DISCUSSION PAPER SERIES

IZA DP No. 222

Betit: A Family That Nests Probit and Logit

Wim P.M. Vijverberg

December 2000

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Betit:
A Family That Nests Probit and Logit

Wim P.M. Vijverberg
University of Texas at Dallas and IZA, Bonn

Discussion Paper No. 222
December 2000

IZA

P.O. Box 7240
D-53072 Bonn
Germany

Tel.: +49-228-3894-0
Fax: +49-228-3894-210
Email: iza@iza.org

IZA Discussion Paper No. 222
December 2000

ABSTRACT

Betit: A Family That Nests Probit and Logit*

This paper proposes a dichotomous choice model that is based on a transformed beta (or "z") distribution. This model, called betit, nests both logit and probit and allows for various skewed and peaked disturbance densities. Because the shape of this density affects the estimated relation between the dichotomous choice variable and its determinants, the greater flexibility of the transformed beta distribution is useful in generating more accurate representations of this relationship. The paper considers asymptotic biases of the logit and probit models under conditions where betit should have been used. It also investigates small sample power and provides two examples of applications that illustrative of the capability of the betit model.

JEL Classification: C25, J22

Keywords: Dichotomous choice model, beta distribution, logit, probit

Wim P.M. Vijverberg
School of Social Sciences
University of Texas at Dallas
P.O. Box 630688
Richardson, TX 75083-0688
Tel.: +1-972-883-2042
Fax: +1-972-883-2735
Email: vijver@utdallas.edu

* Financial support by IZA, Bonn, for work on this paper is gratefully acknowledged. I have benefited from the capable research assistance of Vadym Lepetyuk, as well as the comments of participants of a seminar at IZA where an earlier version of this paper was presented.

Betit: A Family That Nests Probit and Logit

1. Introduction

It is known fact that estimates of dichotomous choice models like probit and logit are sensitive to misspecification of the distribution of the disturbances that drive the data generating process. These models assume that the disturbances are identically and independently distributed and are independent of the explanatory variables of the model. Yatchew and Griliches (1985) illustrate the bias in probit estimates that results from heteroskedasticity and omitted variables. There is a parallel literature on truncated and censored regression models that points out similar results.[1] This literature also experiments with different statistical distributions and designs tests for normality of the disturbance term of the limited dependent variable model (e.g., Bera, Jarque, and Lee, 1984).

The focus of this paper is on the dichotomous choice model—but the generalization proposed here is applicable to censored and truncated regression models as well. Within the economics literature, there is little said about the choice between logit and probit. While bias results from assuming the wrong distribution (e.g., Robinson, 1982; White, 1982), there are typically few theoretical arguments to guide the choice between the two models (Amemiya, 1981). The logistic distribution has slightly thicker tails than the normal, but how often can one argue that a disturbance of a particular data generating process should have thicker-than-normal tails? A more plausible argument might be that, in many cases, the disturbance captures so many different influences that, by virtue of the central limit theorem, a normal distribution assumption seems warranted. However, what constitutes "many," and are these separate influences more or less similarly distributed such that the central limit theorem indeed applies? When one utilizes

[1] E.g., Arabmazar and Schmidt (1981, 1982), Goldberger (1983), Hurd (1979), Robinson (1982), Vijverberg (1987).

individual-level or household-level data, there may be a few dominant factors that violate the conditions for the application of the central limit theorem.

This paper offers a family of distributions that nests both the logistic and normal distributions (Section 2). This family allows for variations in skewness and kurtosis, which, as shown in Section 3, impacts the relationship between the dichotomous choice variable and its determining factors. As the family is derived from the beta distribution, the associated dichotomous choice model is logically named betit. There are two new parameters to estimate, and in the context of maximum likelihood methods, their estimation is relatively straightforward.

The distribution family does have a history in the statistics literature. Its earliest reference is Fisher (1921, 1935), and important work was done by Prentice (1974, 1975, 1976), who in fact also investigated a dichotomous choice model with it. Other references are provided in Section 2.5. To the best of my knowledge, the family has been appeared twice in the economics literature. McDonald and Xu (1995) list it among several other families based on the beta. They use it for fitting the distribution of income and of daily stock returns, as well as in a regression model to estimate the (financial) beta parameter of stocks. Vijverberg (1997) uses two members of the family to generate random numbers and simulate the multivariate normal integral by importance sampling simulation methods.

In this paper, Section 4 offers two applications of the betit model, concerning a labor force participation choice and a choice of employment in the private versus the public sector. Both examples demonstrate the potential of the model: skewness and kurtosis are features of the distribution of the disturbance, and the estimated probability of one or the other choice can differ substantially from that computed with the probit model. Section 5 concludes the paper.

2. The Transformed Beta Distribution

2.1 Construction of the Distribution

In its common form, the probability density function (pdf) of a beta distribution with parameters p and q is

$$g(y) = \frac{1}{B(p,q)} y^{p-1}(1-y)^{q-1} \quad \text{for} \;\; y \in [0,1] \tag{2.1}$$

where $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ and $p,q > 0$. Consider the transformation $x = \ln(y) - \ln(1-y)$: x is defined on the range (-∞,∞). The pdf of x is easily derived to be:[2]

$$\tilde{g}(x) = \frac{1}{B(p,q)} \left(\frac{1}{1+e^{-x}} \right)^p \left(\frac{e^{-x}}{1+e^{-x}} \right)^q \tag{2.2}$$

The moment generating function of x is given by:

$$\begin{aligned} M_{\tilde{g}}(t) &= \int_{-\infty}^{\infty} e^{tx} \tilde{g}(x)dx = \int_{-\infty}^{\infty} \frac{1}{B(p,q)} \left(\frac{1}{1+e^{-x}} \right)^p \frac{\left(e^{-x}\right)^{q-t}}{\left(1+e^{-x}\right)^q} dx \\ &= \frac{B(p+t,q-t)}{B(p,q)} = \frac{\Gamma(p+t)\,\Gamma(q-t)}{\Gamma(p)\Gamma(q)} \end{aligned} \tag{2.3}$$

To obtain moments of this distribution, we must take derivatives of the gamma function. Consider $\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x} dx$. Then (Gradshteyn and Ryzhik, 1980:576):

$$\frac{d\Gamma(r)}{dr} = \int_0^{\infty} e^{-x} x^{r-1} \ln(x) dx = \Gamma(r) y\,(r) \tag{2.4}$$

where one may compute $y\,(r)$ as:

$$y\,(r) = -C - \frac{1}{r} + r \sum_{k=1}^{\infty} \frac{1}{k(r+k)} \tag{2.5}$$

---

[2] In this, by selecting p=q=1 one recognizes the pdf of a logit distribution. This corresponds to a beta pdf $g(y)$=1, the uniform distribution defined over the unit interval.

and C is Euler's constant: $C \approx 0.577216$. This implies that the first two moments equal:

$$t_1 = y(p) - y(q) \tag{2.6}$$

$$t_2 = \left(y(p) - y(q)\right)^2 + y_1(p) + y_1(q) \tag{2.7}$$

where y_1 denotes the first order derivative of y. In general, the n^{th} order derivative of y equals (Gradshteyn and Ryzhik, 1980:944):

$$y_n(r) = (-1)^{n+1} n! \sum_{k=0}^{\infty} \frac{1}{(r+k)^{n+1}} \tag{2.8}$$

In equation (2.6), t_1 is the mean of x and is not equal to 0 unless indeed $p = q$, in which case the distribution is symmetric. Perhaps more importantly, the variance of x, which is denoted as $q_{p,q}^2$ and equals

$$q_{p,q}^2 = t_2 - t_1^2 = y_1(p) + y_1(q) = \sum_{k=0}^{\infty} \left(p+k\right)^{-2} + \sum_{k=0}^{\infty} \left(q+k\right)^{-2}, \tag{2.9}$$

varies with p and q.[3]

Given that we want to compare this transformed beta distribution with the standard normal, it will be necessary to standardize g. Define $z = (x - t_1)/q_{p,q}$ and, from now on, omit the subscripts on t and q. Then the standardized transformed beta pdf is:

$$f(z) = \frac{q}{B(p,q)} \left(\frac{1}{1+e^{-(qz+t)}}\right)^p \left(\frac{e^{-(qz+t)}}{1+e^{-(qz+t)}}\right)^q . \tag{2.10}$$

For the sake of continuity with previous research (Barndorff-Nielsen, Kent and Srrensen, 1982), we shall refer to this as the $z(p,q)$ distribution.

---

[3] For $p = q = 1$, i.e., the logit distribution, we have $q^2 = 2y_1(1) = 2\Sigma_1^{\infty} k^{-2}$. In general, $\Sigma_{k=0}^{\infty} k^{2n} = 2^{2n-1} p^{2n} |B_{2n}| ((2n)!)^{-1}$ (Gradshteyn and Ryshik, 1980:7), where B_{2n} is a Bernouilli number. With $n = 1$ and $B_2 = 1/6$, we find $q_{1,1}^2 = p^2/3$, as is wellknown (e.g., Ben-Akiva and Lerman, 1985:71; Greene, 2000:817).

