Pregibit: A Family of Discrete Choice Models

Chu-Ping C. Vijverberg
Wim P.M. Vijverberg

February 2012

Pregibit:
A Family of Discrete Choice Models

Chu-Ping C. Vijverberg
Wichita State University

Wim P.M. Vijverberg
*City University of New York Graduate Center
and IZA*

ABSTRACT

Pregibit: A Family of Discrete Choice Models

The pregibit discrete choice model is built on a distribution that allows symmetry or asymmetry and thick tails, thin tails or no tails. Thus the model is much richer than the traditional models that are typically used to study behavior that generates discrete choice outcomes. Pregibit nests logit, approximately nests probit, loglog, cloglog and gosset models, and yields a linear probability model that is solidly founded on the discrete choice framework that underlies logit and probit.

JEL Classification: C25, G21, I21

Keywords: discrete choice, asymmetry, logit, probit, post-secondary education,
mortgage application

Corresponding author:

Wim P.M. Vijverberg
Department of Economics
City University of New York Graduate Center
365 5th Avenue
New York, NY 10016-4309
USA
E-mail: wvijverberg@gc.cuny.edu

Pregibit: A Family of Discrete Choice Models

1. INTRODUCTION

When studying dichotomous outcomes of "successes" and "failures," most applied researchers reach for the familiar logit and probit models without giving any thought to the question whether the selected model is appropriate for the data at hand. These two models dictate that responses are symmetric: the marginal effect of an explanatory variable on the probability of success is the same whether the probability of success is $100 \times p$ percent (say, at 30 percent) or $100 \times (1 - p)$ percent (say, 70 percent). Moreover, the choice of the model dictates the magnitude of these marginal effects for different p: those of the logit model are stronger near the median value of $p = 50$ percent and in the tails when p is close to 0 or close to 1, and they are distinctly weaker when p is around 10 percent or 90 percent. But in a study of labor force participation or the purchase of consumer durables or whatever, which of these models is better? There is no way to posit *a priori* that one model is better than the other. In fact, neither of these models may be able to capture the reality that the sample data represent; the relationship may not even be symmetric. Thus, blind reliance on probit and logit models invites instances of model misspecification.

Over the past two decades, econometricians have been developing more flexible models of dichotomous choice. The essence of these models lies in the choice of the function that relates explanatory variables to the dichotomous outcome. On the one hand, this function derives from the distribution of the unobservable factors that drive the outcome—a normal distribution leads to a probit model; a logistic distribution yields a logit model—but there is no reason why this distribution would have to be symmetric.[1] On the other hand, this function may simply be viewed as a link function that connects linear combinations of explanatory variables to dichotomous outcomes, such that a flexible link function yields better predictions of these outcomes. Thus, the aim of these new models is to improve their predictive power and, at the same time, to reduce the likelihood of misspecification that results from a poorly chosen distribution of the unobservables.

The models that are now available in the literature may be classified in various ways. One classification is through the skewness and kurtosis of the underlying assumed distribution. For example, the distributions behind the probit, logit and cauchit models are all symmetric with tails that, between these three models, vary in thickness; loglog and cloglog models have asymmetric distributions with thicker right and left tails, respectively. But as we will see momentarily, the literature offers other, more flexible models that defy this simple classification by skewness and kurtosis.

A second classification method focuses on the number of parameters in the link function: the larger this number of parameters, the more freedom one has in fitting the likelihood of success to its determinants, and the greater is the variation in the shapes of the underlying distribution. A zero-parameter model allows only one shape; examples of such models are probit, logit, cauchit, loglog, and cloglog. A one-parameter model permits shapes with a fixed trade-off of skewness-versus-kurtosis options; this category includes the scobit model of Nagler

---

[1] A normal distribution may be defended on grounds that there are many unobservable factors, with none particularly dominant, such that the aggregate is approximately normally distributed according to Central Limit Theorems. Such assumed conditions are sometimes but not necessarily universally appropriate.

(1994) that generalizes the logit model; the gosset model (Koenker and Yoon, 2009; see also Liu, 2006, who calls it robit), which builds a discrete choice model around the *t*-distribution and then optimizes the degrees of freedom parameter; and the skewed probit model of Bazán et al. (2010) that uses a bivariate normal cdf in a creative way. A two-parameter model opens up a greater range of choices of both skewness and kurtosis; under this category are found the h-family of Stukel (1998) that generalizes the logistic distribution and the betit family of Vijverberg (2000) that builds on the beta distribution.

This paper explores the bounty of another two-parameter model that is derived from the generalized Tukey lambda (GTL) family of distributions (Ramberg and Schmeiser, 1974; Pregibon, 1980). The GTL family is highly flexible (Freimer et al., 1988): the density can be unimodal, U-shaped, J-shaped, symmetric or asymmetric; the range can be finite or infinite; and the tails can be truncated, smooth or abrupt. The resulting discrete choice model nests the logit model (Pregibon, 1980; Koenker and Yoon, 2009), and whether it offers an improvement may be examined with the goodness of link test of Pregibon (1980) or with a likelihood ratio test (Koenker and Yoon, 2009).

