $\leq$ z) = 1-.5 incomplete beta $\left(\frac{\mathbf{s}^2}{\mathbf{s}^2+\mathbf{z}^2},\mathbf{m}-.5,.5\right)$ =

$$=.5+\frac{1}{\mathbf{beta}(\mathbf{m}-.5,.5)}\int_{\frac{\mathbf{s}^2}{\mathbf{s}^2+\mathbf{z}^2}}^1 \mathbf{u}^{\mathbf{m}-\frac{3}{2}}(1-\mathbf{u})^{-\frac{1}{2}}\mathbf{d}$$

For z <0

F(z) = p(x $\leq$ z) = $\frac{\mathbf{s}^{1-2\mathbf{m}}}{2}\int_{\frac{\mathbf{s}^2}{\mathbf{s}^2+\mathbf{z}^2}}^1 \mathbf{u}^{\mathbf{m}-\frac{3}{2}}(1-\mathbf{u})^{-\frac{1}{2}}\mathbf{d}$ u.

Pearson type VIII

$f(x)=$

$$\frac{(1-2m)(x-r_1)^{-2m}}{(r_2-r_1)^{1-2m}}$$ where k = $\frac{1-2\mathbf{m}}{(\mathbf{r}2-\mathbf{r}1)^{1-2\mathbf{m}}}$ then

$r_1 < x < r_2$

$1-2m>0$

F(z) = p(x $\leq$ z) = $\mathbf{k}\int_{\mathbf{r}1}^{\mathbf{z}}(\mathbf{x}-\mathbf{r}1)^{-2\mathbf{m}}\mathbf{dx}=\frac{(\mathbf{z}-\mathbf{r}1)^{1-2\mathbf{m}}}{(\mathbf{r}2-\mathbf{r}1)^{1-2\mathbf{m}}}$ u.

Pearson type XI

$$\text{f(x)}=\frac{(\frac{-x+r_2}{x-r_1})^{m_2}}{\beta(m_2+1,-m_2+1)(-r_1+r_2)}$$

$r_1 < x < r_2$

given k = $\frac{1}{(\mathbf{r}2-\mathbf{r}1)\mathbf{beta}(\mathbf{m}2+1,-\mathbf{m}2+1)}$

then F(z) = p(x$\leq$z) = $\mathbf{k}\int_{\mathbf{r}1}^{\mathbf{z}}(\mathbf{x}-\mathbf{r}1)^{-\mathbf{m}2}(\mathbf{r}2-\mathbf{x})^{\mathbf{m}2}\mathbf{dx}$.

let r2-x=u then F(z) = p(x$\leq$z) = $\mathbf{k}\int_{\mathbf{r}2-\mathbf{z}}^{\mathbf{r}2-\mathbf{r}1}(\mathbf{u})^{\mathbf{m}2}(\mathbf{r}2-\mathbf{r}1-\mathbf{u})^{-\mathbf{m}2}\mathbf{du}$

let z = $\frac{\mathbf{u}}{\mathbf{r}2-\mathbf{r}1}$ which implies F(z) = p(x$\leq$z) = $\frac{1}{\mathbf{beta}(\mathbf{m}2+1,1-\mathbf{m}2)}\int_{\frac{\mathbf{r}2-\mathbf{z}}{\mathbf{r}2-\mathbf{r}1}}^{1}(\mathbf{z})^{\mathbf{m}2}(1-\mathbf{z})^{-\mathbf{m}2}\mathbf{dz}$

incomplete beta