2.2 Characteristics of the Transformed Beta Distribution

The moment generating function of this distribution is given by:

$$M_f(t) = \frac{e^{-t\,t/q}\, B\left(p + \frac{t}{q}, q - \frac{t}{q}\right)}{B(p,q)} = \frac{e^{-t\,t/q}\,\Gamma\left(p + \frac{t}{q}\right)\Gamma\left(q - \frac{t}{q}\right)}{\Gamma(p)\,\Gamma(q)} \tag{2.11}$$

which is more conveniently written as $M_f(t) = H(h(t))$ where

$$H(h) = \frac{1}{\Gamma(p)\Gamma(q)}\, e^{h} \tag{2.12}$$

$$h(t) = -\frac{t}{q}\, t + \ln\Gamma\left(p + \frac{t}{q}\right) + \ln\Gamma\left(q - \frac{t}{q}\right) \tag{2.13}$$

These definitions imply that $H^{(k)}(h) = H(h)$ for every k and that

$$h_1(t) = -\frac{t}{q} + \frac{1}{q}\left(y\left(p + \tfrac{t}{q}\right) - y\left(q - \tfrac{t}{q}\right)\right) \tag{2.14}$$

and

$$h_k(t) = \frac{1}{q^k}\left(y_{k-1}\left(p + \tfrac{t}{q}\right) + (-1)^k\, y_{k-1}\left(q - \tfrac{t}{q}\right)\right) \quad \text{for } k > 1 \tag{2.15}$$

Because of the definition of t and q^2, we have $h_1(0) = 0$ and $h_2(0) = 1$. This of course implies that $E[z]$=0 and $Var(z)$=1. Let us now then examine the skewness and kurtosis of z. The skewness equals

$$\begin{aligned} m_3 &= \left.\frac{d^3 M}{dt^3}\right|_{t=0} = H(h)\left(h_1^3 + 3h_1 h_2 + h_3\right)\Big|_{t=0} \\ &= \frac{2}{q^3}\left(\sum_{k=0}^{\infty}\frac{1}{(q+k)^3} - \sum_{k=0}^{\infty}\frac{1}{(p+k)^3}\right) \\ &= 2\,\frac{\frac{1}{q^3} - \frac{1}{p^3} + \sum_{k=1}^{\infty}\frac{1}{(q+k)^3} - \sum_{k=1}^{\infty}\frac{1}{(p+k)^3}}{\left(\frac{1}{q^2} + \frac{1}{p^2} + \sum_{k=1}^{\infty}\frac{1}{(q+k)^2} + \sum_{k=1}^{\infty}\frac{1}{(p+k)^2}\right)^{3/2}} \end{aligned} \tag{2.16}$$

If $q \to 0$, the numerator of (2.16) is dominated by $1/q^3$. The term inside the parentheses of the denominator is similarly dominated by $1/q^2$. As a result, as $q \to 0$, $m_3 \to 2$ and the distribution of z is right-skewed. In the same way, as $p \to 0$, $m_3 \to -2$ and z is left-skewed.

For other results, it is useful to find bounds for sums like $\sum_{k=0}^{\infty} \frac{1}{(r+k)^n}$. Because $\frac{1}{(r+k)^n}$ is declining at a diminishing rate, it is easy to see that, at least,

$$\int_r^{\infty} \frac{1}{k^n} dk < \sum_{k=0}^{\infty} \frac{1}{(r+k)^n} < \int_{(r-0.5)}^{\infty} \frac{1}{k^n} dk \,, \qquad (2.17)$$

so that we may write

$$\frac{1}{(n-1)r^{n-1}} < \sum_{k=0}^{\infty} \frac{1}{(r+k)^n} \left(= \frac{1}{(n-1)(r-a_{r,n})^{(n-1)}} \right) < \frac{1}{(n-1)(r-0.5)^{(n-1)}} \,, \qquad (2.18)$$

where $a_{r,n}$ is a number between 0 and 0.5 that depends on r and n and makes the expression in parentheses exactly equal to the sum. The bounds provided in (2.18) imply that $\sum_{k=0}^{\infty} \frac{1}{(r+k)^n} \to 0$ as $r \to \infty$.

Then, consider m_3. As $p \to \infty$, we have

$$m_3 \to 2 \frac{\sum_{k=0}^{\infty} \frac{1}{(q+k)^3}}{\left(\sum_{k=0}^{\infty} \frac{1}{(q+k)^2} \right)^{3/2}} = \frac{\left(q - a_{q,2}\right)^{3/2}}{\left(q - a_{q,3}\right)^2} \,. \qquad (2.19)$$

On the basis of (2.18), if $p=\infty$, m_3 lies in the interval $((q\text{-}0.5)^{3/2} q^{-2}, q^{3/2}(q\text{-}0.5)^{-2})$. For example, for q=5 we have $.38 < m_3 < 0.55$; for q=10, $0.29 < m_3 < 0.35$. For these values, z is right-skewed. As p decreases, m_3 diminishes (see equation (2.16)): whenever $p > (<) q$, $m_3 > (<) 0$. Finally, by equation (2.19), if p and q both approach ∞, m_3 approaches 0.

For the kurtosis of *z*, the moment generating function yields:

$$\begin{aligned} m_4 &= \left.\frac{d^4M}{dt^4}\right|_{t=0} = H(h)(h_1^4 + 6h_1^2h_2 + 3h_2^2 + 4h_1h_3 + h_4)\Big|_{t=0} \\ &= 3 + \frac{3!}{q^4}\left(\sum_{k=0}^{\infty}\frac{1}{(q+k)^4} + \sum_{k=0}^{\infty}\frac{1}{(p+k)^4}\right) \end{aligned} \tag{2.20}$$

Similar to above, we have $m_4 \to 9$ as either $p \to 0$ or $q \to 0$. Moreover, for $p \to \infty$ and $q \to \infty$, $m_4 \to 3$. For other values of *p* and *q*, it is obvious that $3 < m_4 < 9$ and diminishing as *p* or *q* increase.

Table 1 and Figure 1 illustrates the flexibility of the *z*(*p*,*q*) distribution for values of *p* and *q* that will be used in simulations in later sections of this paper. Note that the difference between *p* (or *q*) of 5 and 25 is relatively small, and that the kurtosis value of 3.04 for *p* = *q* = 25 is close to that of the normal distribution. Figure 1 shows pdf's for *p* = 0.3 and various values of *q* and allows a visual comparison with the standard normal pdf. For *p* = *q* = 0.3, the distribution is symmetric but substantially more peaked. Increasing *q* yields left-skewness; if *p* were increased instead of *q* (not shown), the densities would skew in the opposite direction. Moreover, a *z*(5,5) density (not shown) already looks very similar to the standard normal one.

2.3 The Normal Distribution as a Limiting Case

This section proves that the distribution of *z* approaches the standard normal distribution as *p* and *q* approach ∞. First, consider $h_n(0)$ for $p = aq$ and $q \to \infty$.

$$h_n(0) = (n-1)!\frac{(-1)^n\sum_{k=0}^{\infty}\frac{1}{(k+aq)^n} + \sum_{k=0}^{\infty}\frac{1}{(k+q)^n}}{\left(\sum_{k=0}^{\infty}\frac{1}{(k+aq)^2} + \sum_{k=0}^{\infty}\frac{1}{(k+q)^2}\right)^{n/2}} \tag{2.21}$$

which, by (2.18), may be rewritten as:

$$h_n(0) = (n-1)!\frac{(-1)^n \frac{1}{(n-1)\left(aq - a_{aq,n}\right)^{n-1}} + \frac{1}{(n-1)\left(q - a_{q,n}\right)^{n-1}}}{\left(\left(aq - a_{aq,2}\right)^{-1} + \left(q - a_{q,2}\right)^{-1}\right)^{n/2}}$$

$$= (n-2)!\,q^{1-(n/2)}\frac{(-1)^n\left(a - \frac{a_{aq,n}}{q}\right)^{-(n-1)} + \left(1 - \frac{a_{q,n}}{q}\right)^{-(n-1)}}{\left(\left(a - \frac{a_{aq,2}}{q}\right)^{-1} + \left(1 - \frac{a_{q,2}}{q}\right)^{-1}\right)^{n/2}} \tag{2.22}$$

Thus, for $n > 2$ and $q \to \infty$, we have $h_n(0) \to 0$.

Second, since the moment generating function $M(t)$ may be written as $M(t) = H(h(t))$, a Taylor expansion of M around $t = 0$ is given by

$$M(t) = \sum_{n=0}^{\infty} \frac{1}{n!} M_n(0) t^n \tag{2.23}$$

where (Gradshteyn and Ryzhik, 1980:19)

$$M_n(0) = \sum_{i_1, i_2, \ldots, i_k} \frac{n!}{i_1!\, i_2!, \ldots, i_k!} \frac{d^m F}{dh^m} \left(\frac{h_1}{1!}\right)^{i_1} \left(\frac{h_2}{2!}\right)^{i_2} \cdots \left(\frac{h_k}{k!}\right)^{i_k} \Big|_{t=0} \tag{2.24}$$

where $m = i_1 + i_2 + \ldots + i_k$ and $n = i_1 + 2i_2 + \ldots + ki_k$. We have $h_1 = 0$, $h_2 = 1$, and, for $p = aq$ and $q \to \infty$, $h_n \to 0$ for $n \geq 3$. Thus, many terms disappear from (2.24). The only remaining terms contain $(h_2/2!)^{i_2}$ with $2i_2 = n$ and $h_2 = 1$. This means that

$$M_n(0) \to \frac{n!}{(n/2)!}\left(\frac{1}{2}\right)^{n/2} \text{ as } p = aq \text{ and } q \to \infty \tag{2.25}$$

when n is even and $M_n(0) = 0$ for odd values of n. Thus,

$$
\begin{aligned}
M(t) &= \sum_{n=0}^{\infty} \frac{1}{n!} M_n t^n = \sum_{n=0}^{\infty} \frac{1}{(2n)!} M_{2n} t^{2n} \\
&= \sum_{n=0}^{\infty} \frac{1}{(2n)!} \frac{(2n)!}{n!} \left(\frac{1}{2}\right)^n (t^2)^n \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{1}{2} t^2\right)^n \\
&= e^{t^2/2}
\end{aligned}
\tag{2.26}
$$

which is the moment generating function of the standard normal distribution.