We make four contributions to this literature on what we shall refer to as the pregibit model. First, we show that the pregibit model is a highly flexible discrete choice model that not only directly nests the logit model but also approximately nests probit, loglog, cloglog, and gosset models. Furthermore, we find that pregibit offers a version of the linear probability model that is more solidly grounded on the typical discrete choice framework than the standard linear probability model that is estimated with Ordinary Least Squares. Thus, the pregibit model offers a comprehensive family of models.

Second, for each of pregibit's special cases, this paper develops a new Lagrange multiplier (LM) test and extends Pregibon's goodness of link (GL) test. The advantage of the LM and GL tests over the customary likelihood ratio test is that they are based on the estimates of the restricted model and thus are straightforward to add as post-estimation commands to existing software. In particular, the LM test is easier to compute.

Third, in order to illustrate the utility of this new model, we apply the pregibit model to six research problems that span different fields in economics. One of these examines enrollment in post-secondary education. Whereas probit and logit point out types of individuals who virtually surely do (or do not) enroll, the pregibit model finds thicker tails: some high school students will not enroll even if all explanatory variables are in their favor, and vice versa. Moreover, pregibit marginal effects are much stronger for the typical student than probit ones. This shows how imposing the probit structure can color the estimation results. Another application explores denial rates of mortgage applications. Unlike probit, pregibit indicates that even high-risk applicants have a positive probability of getting their mortgage application approved. The GTL density proves to be badly skewed, such that it appears that mortgage officers are more inclined to uncover reasons to make a mortgage deal than to send clients away empty-handed. Overall, among the six applications, the flexibility of pregibit fails to lead to further insights in only two cases. In these, the sample size is smaller—and of course probit or logit may well be adequate in some contexts.

Fourth, these applications yield estimates of the two parameters (α, δ) of the GTL distribution that differ from the range of values that were considered in Freimer et al. (1988) and Koenker and Yoon (2009), which warrants a further examination of the characteristics of the

GTL distribution. In particular, the tails of estimated GTL densities can be so thick that the usual skewness and kurtosis cannot be computed. This feature may be attractive for applications particularly in finance but also in more general contexts. Furthermore, there is not one but actually four sets of (α, δ) values that turn the GTL density into a uniform density. Not only does this create a minor identification issue but it also indicates that the log-likelihood function may have local maxima in a portion of the (α, δ)-parameter space.

In the following, Section 2 builds the pregibit model and discusses how it nests the various special cases. In doing so, it makes a case for standardizing the distribution on which the discrete choice model is built, whether this distribution is normal, logistic, Gumbel, uniform, or GTL. Section 3 examines the asymptotic bias that results from inappropriately specifying logit or probit when pregibit is in order. In Section 4, we apply the pregibit model to six different research questions, ranging from mortgage applications to labor market, education and health outcomes. Section 5 concludes.

2. PREGIBIT: A PARAMETRIC FAMILY OF DICHOTOMOUS CHOICE MODELS

2.1. Pregibit and the core design of dichotomous choice models

The dichotomous choice framework connects a dichotomous outcome variable y to a vector of explanatory variables X through a latent variable equation, which in this paper is assumed to be linear, and an index function:

$$y_i^* = X_i'\beta + \epsilon_i \tag{1}$$

$$y_i = I(y_i^* \geq 0) = I(\epsilon_i \geq -X_i'\beta) \tag{2}$$

where $i = 1, \ldots, n$ indexes individuals. Let $F(\epsilon; \theta)$ be the CDF of ϵ, where θ is vector of parameters that either is specified *a priori* or must be estimated along with β, and let $f(\epsilon; \theta)$ be the associated PDF. Then the probabilities of a "success" $y_i = 1$ and a "failure" $y_i = 0$ are given by, respectively,

$$p_i \equiv P[y_i = 1 | X_i] = 1 - F(-X_i'\beta; \theta) \tag{3a}$$

$$1 - p_i \equiv P[y_i = 0 | X_i] = F(-X_i'\beta; \theta) \tag{3b}$$

Equations (3a)-(3b) are the ultimate expression of the relationship between y and X. In this model, F plays two roles. First, F represents the distribution of the unobservables that drive the choice outcome, captured by ϵ. Theories about the relationship between y and X hardly ever dictate the nature of this distribution (Amemiya, 1981). One might defend an assumption of a normal distribution with an appeal to the central limit theorem: subject to regularity conditions, the combined effect of many unobservable factors is approximately normally distributed. But there must be many unobservables and none must dominate. It is really more out of convenience that researchers might assume that ϵ has a normal distribution; an equally common choice, the logistic distribution, does not have a central limit theorem to argue for it.

Second, separate from (but of course flowing out of) its connection with ϵ, the function F plays a more direct role in the relationship between y and X. As $X_i'\beta$ rises, $-X_i'\beta$ declines, $F(-X_i'\beta; \theta)$ falls, and a success becomes more likely. In other words, F helps translate $X_i'\beta$ into p_i and thereby into a dichotomous (or bernoulli) pattern of outcomes y_i for each individual. Notice that the shape of the *left* tail of the distribution of ϵ determines how $X_i'\beta$ drives the

3

probability of success on the *right* side of the S-shaped $(p_i, X_i'\beta)$ curve, and vice versa. This insight is essential for the proper interpretation of pregibit estimates later on in this paper.

The distribution of ϵ_i is often stated in a "natural" form with a mean μ_ϵ not equal to 0 and a standard deviation σ_ϵ not equal to 1; i.e., it is usually not standardized. In fact, candidate distributions may differ greatly in the value of μ_ϵ and σ_ϵ, which is matched by shifts in the intercept and slopes in β. For example, it is well-known that logit parameters are roughly comparable to probit parameters once the latter are multiplied by the standard deviation of the common logistic distribution, $\pi/\sqrt{3} = 1.81$.[2] Moreover, in the notation in equations (3a)-(3b), variations in θ also cause shifts in μ_ϵ and σ_ϵ, making comparison of estimates of β difficult even within the same family of discrete choice models. Expressing the model in standardized form avoids this problem.[3] Define $\tilde{\epsilon} = (\epsilon - \mu_\epsilon)/\sigma_\epsilon$, and let $\tilde{F}(\tilde{\epsilon}; \theta) = F(\mu_\epsilon + \sigma_\epsilon \tilde{\epsilon}; \theta)$ denote the CDF of $\tilde{\epsilon}$. Let a success occur when the standardized disturbance $\tilde{\epsilon}_i$ exceeds $-X_i'\tilde{\beta}$, or when ϵ_i exceeds $\mu_\epsilon - \sigma_\epsilon X_i'\tilde{\beta} \equiv -X_i'\beta$: equations (3.a) and (3b) are simply restated as

$$p_i \equiv P[y_i = 1|X_i] = 1 - \tilde{F}\left(-X_i'\tilde{\beta}; \theta\right) = 1 - F(-X_i'\beta; \theta) \tag{3a'}$$

$$1 - p_i \equiv P[y_i = 0|X_i] = \tilde{F}\left(-X_i'\tilde{\beta}; \theta\right) = F(-X_i'\beta; \theta) \tag{3b'}$$

The search in the literature for different families of dichotomous choice models arises out of a dissatisfaction with the common probit and logit models that assume normal and logistic distribution, respectively, for ϵ. The desire is to find a flexible functional form for F that imposes fewer restrictions on the estimated relationship between y_i and $X_i'\beta$. Thus, the literature diverges into two directions: one offers new families of dichotomous choice models that are derived from a particular function $F(\cdot; \theta)$; the other branch starts off with a so-called link function G that implies a distribution of ϵ, i.e., $\epsilon_i = G(q_i; \theta)$ with q_i a uniform random variable over the range $[0,1]$. In essence, this aims to translate p_i into $X_i'\beta$ in a flexible way: $X_i'\beta = -G(1 - p_i; \theta)$.[4] Since $G^{-1} = F$, the two approaches are related, but new families of dichotomous choice models may sometimes be more easily specified through G than through F, and vice versa. In all this, what these specifications of $F(\cdot; \theta)$ and $G(\cdot; \theta)$ mean for the pattern of ϵ_i in the data—its distribution, more particularly—is essentially secondary.

Pregibon (1980) proposed the generalized Tukey lambda (GTL) link function as a way to introduce asymmetric disturbances into a linear model and proceeded to explore this link function in the context of a logit model of grouped data. In this section, we build up this discrete choice model with generalized Tukey lambda disturbances, and we refer to this discrete choice

---

[2] Amemiya (1981) suggests a value of 1.6 instead, which more closely links the marginal effects at the center of the distribution. Greene (2012:695) suggests a value of 1.7 instead, as this make the average marginal effect more comparable. Since the marginal effect is intimately related to the PDF, Greene's suggestion amounts to selecting a factor that minimizes the difference between the PDFs, fitting a transformed logistic density to the standard normal density.

[3] Standardization has another benefit. If the distribution is specified in its natural form, the surface of the likelihood function exhibits a ridge-like shape, such that an iterative search may traverse tediously through a narrow range of values of $(\hat{\beta}, \hat{\theta})$: as $\hat{\theta}$ increases, the mean and standard deviation rise, say, which necessitates a parallel adjustment in the searched value of $\hat{\beta}$. Standardization makes these search directions more independent of each other: indeed, in our experience with the pregibit model, likelihood functions formulated with distributions in standardized form (when feasible) converge more quickly and with greater success.

[4] Identically in approach but with a slight difference in the precise formulation, the link function is sometimes specified in the following form: $X_i'\beta = G(p_i)$. We prefer the approach in the text, tying G to F^{-1}.

model as the pregibit model. As will be seen below, the pregibit model fits in the class of two-parameter models. The two parameters endow the pregibit model with a capability to fit data generated with asymmetric disturbances. This is clearly beyond the capability of the popular probit and logit models, which rely on symmetric disturbances that are generated by the standard normal and logistic distribution, respectively. It is also more general than the loglog and cloglog models that are sometimes used when disturbances are presumed to be asymmetric. The pregibit model directly or approximately nests all of these four discrete choice models as well as the gosset model of Liu (2006) and Koenker and Yoon (2009). And finally, it offers a version of a linear probability model that properly fits within the general discrete choice framework.