2.4 The Cumulative Distribution Function of the Transformed Beta Distribution

To employ the transformed beta distribution in limited dependent variable applications, it is imperative to have an accurate approximation of the cumulative distribution function (cdf). Let F denote the cdf of z, and let G be the cdf of a beta-distributed random variable y, as in equation (2.1). Then:

$$
F(Z) = \int_{-\infty}^{Z} f(z)dz = \int_{0}^{Y} g(y)dy = G(Y) \tag{2.27}
$$

where

$$
Y = \frac{1}{1 + e^{-(qZ+t)}}
$$

An numerical approximation algorithm of $G(Y)$ is found in Spanier and Oldham (1987, Ch. 58).[4] Write $G(Y)$ as the ratio of the incomplete beta function $B(p,q;Y)$ and the complete beta function $B(p,q)$. Define $f_0 = 0$, $t_0 = Y^p(1-Y)^q / p$, $f_j = f_{j-1} + t_{j-1}$, and $t_j = t_{j-1}Y(p+q+j-1)/(p+j)$, where j = 1, 2, ... Compute f_j and t_j iteratively until $\left| t_j Y(q-1) / \left[f_j j(1-Y)^2 \right] \right| < 10^{-8}$; refer to this value of j as J. The approximate solution is then

[4] Johnson, Kotz, and Balakrishnan (1995, section 25.6.1) present several other approximations. There is always a desire for lower computational burden and greater speed, but faster algorithms may not be as accurate generally. Some do not work for values of p or q between 0 and 1.

found as $B(p,q;Y) = f_J + t_J[1+(q-1)Y/J(1-Y)]/(1-Y)$. Spanier and Oldham note that this solution does not work as well for values of Y near 1. For that reason, this algorithm should be used for $Y < 0.7$, say. If $Y \geq 0.7$, we may write $B(p,q;Y) = B(p,q) - B(q,p,1-Y)$ and evaluate $B(q,p;1-Y)$ with the algorithm above.

If q is an integer and p is any positive real number, an exact solution of $G(Y)$ is given by:

$$G(Y) = \frac{1}{p+q}\left(\sum_{r=1}^{q-1} g\left(Y; p+r, q+1-r\right)\right) + Y^{p+q-1} \tag{2.28}$$

A similar formula exists when p is an integer and q is any positive real number. Equation (2.28) is useful to gauge the precision of the algorithm above.

This study uses the function that is built into Gauss 3.2. A check of Gauss's function by means of the Spanier and Oldham algorithm with a stoppage criterion of 10^{-8} over a wide range of values of p, q, and Y revealed showed for the most part virtual correspondence: (1) the largest absolute difference between the two results was no larger than 10^{-8}; (2) the proportional difference in G was no larger than 10^{-3}; and (3) the largest proportional difference in $1-G$ equaled 2. When the stoppage criterion on the Spanier and Oldham algorithm was lowered to 10^{-12}, the differences diminished to 10^{-10}, 10^{-6}, and 0.7 respectively. The only worrisome inaccuracy is measured with regard to the proportional difference in $1-G$, which occurs in extremely skewed beta distributions (p = 0.2, q = 25) when both routine compute $G(Y)$ in the neighborhood of $1-10^{-11}$ and one algorithm approaches 1 faster than another. The fact that the outcomes of the Spanier and Oldham algorithm move closer to the Gauss values when the stoppage criterion is tightened is reassuring of the accuracy of the Gauss function. Nevertheless, it is necessary to pay attention to these issues, because numerical optimization of log-likelihood functions is sensitive to numerical inaccuracies in the underlying function routines.

2.5 Summary and Discussion

The family of transformed beta distributions depends on two parameters, p and q. This family contains the logit distribution as a special case, namely for $p = q = 1$. The standard normal distribution is a limiting case when both p and q approach ∞. The distribution allows skewness values between -2 and 2, and kurtosis values ranging from 3 to 9.[5] Thus, it has considerable flexibility that one may wish to exploit in estimating econometric models.

The transformed beta distribution has a history in the statistics literature. It first appeared in work by Fisher (1921, 1935) under the name of a z distribution. It was, in essence, an unstandardized transformed beta distribution where the mean and spread were inherent functions of p and q. It was Prentice (1975) who introduced a location and scale parameter to make it a more general density that is suitable for density fitting and regression analysis. There is a difference between his notation (which appears in Barndorff-Nielsen, Kent and Srrensen (1982) and McDonald and Xu (1995) as well) and the exposition in Section 2.1: the 'parameters' t and q are functions of p and q and serve to fix location and spread at 0 and 1, rather than to generalize it. Consequently, the parameters of a regression model (intercept, slope, and standard deviation of the disturbance) are not affected by p and q; they are, in a sense, orthogonal to p and q and therefore more easily estimated.

The transformed beta distribution is more than a mere data fitting tool. Barndorff-Nielsen, Kent and Srrensen (1982) point out that it is a "normal variance-mean mixture." More specifically, let h be a random variable with a probability distribution $\hat{G}$ defined for $h \geq 0$. z is called a normal variance-mean mixture with mixing distribution $\hat{G}$ if, given h, z is distributed $N(m_1 + m_2 h, m_3 h)$,

---

[5] Another family that allows skewness and kurtosis is the Gram-Charlier density (e.g., Ord, 1972, Ch. 2; Kendall and Stuart, 1977, Ch. 6). The range of skewness and kurtosis is not as large, and the additional parameters are subject to a nonlinear parameter restriction in order to ensure that the density is globally nonnegative (Barton and Dennis, 1952).

where m_1, m_2 and m_3 are constants with $m_3 > 0$. The transformed beta distribution is therefore a mixture of normal distributions with a specific mixing distribution. For $m_2 = 0$, the unconditional distribution of z is symmetric, and it becomes left-skewed (right-skewed) if $m_2 < 0\ (> 0)$. This implies that the disturbance of a regression model can be a draw from one of various normal distributions, each with their own location and spread: for some sample members there may be more underlying factors than for other members. The overall mix may not be normal and may indeed be skewed. Together with the required specification of $\widehat{G}$ (which itself contains two parameters), m_1, m_2 and m_3 create a transformed beta distribution with a general mean and standard deviation. For future research, this suggests an avenue to generalize disturbance distributions: mix normals with convenient mixing distributions $\widehat{G}$.

3. A class of limited dependent variable models

3.1 The Betit Model

This section applies the transformed beta distribution to limited dependent variable models. These models center around an index function

$$Y_i^* = X_i' b + u_i \tag{3.1}$$

The dichotomous outcome variable Y relates to Y_i^* as follows:

$$\begin{aligned} Y_i &= 1 \quad \text{iff} \quad Y_i^* \geq 0, \text{ i.e., } u_i \geq -X_i' b \\ &= 0 \quad \text{iff} \quad Y_i^* < 0, \text{ i.e., } u_i < -X_i' b \end{aligned} \tag{3.2}$$

Define $P = \text{Prob}\left[Y = 1\right]$. For a sample of n observations, the loglikelihood function may then be written as:

$$\ln L = \sum_{i=1}^{n} \left\{ Y_i \ln P_i + (1-Y_i)\ln(1-P_i) \right\} \tag{3.3}$$

The common approach is to assume that u_i has a standard normal or a logit distribution. This generates the familiar probit and logit models, respectively. This paper proposes a *z*(*p*,*q*) distribution, yielding a model that, for obvious reasons, we shall name betit. This betit model nests the probit and logit models through the parameters *p* and *q*. Thus, P_i is written as:

$$P_i = 1 - \text{Prob}\left[u_i < -X_i{}'b \right] = 1 - F\left(-X_i{}'b \right) = 1 - \frac{B\left(p,q;Z_i \right)}{B(p,q)} = \frac{B\left(q,p;1-Z_i \right)}{B(p,q)} \tag{3.4}$$

where $B\left(p,q;Z_i\right)$ is the incomplete Beta function with argument

$$Z_i = \frac{1}{1 + e^{-(t - qX_i{}'b)}}$$

For $p = q = 1$, this simplifies to the standardized logit model. When $p = q = \infty$, the probit model results.

The betit model has already been suggested as a generalization to probit and logit by Prentice (1976) and is mentioned without further discussion by Amemiya (1981:1487). There has not been any application of the model, likely because of its computational complexity which, at the time, still was a formidable barrier. Prentice also designs score tests for probit and logit models. For probit this test is complicated by the behavior of the loglikelihood function when *p* and *q* go to ∞. Since *p* and *q* may be estimated by maximum likelihood, standard tests can be used for hypothesized values on $\mathbb{R}^+$ but less than infinite. For large values of *p* and *q*, say at 10 or 20, the difference between *z*(*p*,*q*) and the standard normal distribution becomes minor; indeed, likelihood function values are very similar unless samples are extremely large. Thus, a test of normality could in principle be done by a comparison of an estimated $\text{betit}\left(\hat{p},\hat{q}\right)$ model with betit(10,10) or betit(20,20). Because it matters so little, we shall test by means of likelihood ratio statistics that compare $\text{betit}\left(\hat{p},\hat{q}\right)$ with betit(∞,∞).[6]

3.2 Why Another Dichotomous Choice Model?

Is there actually a need for another dichotomous choice model? The answer to this question is found in the fact that the objective of dichotomous choice models is to quantify the relationship between explanatory variables X and the dichotomous variable Y. The relationship is described by the expression $\text{Prob}[Y=1]=1-F\left(-X'b\right)$ where F is the cdf of the distribution assigned to the disturbance term of the model. Thus, the distributional assumption plays a role in determining how X influences the likelihood of a "success" in the form of $Y = 1$. A more flexible distribution enables one to more accurately describe this relationship.