2.2. Generalized Tukey lambda disturbances

The pregibit model of Pregibon (1980) builds on the generalized Tukey lambda (GTL) distribution that was first proposed by Ramberg and Schmeiser (1974). The versatility of the pregibit model is better understood if the great flexibility of the GTL distribution is clear. Thus, this section augments the literature on GTL distributions. In particular, we will illustrate the GTL distributions for a wider range of parameter values; we will make a case for standardization of the distribution; and we will gain a better perspective on the type of data that facilitate estimation of the pregibit model.

In the notation of the present paper, the GTL link function is given by

$$\epsilon = G(q; \theta) = \frac{q^{\alpha-\delta}-1}{\alpha-\delta} - \frac{(1-q)^{\alpha+\delta}-1}{\alpha+\delta} \tag{4}$$

for $q \in [0,1]$, $\theta = (\alpha, \delta)$, $\alpha - \delta \neq 0$, and $\alpha + \delta \neq 0$.[5] The link function in equation (4) translates a uniformly distributed random variable q into a GTL disturbance ϵ. The inverse of G does not have an analytical solution for general values of α and δ but is of practical value: $q = G^{-1}(\epsilon)$ is the CDF of ϵ since $0 \leq q \leq 1$ and since the derivative

$$\frac{\partial G^{-1}(\epsilon)}{\partial \epsilon} = \frac{1}{\partial G(q)/\partial q} = \frac{1}{q^{\alpha-\delta-1}+(1-q)^{\alpha+\delta-1}} \tag{5}$$

is positive for all values of $q \in [0,1]$. Thus, in terms of notation introduced in Section 2.1, $G^{-1} = F$ and $\partial G^{-1}/\partial \varepsilon = f$.

This link function yields a very rich and diverse family of GTL density functions (Freimer et al., 1988): f can be symmetric or non-symmetric, unimodal or U-shaped, J-shaped or monotone; the tails of f can be smooth or truncated or abrupt and thus can be thick, thin or non-existing; and the range of ϵ can be finite or infinite.

Let us first consider the range of ϵ. When $\alpha > \delta$, the lower bound of the range of ϵ is finite and equals $-1/(\alpha - \delta)$; otherwise, when $\alpha \leq \delta$, the lower bound of ϵ is $-\infty$. Similarly, when $\alpha > -\delta$, the upper bound is finite and equals $1/(\alpha + \delta)$; otherwise, when $\alpha \leq -\delta$, the upper bound of ϵ is ∞. (Proof of these and other properties is provided in Appendix A.) Figure 1 demonstrates this property. In the (α, δ) plane, to the right of the 45 degree line in the direction of the small arrows, the lower bound is finite, and to the right of the 135 degree line in the direction of the large arrows, the upper bound is finite.

---

[5] When $\alpha - \delta = 0$, the first term in (4) is replaced with $\ln q$ according to L'Hôpital's Rule. Similarly, when $\alpha + \delta = 0$, the second term in (4) is replaced with $\ln(1 - q)$.

Figure 1: Feasible parameter regions
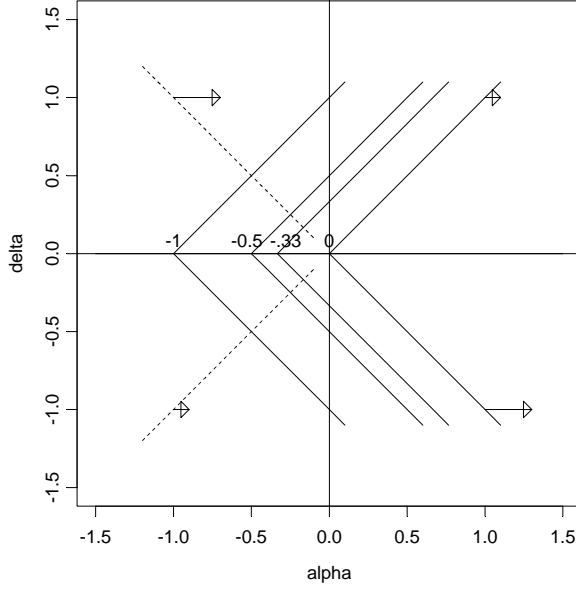
Next, consider four moments of the GTL distribution: mean, variance, skewness and kurtosis. From equations (4) and (5), we may derive the mean and the variance of the GTL distribution as follows.

$$E(\epsilon) = \mu_\epsilon = -\frac{2\delta}{(\alpha+1)^2 - \delta^2} = -\frac{2\delta}{(a+1)(b+1)} \tag{6}$$

$$Var(\epsilon) = \sigma_\epsilon^2 = E(\epsilon^2) - \mu_\epsilon^2 \tag{7}$$

where $a = \alpha - \delta$; $b = \alpha + \delta$; $E(\epsilon^2) = \frac{2}{(2a+1)(a+1)} + \frac{2}{(2b+1)(b+1)} - \frac{2}{ab}\left(B + \frac{(ab-1)}{(a+1)(b+1)}\right)$; and $B = \Gamma(a+1)\Gamma(b+1)/\Gamma(a+b+2)$.[6] This implies that the mean of ϵ equals 0 only when δ equals 0 and that the variance of ϵ can be any positive number.