Breaking loose from the restrictions imposed by the distributional assumption is indeed the objective of semiparametric dichotomous choice models such as Klein and Spady (1993) or Horowitz (1998). The disadvantage of the semiparametric model is their generally slower rate of convergence, their computational burden, and the difficulty presented when some of the explanatory variables are dichotomous. In contrast, the betit model shares the faster ($n^{1/2}$) rate of convergence of probit and logit, is easy to compute, and handles any type of explanatory variables.

The difference with logit and probit models is illustrated in Figure 2, which, along with the normal cdf curve, draws $z(0.3,q)$ curves for q = 0.3, 1, 5 and 25. At the various points along the horizontal $\left(X'b\right)$ axis, the marginal impact on Y clearly differs. Considering that the marginal impact can be computed as

---

[6] In the vast majority of the scenarios represented in Tables 2 and 3 below, the log-likelihood values of betit(10,10) and betit(∞,∞) (i.e., probit) differ by less than 0.50. The only exceptions occur when probit is convincingly rejected anyway and the sample is large.

$$\frac{d\text{Prob}\left[Y=1\right]}{dX}=f\left(-X'b\right)b \tag{3.5}$$

where f is the $z(p,q)$ density defined in (2.10), the variation among the curves in Figure 1 indicates the potential gain from the betit(p,q) model.

It should be noted that if the assumed distributional assumption misspecifies f, the estimate of b will attempt to compensate, such that the marginal impact is, more or less on average across the sample, similar to that of equation (3.5). Indeed, this is the rationale for misspecification bias. Table 2 quantifies this misspecification bias asymptotically for various scenarios. The model is assumed to contain only one explanatory variable, and the data generating process uses $b_0 = 1$ and $b_1 = 1$. The scenarios differ according to the stated values of p and q and the range of the explanatory variable $X \in \left[X_L, X_U\right]$, which is selected such that $\text{Prob}\left[Y = 1 | X\right]$ falls in specified intervals. There are 500 equally spaced X values within each range.

As one would suspect, the asymptotic bias in the logit and probit estimators of b_1 differs according to the data range. Especially when successes are likely (or unlikely) throughout the sample is the bias large. But once again, the relationship between X and Y is more subtle, depending on the cdf of the assigned distribution. Thus, Table 2 also summarizes the difference of the estimated marginal logit and probit impacts and the true betit impact, expressed as a ratio for ease of comparison; the table reports the lowest value of this ratio, the highest value, and the median value. Depending on the values of p and q and the data range, the marginal impact may be underestimated by as much as 40 percent or overestimated by more than 100 percent. The median value of the ratio is mostly close to 1, again as one would expect, though in a few cases even the median impact is overestimated by 10 percent.

3.3 Small Sample Power

In a large sample, the estimated probit or logit relationship between X and Y can be substantially biased when the true data generating process contains transformed-beta disturbances. The next question is whether the difference between the true betit and the hypothesized probit or logit model can be detected when the sample is small.

The first order of business is to produce estimates of the parameters of the regression model, including ones for p and q. From experimentation, it is clear that the log-likelihood function is not always globally concave. In particular, the concentrated log-likelihood function $\ln L(p,q)$ sometimes shows locally convex shapes. Furthermore, the likelihood function is virtually flat for larger values of p and q, and beta probabilities are difficult to compute in the tail for small p and/or q. This creates convergence problems both when p or q is small and when p or q grows large. Because little is gained empirically by letting p and/or q rise without bounds (see Figure 1), these parameters are restricted *a priori* to the range [0.25, 10].

Table 3 examines the power of the betit model over probit and logit under the various data scenarios. The underlying model contains a single explanatory variable X; the Prob[Y=1] range determines the length of the interval of X. The individual observations of X_i are equally spread over the interval. For each scenario, 100 random samples are created with the stated $z(p,q)$ distribution.[7] The table indicates the proportion of these 100 runs that the probit or logit model is rejected at the 5 percent level of significance.[8]

Generally, the betit model is the easiest to distinguish when the data generating process uses highly skewed disturbances (the third line of each Prob[Y=1] group). The greatest power is found

[7] In practice, a random $z(p,q)$ disturbance is computed as a transformation from a uniform random draw. The 100 simulated samples of $z(p,q)$ disturbances for each scenario start with the same seed and therefore use the same uniform random values. This ensures a degree of comparability across scenarios.

[8] Note that each of these power values are themselves outcomes of random variables. This explains the occasional inconsistency where an increase in the sample size seems to lower the power of the likelihood ratio test.

when the probability of success varies over the full unit range, i.e., when the sample is balanced. When success is quite likely for all sample members (or, by symmetry, quite unlikely), or when success is not well defined for any subsample, the betit model has a capacity to differentiate only when the sample is moderately large and the distribution is sharply skewed.

4. Applications

To get a sense of how the betit model might modify outcomes of limited dependent variables models, let us turn now to a couple of illustrative examples with real rather than simulated data. The first considers women's labor force participation and uses data of a sample of married white women drawn from the Panel Study of Income Dynamics in 1975, as used by Mroz (1987) and made available through Berndt (1991). Table 4.A defines the variables and their descriptive statistics. Other Income is defined as household income minus the wife's own earnings. Other variables are self-explanatory.

The probit and logit models achieve virtually identical maximized log-likelihood values. The betit model comes up with one that is statistically superior to both at the 10 percent level. As it turns out, unrestricted estimation of p and q iterates toward a value of $(p, q) = (0.68, 0.21)$ before stalling; restriction to the parameter space of $0.25 \leq p, q \leq 10$ leads to a corner solution at $\hat{p} = 0.8357$ and $\hat{q} = 0.25$. This suggests that the distribution of the disturbance is strongly peaked and mildly right-skewed. As mentioned in Section 3, this impacts the relationship between the determining factors and the labor force participation outcome. The difference between the probit and betit probability of labor force participation across women in the sample can be as high as 7 percentage points either way.[9] Not reported in the table are two other specifications of the model. One adds the log of the woman's wage (observed if working, imputed if not working). This model yields $(\hat{p},\hat{q})=(10,0.25)$, which is significantly different from either logit or probit at the 1 percent significance level; it generates greatly probit and betit probabilities that differ by as much as 25 percentage points both ways.[10] Another specification contains the log of husband's wage, nonwage income, and the imputed value of the log of the wife's wage. It converges to values of $(\hat{p},\hat{q})=(10,3.9)$ and was not significantly different from logit or probit estimates.

As a second illustration, Table 5 looks at sectoral employment choices in Tanzania. The sample derives from the National Urban Mobility, Employment, and Income Survey of Tanzania (NUMEIST), conducted in 1971 (see Vijverberg and Zeager, 1994). This survey collected a random sample of households in seven urban areas including the capital city of Dar es Salaam (Sabot 1979). The education variables describe the schooling attainment as cumulative dummy variables; for parsimony, the limited dependent variables models assume a constant effect for the three highest levels. Compared to probit and logit, the estimates of the betit model indicate a significant effect of marital status and residence in Dar es Salaam, and insignificant values for schooling in Standard 1-4 and religious orientation. The likelihood function reaches its maximum at the imposed boundary value of $(\hat{p},\hat{q})=(0.20,10)$, indicating a sharply right-skewed and peaked disturbance distribution.[11] The probit probability of public sector employment differs by as much as 15 percentage points from the betit probability. However, the fit of the betit model is not

[9] By comparison, the probit and logit probabilities are nearly identical: the probit and logit models are fully equivalent.
[10] One might make the argument that the rejection of the logit and probit models owes to an objectionable treatment of the wage variable—but this is the way labor force participation models used to be estimated at one time. The point is that the betit model is capable of determining skewness and kurtosis in the disturbances, and that predicted probabilities can be greatly influenced.
[11] For this sample, the lower boundary of $p = 0.2$ did not appear to yield convergence problems, unlike other estimation runs made for this paper. The estimated probability of public sector employment ranges from 0.0036 to 0.9439 and is therefore not as extreme.

statistically superior to probit or logit: the likelihood ratio test does not reach the 10 percent significance level. In a more elaborate specification of the model, one that addresses the focus of the Vijverberg-Zeager paper,[12] the probit model is rejected by the betit model, once again with $(\hat{p},\hat{q})=(0.20,10)$, at the 6.5 percent significance level. The probit probabilities deviate by up to ±22 percentage points. A scatterplot of the betit and probit probabilities (Figure 3) illustrates that the deviation between the probabilities contain systematic patterns that may impact the substantive outcome of the analysis.

In closing, it is worth noting that, for both applications, the semiparametric models of Klein and Spady (1993) and Horowitz (1998) would not converge to meaningful solutions. Experimentation suggested that this likely owes to the large number of dichotomous explanatory variables in the model.

5. Concluding Remarks

This paper proposes a family of dichotomous choice models that constitute a generalization of probit and logit. Its advantage is the greater flexibility of the distribution of the disturbance term. The shape of this distribution impacts the relation between the dichotomous choice variable and its determinants.

Prentice (1976) makes the point that it is important to estimate the shape of this distribution precisely in order to make accurate predictions of the probability of success outside the range of observation. This is obviously valuable in the context of dose response functions, but similarly one

[12] The sectoral choice equation is identical to specification 1 of Table 1 in Vijverberg and Zeager (1994). It adds information on father's occupation, migration status, whether the migrant still had land in the place of origin, the number of years since leaving the village and its square, and a number of interaction terms with migrant and land status.

encounters policy questions in the social sciences that require answers to "what if" questions that would take some members of the target population outside the range observed in the sample.

Even within the range of observation, one desires to get accurate estimates of the impact of determining factors on the dichotomous choice variable. Assuming the wrong distribution leads to specification bias and incorrect perceptions about this impact. The paper offers two examples where the probit probability of a particular choice being made differed by as much as 20 percentage points from the probability computed from the statistically preferred betit model.

The betit model has one main drawback, a problem that has long been recognized in the context of comparing logit and probit models (Cox and Chambers, 1967). The sample generally needs to be large in order to differentiate the various alternatives. Here, "large" means a sample of, preferably, several thousands of observations. Substantial skewness and peakedness is noticeable with fewer observations, provided the sample is more balanced and the probability of a success varies over the full unit range.

References

Amemiya, Takeshi, (1981). "Qualitative Response Models: A Survey." *Journal of Economic Literature*, December, 19:4, 1483-1536.

Arabmazar, A., and P. Schmidt (1981). "Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity." *Journal of Econometrics*, November, 17:2, 253-258.

Arabmazar, A., and P. Schmidt (1982). "An Investigation of the Robustness of the Tobit Estimator to Non-normality." *Econometrica*, July, 50:4, 1055-1064.

Barndorff-Nielsen, O., J. Kent, and M. Srrensen (1982). "Normal Variance-Mean Mixtures and *z* Distributions." *International Statistical Review*, 50, 145-159.

Barton, D.E., and K.E. Dennis (1952). "The Conditions Under Which Gram-Charlier and Edgeworth Curves are Positive Definite and Unimodal." *Biometrika*, 39, 425-427.

Ben-Akiva, Moshe, and Steven R. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. (Cambridge, Mass: The MIT Press).

Bera, Anil K., Carlos M. Jarque, and Lung-Fei Lee (1984). "Testing the Normality Assumption in Limited Dependent Variables." *International Economic Review*, October, 25:3, 563-578.

Berndt, Ernst R. (1991). The Practice of Econometrics, Classic and Contemporary. (Reading, Mass.: Addison-Wesley Publ. Co.).

Chambers, E.A., and D.R. Cox (1967). "Discrimination between Alternative Binary Response Models." *Biometrika,* 54, 573-578.

Fisher, R.A. (1921). "On the 'Probable Error' of a Coefficient of Correlation Deduced from a Small Sample." *Metron*, 1, 3-32.

Fisher, R.A. (1935). "The Mathematical Distributions Used in the Common Tests of Significance." *Econometrica*, 3, 353-363.

Goldberger, Arthur, (1983). "Abnormal Selection Bias." In S. Karlin and T. Amemiya, eds., *Studies in Econometrics, Time Series, and Multivariate Statistics* (New York: Academic Press).

Gradshteyn, I.S., and I.M. Ryzhik (1980). *Table of Integrals, Series, and Products.* (New York: Academic Press).

Greene, William H. (2000). *Econometric Analysis*. (Upper Saddle River, NJ: Prentice Hall).

Horowitz, Joel L. (1998). *Semiparametric Methods in Econometrics*. (New York: Springer Verlag).

Hurd, Michael (1979). "Estimation in Truncated Samples When There Is Heteroskedasticity." *Journal of Econometrics*, October/December, 11:2-3, 247-258.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions,* Vol. 2. (New York: John Wiley and Sons, Inc.).

Kendall, M., and A. Stuart (1977). *The Advanced Theory of Statistics*, Vol. 1, 4th ed. (New York: MacMillan Publ.).

Klein, Roger W., and Richard H. Spady (1993). "An Efficient Semiparametric Estimator for Binary Response Models." *Econometrica*, 61:2, 387-421.

Mroz, Thomas A. (1987). "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economics and Statistical Assumptions." *Econometrica*, July, 55:4, 765-799.

Ord, J.K. (1972). *Families of Frequency Distributions* (New York: Hafner Publ. Co.).

Prentice, R.L. (1974). "A Log Gamma Model and its Maximum Likelihood Estimation." *Biometrika,* 61, 539-544.

Prentice, R.L. (1975). "Discrimination among some Parametric Models." *Biometrika*, 62:3, 607-614.

Prentice, R.L. (1976). "A Generalization of the Probit and Logit Methods for Dose Response Curve." *Biometrics*, December, 32, 761-768.

Robinson, Peter M. (1982). "On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables." *Econometrica*, January 1982, 50:1, 27-42

Sabot, Richard (1979). *Economic Development and Urban Migration: Tanzania 1900-1971*. (Oxford: Clarendon Press).

Spanier, Jerome, and Keith B. Oldham (1987). *An Atlas of Functions*. (New York: Hemisphere Publ. Corp.).

Vijverberg, Wim P.M. (1987). "Non-normality as Distributional Misspecification in Single-Equation Limited Dependent Variable Models." *Oxford Bulleting of Economics and Statistics*, November, 49:4, 417-430.

Vijverberg, Wim P.M. (1997). "Monte Carlo Evaluation of Multivariate Normal Probabilities." *Journal of Econometrics*, January, 76:1-2, 281-307.

Vijverberg, Wim P.M., and Lester A. Zeager (1994). "Comparing Earnings Profiles in Urban Areas of an LDC: Rural-to-Urban Migrants vs. Native Workers." *Journal of Development Economics*, December, 45:2, 177-199.

White, Halbert (1982). "Maximum Likelihood Estimation of Misspecified Models." *Econometrica*, January 1982, 50:1, 1-26.

Yatchew, Adonis, and Zvi Griliches (1985). "Specification Error in Probit Models." *Review of Economics and Statistics*, February, 67:1, 134-139.

Table 1: Skewness and Kurtosis of the *z*(*p*,*q*) Distribution for Various Values of *p* and *q*

p	*q* 0.3	1	5	25
	Skewness			
0.3	0.00	-1.41	-1.71	-1.75
1	1.41	0.00	-0.92	-1.10
5	1.71	0.92	0.00	-0.35
25	1.75	1.10	0.35	0.00
	Kurtosis			
0.3	5.48	6.89	7.78	7.92
1	6.89	4.2	4.87	5.29
5	7.78	4.87	3.22	3.32
25	7.92	5.29	3.32	3.04

Table 2: Asymptotic Comparison of Logit, Probit, and Betit

Scenario		logit				probit			
		asy.value	$\Delta P_{logit} / \Delta P_{betit}$			asy.value	$\Delta P_{probit} / \Delta P_{betit}$		
p	q	of b_1	min	max	med	of b_1	min	max	med
Range of P[Y=1] = (0.01–0.99)									
0.3	0.3	1.023	0.815	1.178	1.029	1.023	0.599	1.422	1.098
0.3	1	0.972	0.414	2.343	0.996	0.966	0.191	2.788	1.067
0.3	5	0.915	0.435	1.867	0.952	0.901	0.222	2.123	1.008
1	1	1.000	1.000	1.000	1.000	1.018	0.814	1.206	1.053
1	5	0.968	0.611	1.662	0.963	0.983	0.352	1.904	1.013
5	5	0.972	0.882	1.086	0.966	1.005	0.974	1.042	1.012
Range of P[Y=1] = (0.333–0.99)									
0.3	0.3	1.007	0.832	1.154	1.001	0.980	0.664	1.360	1.035
0.3	1	1.430	0.791	1.438	1.065	1.500	0.869	1.415	1.078
0.3	5	1.526	0.633	1.573	1.062	1.654	0.708	1.570	1.078
1	1	1.000	1.000	1.000	1.000	1.002	0.835	1.189	1.022
1	5	1.209	0.777	1.290	1.011	1.273	0.885	1.250	1.027
5	5	0.978	0.889	1.085	0.987	1.002	0.977	1.040	1.006
Range of P[Y=1] = (0.5–0.99)									
0.3	0.3	0.964	0.799	1.119	1.003	0.904	0.653	1.302	1.041
0.3	1	1.605	0.852	1.435	1.030	1.604	0.934	1.269	1.044
0.3	5	1.878	0.678	1.688	1.036	1.957	0.768	1.559	1.043
1	1	1.000	1.000	1.000	1.000	0.964	0.843	1.178	1.026
1	5	1.331	0.799	1.356	1.000	1.346	0.916	1.212	1.012
5	5	1.007	0.892	1.118	0.987	0.994	0.968	1.039	1.006
Range of P[Y=1] = (0.667–0.99)									
0.3	0.3	0.919	0.791	1.078	1.005	0.825	0.625	1.236	1.041
0.3	1	1.759	0.907	1.320	1.008	1.645	0.965	1.087	1.027
0.3	5	2.328	0.729	1.687	1.009	2.271	0.828	1.451	1.016
1	1	1.000	1.000	1.000	1.000	0.914	0.804	1.158	1.029
1	5	1.481	0.826	1.378	0.990	1.409	0.943	1.155	1.004
5	5	1.051	0.901	1.164	0.986	0.982	0.957	1.038	1.006
Range of P[Y=1] = (0.333–0.667)									
0.3	0.3	1.183	0.987	1.046	1.009	1.338	0.982	1.066	1.012
0.3	1	1.096	0.814	1.198	1.009	1.239	0.826	1.225	1.012
0.3	5	1.078	0.757	1.228	1.010	1.219	0.769	1.257	1.014
1	1	1.000	1.000	1.000	1.000	1.131	0.995	1.020	1.003
1	5	0.974	0.882	1.097	1.000	1.101	0.898	1.121	1.004
5	5	0.906	0.984	1.004	0.997	1.025	0.999	1.004	1.001