Now, the mean is defined only for parameter values for which $\alpha - \delta > -1$ and $\alpha + \delta > -1$, or more succinctly, $-\alpha - 1 < \delta < \alpha + 1$. In Figure 1, this constraint is represented by the kinked solid line that intersects with the horizontal axis at $\alpha = -1$: to the right of this kinked line, the mean is a finite number; to the left the expected value of ϵ does not exist. Similarly, the variance exists only if $\alpha - \delta > -\frac{1}{2}$ and $\alpha + \delta > -\frac{1}{2}$, i.e., if, $-\alpha - \frac{1}{2} < \delta < \alpha + \frac{1}{2}$. This constraint shows up in Figure 1 as the kinked solid line that intersects with the horizontal axis at $\alpha = -\frac{1}{2}$: the variance exists only to the right of this kinked line.

The third and fourth moments exist only if $-\alpha - 1/3 < \delta < \alpha + 1/3$ and $-\alpha - 1/4 < \delta < \alpha + 1/4$, respectively (Appendix A). Thus, skewness may be computed for parameter values to the right of the kinked line that intersects the horizontal axis at $\alpha = -1/3$, and kurtosis

---

[6] Note that the equation of the second moment changes in the case of $a = 0$ and/or $b = 0$. Expressions are given in Appendix A.

to the right of a similar kinked line at $\alpha = -1/4$ (not shown in Figure 1). In the following, we shall refer to skewness as κ_3 and to kurtosis as κ_4.[7]

Table 1 reports the four moments for various values of α and δ, or more precisely, for values of $\alpha - \delta$ and $\alpha + \delta$ to make the table more effective. Thus, within each block of the table, on the diagonal we find cases with $\delta = 0$ and α taking on values as indicated in either margin; below the diagonal δ is positive; and above the diagonal δ is negative. Thus, with $\delta = 0$, the distribution is symmetric—both μ_ϵ and κ_3 equal zero—but can range from extremely peaked with long tails ($\sigma_\epsilon = 2.74$, $\kappa_4 = 22.21$) to not peaked at all and all bunched up ($\sigma_\epsilon = 0.29$, $\kappa_4 = 1.80$). A positive δ yields negative skewness, and a negative δ implies positive skewness.[8] Each moment has an asymptote of $+\infty$ or $-\infty$ at the boundary of the feasible parameter area. Outside the feasible area, the moment is not defined—but that does not mean that the distribution does not exist. As we will see later on, applications sometimes yield estimates of α and δ outside these feasible areas.

As seen in Figure 1, the (α, δ)-value of $(0,0)$ is an interesting special case, turning GTL$(0,0)$ into the logistic distribution (Pregibon, 1980). The value of $(0,0)$ is on the boundary of the parameter area where the range of ϵ is infinite in both directions while achieving symmetry and the lowest degree of kurtosis among densities with an infinite range. Freimer et al. (1988) examine the GTL distribution primarily for positive values of $\alpha - \delta$ and $\alpha + \delta$, which corresponds to the area to the right of the kinked line through the origin, and they ignore parameter values for which coefficients of skewness and kurtosis do not exist. Without commenting on coefficients of skewness and kurtosis, Koenker and Yoon (2009) illustrate the GTL distribution for $\alpha = -0.25, 0, 0.25$ and $\delta = -0.25, 0, 0.25$, and they perform simulations for (α, δ)-pairs equal to $(0,0)$, $(-0.25,0)$, $(-0.15,0.15)$, and $(-0.40,0.30)$. These values are all close to $(0,0)$. In our applications, we find estimates of α and δ that are much farther away from these well-studied values.[9] Let us therefore examine the GTL distribution in detail for a broader range of parameter values.

Figure 2 shows two kinds of GTL distributions. In Figure 2A, δ equals 0 and α varies from -2 to $+2$: the CDF curve becomes steeper as the value of α increases. Thus, it seems to imply that, as α rises, the marginal increment in probability increases for a given unit change in ϵ. However, the standard deviation actually decreases substantially when the value of α increases from -0.5 to $+2$; similarly, to the left of $\alpha = -0.5$ where the standard deviation is not defined, the distribution becomes more compact as α increases. Thus, one must be cautious when interpreting the implication of a unit change of the argument of the CDF. In Figure 2B, α equals 0 and δ ranges from -1 to $+1$. Here, the CDF curve shifts leftward as the value of δ increases. For $\delta = 0.5$ and $\delta = 1$, the upper bound is a finite value and the lower tail is not bounded, while for $\delta = -0.5$ and $\delta = -1$, the lower bound is finite and the upper tail is not bounded.

---

[7] For clarity, $\kappa_3 = m_3/\sigma_\epsilon^3$ and $\kappa_4 = m_4/\sigma_\epsilon^4$ where m_k denotes the kth moment around the mean of ϵ.
[8] For larger values of $\alpha + \delta$ and $\alpha - \delta$ with $\alpha + \delta < \alpha - \delta$, the sign of the skewness measure switches. This unexpected phenomenon is also reported in Ramberg (1979:206) and Freimer et al. (1988:3552).
[9] In an application, Koenker and Yoon (2009) also find some of such estimates.