Table 3: Small-Sample Power of Betit Relative to Probit and Logit: Simulations

Scenario		Power of betit at 5 percent significance level against:					
		N=500		N=1500		N=2500	
p	q	probit	logit	probit	logit	probit	logit
Range of P[Y=1] = 0.01 - 0.99							
0.3	0.3	0.26	0.01	0.69	0.15	0.79	0.21
0.3	1	0.90	0.76	1.00	0.99	1.00	1.00
0.3	5	0.92	0.83	1.00	1.00	1.00	1.00
1	1	0.04	0.00	0.24	0.03	0.30	0.07
1	5	0.59	0.36	0.91	0.83	0.99	0.99
5	5	0.02	0.01	0.06	0.03	0.03	0.14
Range of P[Y=1] = 0.33 - 0.99							
0.3	0.3	0.17	0.00	0.53	0.09	0.67	0.08
0.3	1	0.02	0.13	0.13	0.43	0.24	0.77
0.3	5	0.22	0.42	0.70	0.94	0.96	1.00
1	1	0.05	0.01	0.16	0.02	0.15	0.03
1	5	0.01	0.11	0.06	0.40	0.23	0.73
5	5	0.00	0.03	0.03	0.06	0.03	0.13
Range of P[Y=1] = 0.50 - 0.99							
0.3	0.3	0.04	0.01	0.25	0.02	0.35	0.02
0.3	1	0.00	0.05	0.02	0.16	0.04	0.29
0.3	5	0.04	0.14	0.31	0.61	0.64	0.87
1	1	0.03	0.02	0.06	0.01	0.12	0.02
1	5	0.00	0.06	0.05	0.18	0.13	0.40
5	5	0.00	0.04	0.02	0.08	0.02	0.13
Range of P[Y=1] = 0.67 - 0.99							
0.3	0.3	0.00	0.00	0.03	0.00	0.06	0.00
0.3	1	0.00	0.01	0.00	0.04	0.01	0.07
0.3	5	0.00	0.03	0.05	0.23	0.15	0.50
1	1	0.00	0.01	0.01	0.00	0.02	0.02
1	5	0.00	0.00	0.00	0.08	0.03	0.20
5	5	0.00	0.01	0.00	0.03	0.01	0.05
Range of P[Y=1] = 0.33 - 0.67							
0.3	0.3	0.00	0.00	0.00	0.00	0.00	0.00
0.3	1	0.00	0.00	0.00	0.00	0.01	0.01
0.3	5	0.00	0.00	0.00	0.00	0.01	0.01
1	1	0.00	0.00	0.00	0.00	0.00	0.00
1	5	0.00	0.00	0.00	0.00	0.00	0.00
5	5	0.00	0.00	0.00	0.00	0.00	0.00

Table 4: Application: Women's Labor Force Participation

A: Definition and Descriptive Statistics

Variable	Definition	Mean	Std.Dev.	Minimum	Maximum
LFP	Labor force participation in 1975	0.57	0.50	0.00	1.00
KL6	Number of children in household under 6 years of age	0.24	0.52	0.00	3.00
K618	Number of children in household between 6 and 18 years of age	1.35	1.32	0.00	8.00
AGE	Age (years)	42.54	8.07	30.00	60.00
EDUC	Schooling (years)	12.29	2.28	5.00	17.00
URATE	Unemployment rate in the county of residence	8.62	3.11	3.00	14.00
SMSA	Dummy, =1 if person lives in an SMSA	0.64	0.48	0.00	1.00
OTHINC	Other Household Income ($000s)	20.13	11.63	-0.03	96.00

The number of observations is 753.
Source: Panel Study of Income Dynamics: Mroz (1987), Berndt (1991)

B: Estimates of Probit, Logit, and Betit Models

	Probit		Logit (a)		Betit	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
CONST	0.4802	1.00	0.4474	1.01	0.5430	1.59
KL6	-0.8909	-7.78	-0.8167	-7.48	-0.7061	-6.39
K618	-0.0364	-0.90	-0.0352	-0.94	-0.0390	-1.29
AGE	-0.0342	-4.47	-0.0314	-4.44	-0.0255	-4.03
EDUC	0.1566	6.50	0.1430	6.32	0.1210	5.84
URATE	-0.0109	-0.69	-0.0098	-0.68	-0.0094	-0.83
SMSA	0.0246	0.23	0.0210	0.22	0.0390	0.50
OTHINC	-0.0210	-4.49	-0.0193	-4.34	-0.0183	-4.36
p					0.8357	1.01
q					0.2500	(b)
Log-Likelihood	-453.982		-453.949		-451.54	
χ^2 statistic (c)	4.884		4.819			

Notes:

(a) The logit model is standardized by the factor $p/\sqrt{3}$. Thus the logit parameters are comparable to probit and betit ones.

(b) q is restricted to these parameter values on the border of the defined parameter space, which is the location of the maximum of the log-likelihood function over the space.

(c) The χ^2 statistic reports the likelihood ratio test of the given model against the betit model. The 5-percent significant value equals 5.99; the 10-percent significance value is 4.61.

Table 5: Application: Employment in the Public Sector in Tanzania

A: Definition and Descriptive Statistics

Variable	Definition	Mean	Std. Dev	Minimum	Maximum
SECTOR	=1 if employed in the public sector	0.605	0.489	0	1
EDST1	Education: =1 if Standards 1-4 or more	0.826	0.379	0	1
EDST5	Education: =1 if Standards 5-8 or more	0.608	0.488	0	1
EDFM1	Education: =1 if Forms I-IV or more	0.203	0.402	0	1
EDFM5	Education: =1 if Forms V-VI or more	0.024	0.153	0	1
EDUNI	Education: =1 if university	0.011	0.105	0	1
AGE	Age in years	29.614	9.701	14	78
SEX	=1 if female	0.129	0.335	0	1
MARRIED	=1 if married	0.574	0.495	0	1
RELIG	=1 if Christian religion	0.463	0.499	0	1
SKILLED	=1 if skilled occupation	0.644	0.479	0	1
CITIZEN	=1 if citizen of Tanzania	0.961	0.195	0	1
SALAAM	=1 if in Dar es Salaam	0.661	0.473	0	1

The number of observations is 1721.
Source: National Urban Mobility, Employment, and Income Survey of Tanzania (NUMEIST): Vijverberg and Zeager (1994)

B: Estimates of Probit, Logit, and Betit Models

	Probit		Logit (a)		Betit	
parameter	estimate	t-stat	estimate	t-stat	estimate	t-stat
const	-2.1654	-5.57	-1.9606	-5.51	-2.1086	-7.33
EDST1	-0.2376	-2.32	-0.2096	-2.29	-0.1143	-1.52
EDST5	0.4694	5.45	0.4214	5.42	0.3914	5.66
EDFMU (d)	0.3940	5.01	0.3704	4.82	0.4704	4.59
AGE	0.0499	2.56	0.0455	2.59	0.0485	3.21
AGESQ	-0.0467	-1.80	-0.0426	-1.82	-0.0465	-2.27
SEX	-0.0018	-0.02	0.0030	0.03	-0.0011	-0.01
MARRIED	0.0858	1.08	0.0813	1.13	0.1100	1.66
RELIG	0.1555	2.26	0.1331	2.12	0.0722	1.19
SKILLED	0.3864	5.35	0.3416	5.26	0.3021	4.84
CITIZEN	0.8293	4.82	0.7469	4.64	0.5624	5.05
SALAAM	0.1289	1.87	0.1166	1.86	0.1297	2.25
p					0.2000	(b)
q					10.0000	(b)
Log-Likelihood	-1047.38		-1047.67		-1046.50	
χ^2 statistic (c)	1.77		2.36			

Notes:

(a) The logit model is standardized by the factor $p/\sqrt{3}$. Thus the logit parameters are comparable to probit and betit ones.

(b) p and q is restricted to these parameter values on the border of the defined parameter space, which is the location of the maximum of the log-likelihood function over the space.

(c) The χ^2 statistic reports the likelihood ratio test of the given model against the betit model. The 5-percent significant value equals 5.99; the 10-percent significance value is 4.61.

(d) EDFMU = EDFM1 + EDFM2 + EDUNI.

Figure 1: Several *z*(*p*,*q*) density functions, compared to the standard normal pdf