Figure 2: Various Non-Standardized Cases
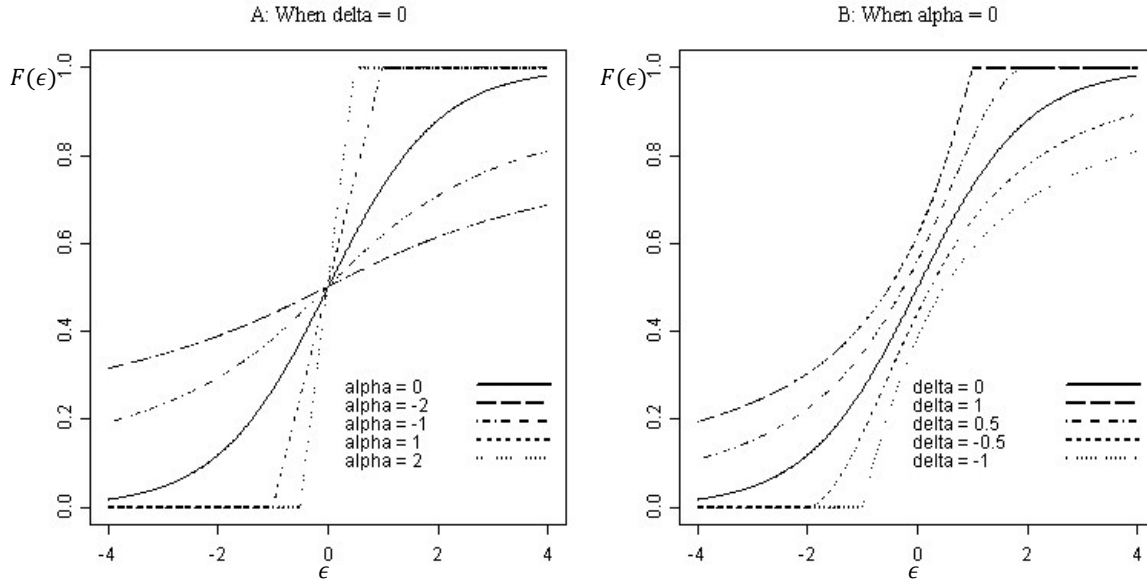
The differences between these distributions are clearly accentuated by the lack of standardization. Let us therefore also compare standardized distributions (i.e., of $\tilde{\epsilon}$), scaled such that the mean is 0 and the standard deviation is 1 for all distributions, which, in view of Figure 1, requires us to restrict (α, δ) to be such that $-\alpha - \frac{1}{2} < \delta < \alpha + \frac{1}{2}$.[10] First, consider the case where $\a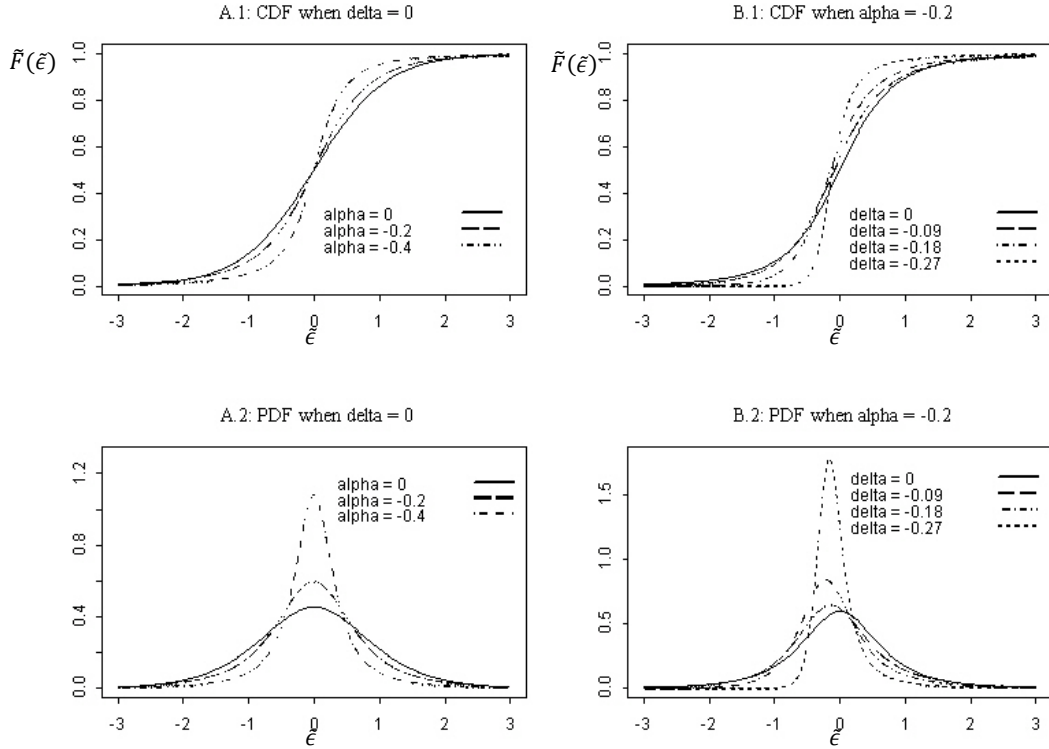lpha$ is negative. Figure 3A.1 indicates that, for $\delta = 0$, the CDF curve becomes flatter in the neighborhood of $\tilde{\epsilon} = 0$ as α increases from -0.4 to 0. Accordingly, the density curves of Figure 3A.2, which are all symmetric since $\delta = 0$, are less peaked and have thinner tails as α increases (since the probability mass in the middle must be balanced with probability mass in the tails to yield the same standard deviation). Figure 3B.1 shows that, in the case $\alpha = -0.2$, the CDF curve steepens as well when δ declines, but the way the CDF approaches 1 as ϵ rises is more gradual than the way it approaches 0 as ϵ falls. The PDF gets a fatter right tail and a thinner left tail as δ decreases from 0 to -0.27: the distribution is right-skewed when δ is negative.

Next, let α be positive. Figure 4A.1 shows the CDFs of varying positive α values ranging from 0 to 1.5 with $\delta = 0$. Of the four curves in Figure 4A.1, a major divergence occurs between $\alpha = 0$ and $\alpha = 0.5$; all of the curves associated with $\alpha = 0.5$, $\alpha = 1$ and $\alpha = 1.5$ stay close to each other except at the tail areas. The corresponding PDF curves in Figure 4A.2 show that, except for $\alpha = 0$, all have finite tails on both sides, which is consistent with Figure 1. Figure 4B.1 illustrates that as δ moves from 0 to -0.75, the CDF rises more abruptly from 0 and approaches 1 more gradually. The left tail of the PDF becomes shorter and the right tail becomes longer (Figure 4B.2). Indeed, given that $\alpha = 0.5$, we are traversing from the right to the left of the 135 degree line through the origin in Figure 1 as δ decreases: note that the upper

---

[10] If this restriction is violated and (α, δ) occurs to the left of the kinked line that intersects the horizontal axis in Figure 1 at $\alpha = -\frac{1}{2}$, standardization with mean and standard deviation is no longer feasible, but as we will see later, the distribution may be standardized in a different way with fruitful results.

Figure 3: Various Standardized Cases with $\alpha \leq 0$

A.1: CDF when delta = 0   B.1: CDF when alpha = -0.2

A.2: PDF when delta = 0   B.2: PDF when alpha = -0.2

bound of $\tilde{\epsilon}$ is infinite on the left of the 135 degree line. Thus, for $\delta = -0.5$ and $\delta = -0.75$, the upper limit of $\tilde{\epsilon}$ is ∞ and the PDF has an infinite right tail.

When either $\delta < -\alpha - 1/2$ or $\delta > \alpha + 1/2$, standardization by μ_ϵ and σ_ϵ is no longer feasible. But a degree of standardization may still be obtained with other, familiar statistics that are available for any distribution: the median m_ϵ as a measure of centrality, and one half of the interquartile range s_ϵ as a measure of spread. Figure 5 illustrates this type of "iqr-standardization" with two GTL distributions that also appeared in Figure 2A, with $(\alpha, \delta) = (0,0)$ and $(-2,0)$, and two that use the more "extreme" values of α and δ of $(-2,-1)$ and $(-3,1)$. In their natural form, the value of s_ϵ equals 1.10, 7.11, 11.60 and 35.16, respectively: these distributions are so spread out that they are difficult to compare. With the iqr-standardization, the CDF- and PDF-curves illustrate the relative skewness and peakedness of the distributions well, even if κ_3 and κ_4 cannot be computed.

2.3. Pregibit: a family of discrete choice models with GTL disturbances

The pregibit model is simply a dichotomous choice model with disturbances that follow the generalized Tukey lambda distribution, which we shall denote as GTL(α, δ). The parameters β and $\theta = (\alpha, \delta)$ are estimated simultaneously, which allows for a rich family of dichotomous choice models that relate success to its determinants in many different ways. In particular, members of this family distinguish themselves in two ways: (i) the impact of $X\beta$ on the probability of success p in the intermediate range of p, and (ii) the shape of the upper and lower tail of the relationship between $X\beta$ and p. In regard to the latter, recall that the *left* tail of the