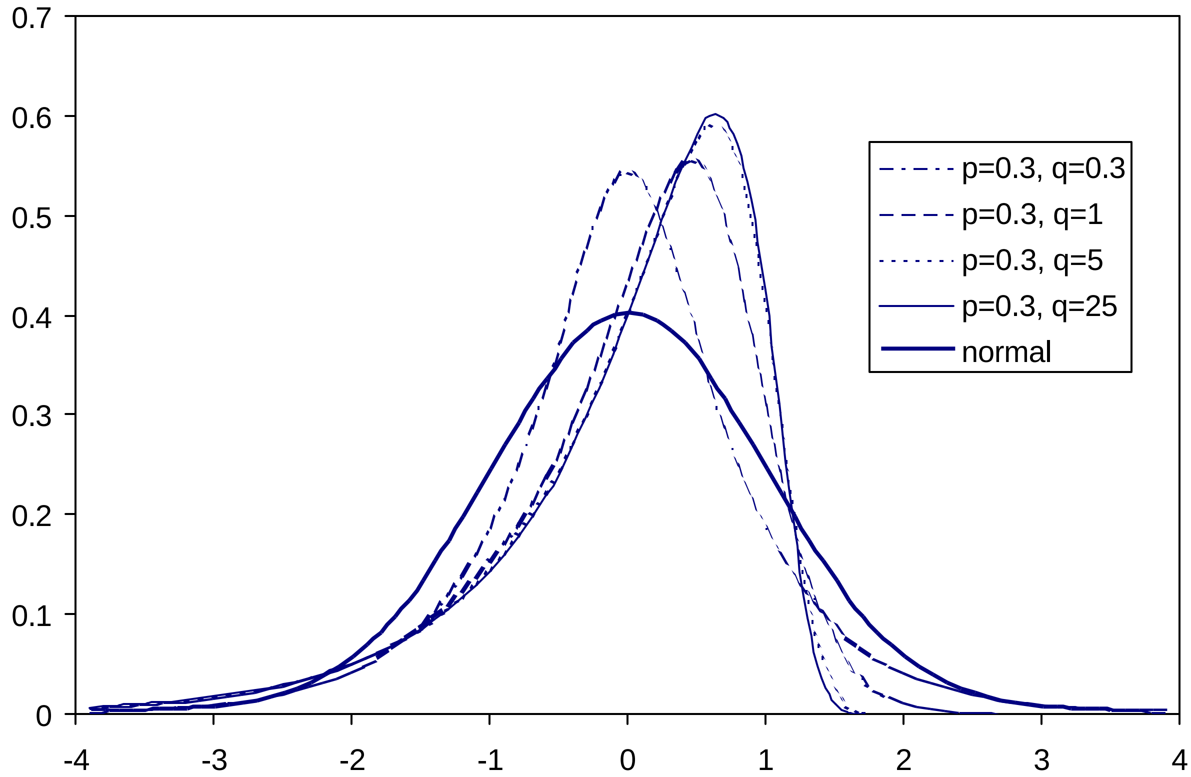


Figure 2: The relationship between $X'b$ and Y, modeled with $z(p,q)$ and normal distributions

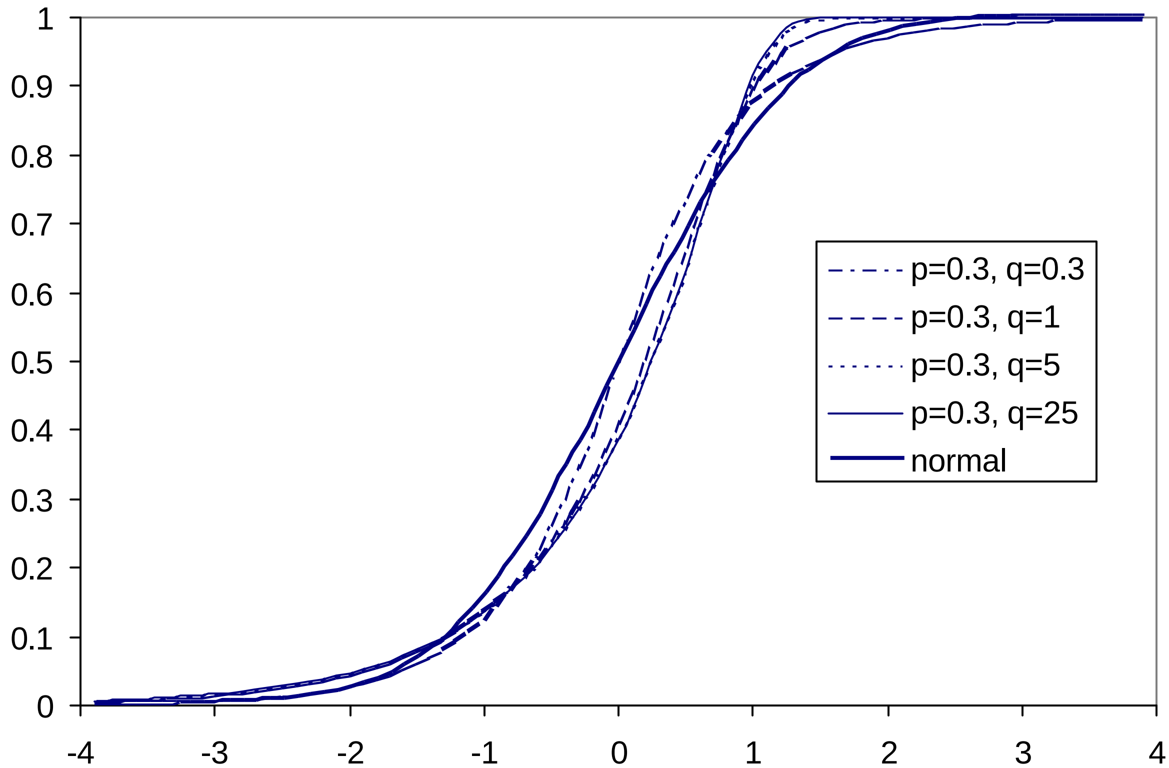
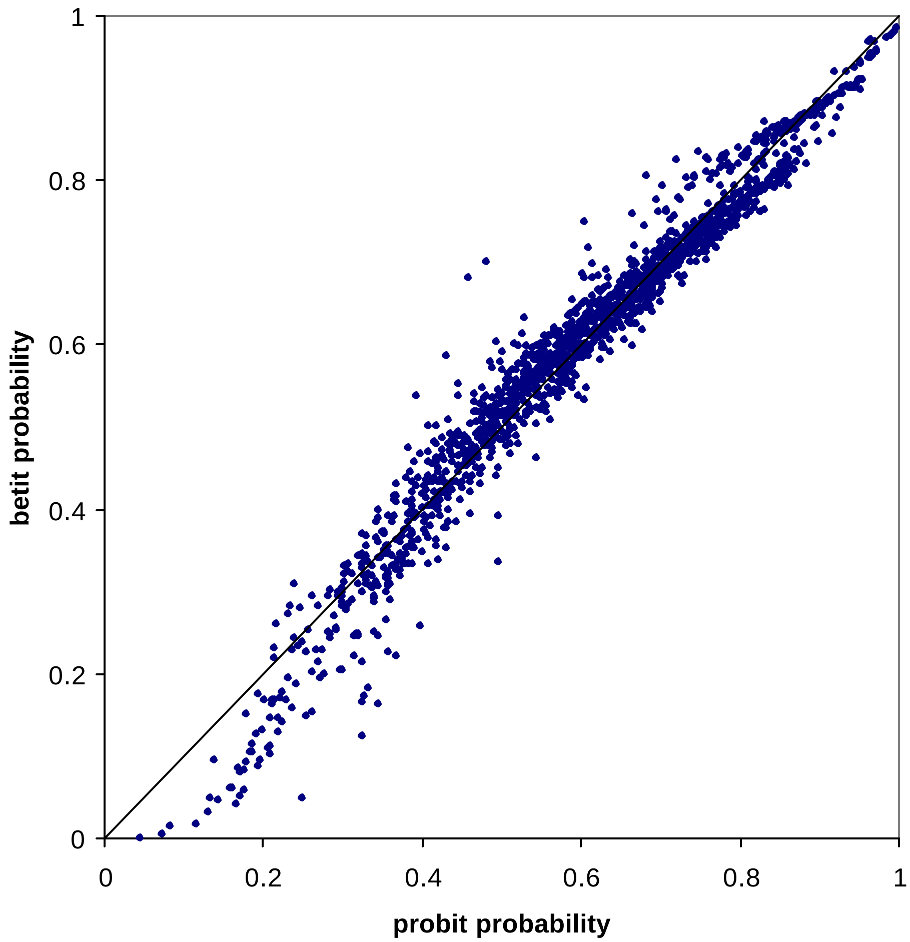


Figure 3: Probability of Employment in the Public Sector, Tanzania 1971: A Comparison of Probit and Betit Probabilities



IZA Discussion Papers

No	Author(s)	Titel	Area	Date
131	B. R. Chiswick	Are Immigrants Favorably Self-Selected? An Economic Analysis	1	3/00
132	R. A. Hart	Hours and Wages in the Depression: British Engineering, 1926-1938	7	3/00
133	D. N. F. Bell R. A. Hart O. Hübler W. Schwerdt	Paid and Unpaid Overtime Working in Germany and the UK	1	3/00
134	A. D. Kugler G. Saint-Paul	Hiring and Firing Costs, Adverse Selection and Long-term Unemployment	3	3/00
135	A. Barrett P. J. O'Connell	Is There a Wage Premium for Returning Irish Migrants?	1	3/00
136	M. Bräuninger M. Pannenberg	Unemployment and Productivity Growth: An Empirical Analysis within the Augmented Solow Model	3	3/00
137	J.-St. Pischke	Continuous Training in Germany	5	3/00
138	J. Zweimüller R. Winter-Ebmer	Firm-specific Training: Consequences for Job Mobility	1	3/00
139	R. A. Hart Y. Ma	Wages, Hours and Human Capital over the Life Cycle	1	3/00
140	G. Brunello S. Comi	Education and Earnings Growth: Evidence from 11 European Countries	2/5	4/00
141	R. Hujer M. Wellner	The Effects of Public Sector Sponsored Training on Individual Employment Performance in East Germany	6	4/00
142	J. J. Dolado F. Felgueroso J. F. Jimeno	Explaining Youth Labor Market Problems in Spain: Crowding-Out, Institutions, or Technology Shifts?	3	4/00
143	P. J. Luke M. E. Schaffer	Wage Determination in Russia: An Econometric Investigation	4	4/00
144	G. Saint-Paul	Flexibility vs. Rigidity: Does Spain have the worst of both Worlds?	1	4/00
145	M.-S. Yun	Decomposition Analysis for a Binary Choice Model	7	4/00
146	T. K. Bauer J. P. Haisken-DeNew	Employer Learning and the Returns to Schooling	5	4/00

147	M. Belot J. C. van Ours	Does the Recent Success of Some OECD Countries in Lowering their Unemployment Rates Lie in the Clever Design of their Labour Market Reforms?	3	4/00
148	L. Goerke	Employment Effects of Labour Taxation in an Efficiency Wage Model with Alternative Budget Constraints and Time Horizons	3	5/00
149	R. Lalive J. C. van Ours J. Zweimüller	The Impact of Active Labor Market Programs and Benefit Entitlement Rules on the Duration of Unemployment	3/6	5/00
150	J. DiNardo K. F. Hallock J.-St. Pischke	Unions and the Labor Market for Managers	7	5/00
151	M. Ward	Gender, Salary and Promotion in the Academic Profession	5	5/00
152	J. J. Dolado F. Felgueroso J. F. Jimeno	The Role of the Minimum Wage in the Welfare State: An Appraisal	3	5/00
153	A. S. Kalwij M. Gregory	Overtime Hours in Great Britain over the Period 1975-1999: A Panel Data Analysis	3	5/00
154	M. Gerfin M. Lechner	Microeconometric Evaluation of the Active Labour Market Policy in Switzerland	6	5/00
155	J. Hansen	The Duration of Immigrants' Unemployment Spells: Evidence from Sweden	1/3	5/00
156	C. Dustmann F. Fabbri	Language Proficiency and Labour Market Performance of Immigrants in the UK	1	5/00
157	P. Apps R. Rees	Household Production, Full Consumption and the Costs of Children	7	5/00
158	A. Björklund T. Eriksson M. Jäntti O. Raaum E. Österbacka	Brother Correlations in Earnings in Denmark, Finland, Norway and Sweden Compared to the United States	5	5/00
159	P.- J. Jost M. Kräkel	Preemptive Behavior in Sequential Tournaments	5	5/00
160	M. Lofstrom	A Comparison of the Human Capital and Signaling Models: The Case of the Self-Employed and the Increase in the Schooling Premium in the 1980's	5	6/00
161	V. Gimpelson D. Treisman G. Monusova	Public Employment and Redistributive Politics: Evidence from Russia's Regions	4	6/00
162	C. Dustmann M. E. Rochina-Barrachina	Selection Correction in Panel Data Models: An Application to Labour Supply and Wages	6	6/00

163	R. A. Hart Y. Ma	Why do Firms Pay an Overtime Premium?	5	6/00
164	M. A. Shields S. Wheatley Price	Racial Harassment, Job Satisfaction and Intentions to Quit: Evidence from the British Nursing Profession	5	6/00
165	P. J. Pedersen	Immigration in a High Unemployment Economy: The Recent Danish Experience	1	6/00
166	Z. MacDonald M. A. Shields	The Impact of Alcohol Consumption on Occupa-tional Attainment in England	5	6/00
167	A. Barrett J. FitzGerald B. Nolan	Earnings Inequality, Returns to Education and Immigration into Ireland	5	6/00
168	G. S. Epstein A. L. Hillman	Social Harmony at the Boundaries of the Welfare State: Immigrants and Social Transfers	3	6/00
169	R. Winkelmann	Immigration Policies and their Impact: The Case of New Zealand and Australia	1	7/00
170	T. K. Bauer K. F. Zimmermann	Immigration Policy in Integrated National Economies	1	7/00
171	C. Dustmann F. Windmeijer	Wages and the Demand for Health – A Life Cycle Analysis	5	7/00
172	D. Card	Reforming the Financial Incentives of the Welfare System	3	7/00
173	D. S. Hamermesh	Timing, Togetherness and Time Windfalls	5	7/00
174	E. Fehr J.-R. Tyran	Does Money Illusion Matter? An Experimental Approach	7	7/00
175	M. Lofstrom	Self-Employment and Earnings among High-Skilled Immigrants in the United States	1	7/00
176	O. Hübler W. Meyer	Industrial Relations and the Wage Differentials between Skilled and Unskilled Blue-Collar Workers within Establishments: An Empirical Analysis with Data of Manufacturing Firms	5	7/00
177	B. R. Chiswick G. Repetto	Immigrant Adjustment in Israel: Literacy and Fluency in Hebrew and Earnings	1	7/00
178	R. Euwals M. Ward	The Renumeration of British Academics	5	7/00
179	E. Wasmer P. Weil	The Macroeconomics of Labor and Credit Market Imperfections	2	8/00
180	T. K. Bauer I. N. Gang	Sibling Rivalry in Educational Attainment: The German Case	5	8/00
181	E. Wasmer Y. Zenou	Space, Search and Efficiency	2	8/00

182	M. Fertig C. M. Schmidt	Discretionary Measures of Active Labor Market Policy: The German Employment Promotion Reform in Perspective	6	8/00
183	M. Fertig C. M. Schmidt	Aggregate-Level Migration Studies as a Tool for Forecasting Future Migration Streams	1	8/00
184	M. Corak B. Gustafsson T. Österberg	Intergenerational Influences on the Receipt of Unemployment Insurance in Canada and Sweden	3	8/00
185	H. Bonin K. F. Zimmermann	The Post-Unification German Labor Market	4	8/00
186	C. Dustmann	Temporary Migration and Economic Assimilation	1	8/00
187	T. K. Bauer M. Lofstrom K. F. Zimmermann	Immigration Policy, Assimilation of Immigrants and Natives' Sentiments towards Immigrants: Evidence from 12 OECD-Countries	1	8/00
188	A. Kapteyn A. S. Kalwij A. Zaidi	The Myth of Worksharing	5	8/00
189	W. Arulampalam	Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages	3	8/00
190	C. Dustmann I. Preston	Racial and Economic Factors in Attitudes to Immigration	1	8/00
191	G. C. Giannelli C. Monfardini	Joint Decisions on Household Membership and Human Capital Accumulation of Youths: The role of expected earnings and local markets	5	8/00
192	G. Brunello	Absolute Risk Aversion and the Returns to Education	5	8/00
193	A. Kunze	The Determination of Wages and the Gender Wage Gap: A Survey	5	8/00
194	A. Newell F. Pastore	Regional Unemployment and Industrial Restructuring in Poland	4	8/00
195	F. Büchel A. Mertens	Overeducation, Undereducation, and the Theory of Career Mobility	5	9/00
196	J. S. Earle K. Z. Sabirianova	Equilibrium Wage Arrears: A Theoretical and Empirical Analysis of Institutional Lock-In	4	9/00
197	G. A. Pfann	Options to Quit	1	9/00
198	M. Kreyenfeld C. K. Spiess G. G. Wagner	A Forgotten Issue: Distributional Effects of Day Care Subsidies in Germany	3	9/00

199	H. Entorf	Rational Migration Policy Should Tolerate Non-Zero Illegal Migration Flows: Lessons from Modelling the Market for Illegal Migration	1	9/00
200	T. Bauer G. S. Epstein I. N. Gang	What are Migration Networks?	1	9/00
201	T. J. Dohmen G. A. Pfann	Worker Separations in a Nonstationary Corporate Environment	1	9/00
202	P. Francois J. C. van Ours	Gender Wage Differentials in a Competitive Labor Market: The Household Interaction Effect	5	9/00
203	J. M. Abowd F. Kramarz D. N. Margolis T. Philippon	The Tail of Two Countries: Minimum Wages and Employment in France and the United States	5	9/00
204	G. S. Epstein	Labor Market Interactions Between Legal and Illegal Immigrants	1	10/00
205	A. L. Booth M. Francesconi J. Frank	Temporary Jobs: Stepping Stones or Dead Ends?	1	10/00
206	C. M. Schmidt R. Baltussen R. Sauerborn	The Evaluation of Community-Based Interventions: Group-Randomization, Limits and Alternatives	6	10/00
207	C. M. Schmidt	Arbeitsmarktpolitische Maßnahmen und ihre Evaluierung: eine Bestandsaufnahme	6	10/00
208	J. Hartog R. Winkelmann	Dutch Migrants in New Zealand: Did they Fare Well?	1	10/00
209	M. Barbie M. Hagedorn A. Kaul	Dynamic Effciency and Pareto Optimality in a Stochastic OLG Model with Production and Social Security	3	10/00
210	T. J. Dohmen	Housing, Mobility and Unemployment	1	11/00
211	A. van Soest M. Das X. Gong	A Structural Labour Supply Model with Nonparametric Preferences	5	11/00
212	X. Gong A. van Soest P. Zhang	Sexual Bias and Household Consumption: A Semiparametric Analysis of Engel Curves in Rural China	5	11/00

213	X. Gong A. van Soest E. Villagomez	Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico	1	11/00
214	X. Gong A. van Soest	Family Structure and Female Labour Supply in Mexico City	5	11/00
215	J. Ermisch M. Francesconi	The Effect of Parents' Employment on Children's Educational Attainment	5	11/00
216	F. Büchel	The Effects of Overeducation on Productivity in Germany — The Firms' Viewpoint	5	11/00
217	J. Hansen R. Wahlberg	Occupational Gender Composition and Wages in Sweden	5	11/00
218	C. Dustmann A. van Soest	Parametric and Semiparametric Estimation in Models with Misclassified Categorical Dependent Variables	1	11/00
219	F. Kramarz T. Philippon	The Impact of Differential Payroll Tax Subsidies on Minimum Wage Employment	5	11/00
220	W. A. Cornelius E. A. Marcelli	The Changing Profile of Mexican Migrants to the United States: New Evidence from California and Mexico	1	12/00
221	C. Grund	Wages as Risk Compensation in Germany	5	12/00
222	W.P.M. Vijverberg	Betit: A Family That Nests Probit and Logit	7	12/00

An updated list of IZA Discussion Papers is available on the center's homepage *www.iza.org*.