9

Figure 4: Various Standardized Cases with $\alpha \geq 0$

A.1: CDF when delta = 0

B.1: CDF when alpha = 0.5

A.2: PDF when delta = 0

B.2: PDF when alpha = 0.5
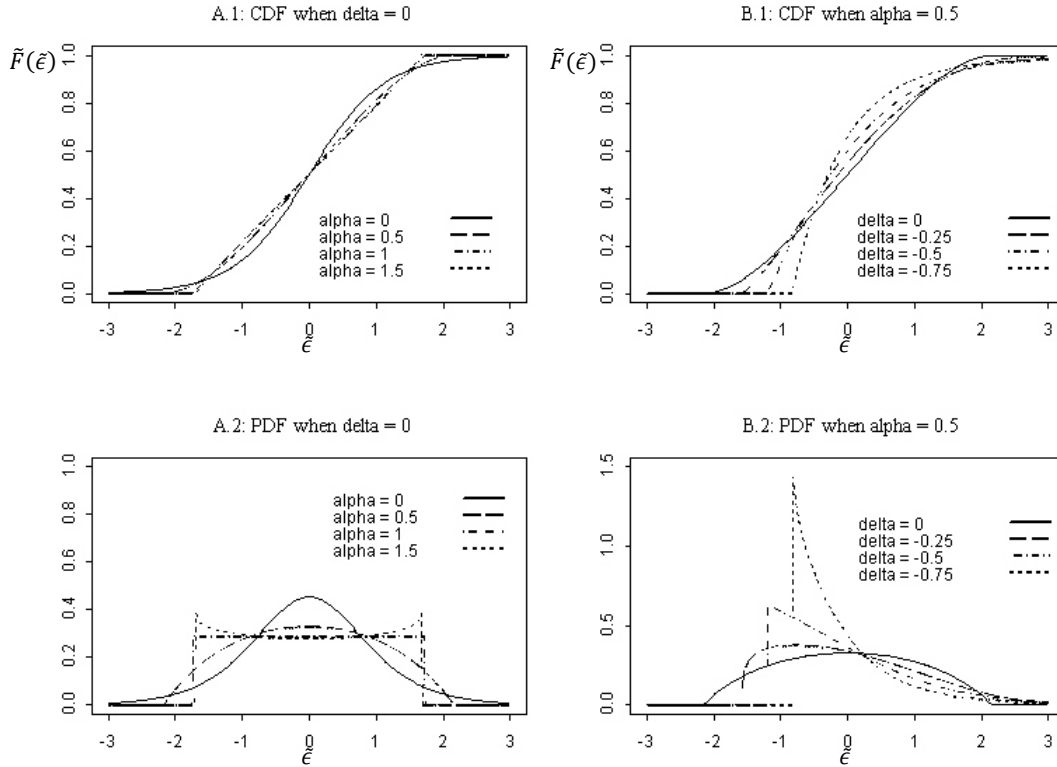
Figure 5: GTL Distributions Standardized with m_ϵ and s_ϵ for various (α, δ)

A: CDF

B: PDF
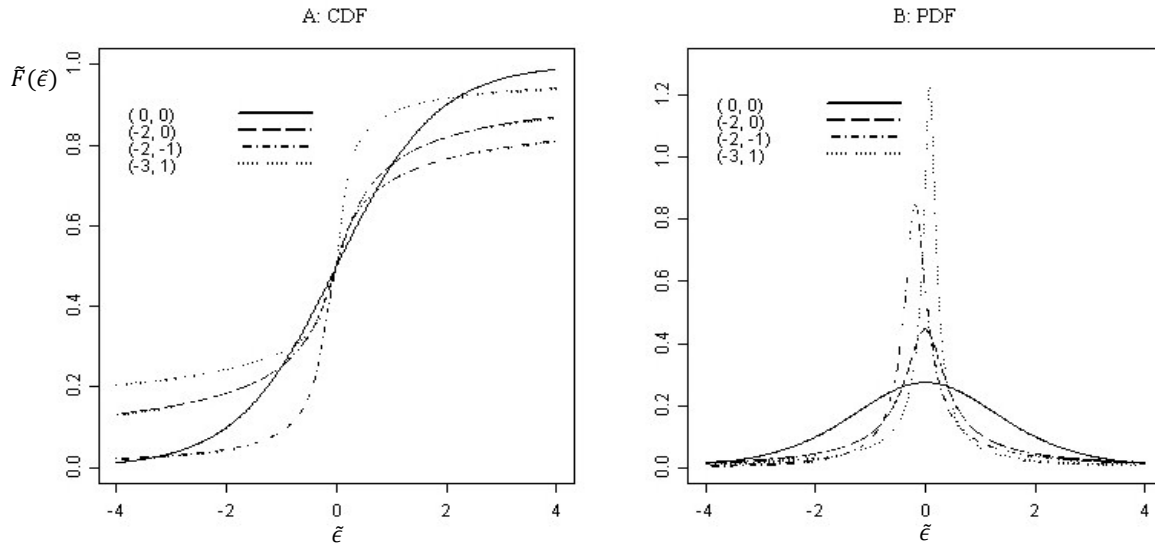
CDF of ϵ determines the shape at the *right* end of the relationship between $X\beta$ and p and that the *right* tail of the CDF determines the shape at the *left* end (equation (3a)).

Five members of this family stand out. First, Pregibon (1980) already highlighted the fact that the GTL distribution becomes the logistic distribution when $\alpha = \delta = 0$. Thus, logit is a special case of pregibit. Second, Table 1 already hints at the fact that there is a member in this

family for which $\kappa_3 = 0$ and $\kappa_4 = 3$, similar to the standard normal distribution: indeed, as Joiner and Rosenblatt (1971), Ramberg et al. (1979), and Freimer et al. (1988) noted, this happens for $(\alpha, \delta) = (0.1349, 0)$. This distribution has finite tails: the density is positive for $-7.41 < \epsilon < 7.41$ with $\sigma_\epsilon = 1.46$, or, in a standardized form, for $-5.06 < \tilde{\epsilon} < 5.06$. Given that these endpoints are finite and that the GTL link function in equation (4) and the derivative in equation (5) leave the density function undefined outside the endpoints, we define the density to equal 0 for all values outside the endpoints. The similarity of the moments expresses that these distributions are similar, but the accuracy of the approximation may also be expressed by two indices: the absolute difference in the CDF, $L(\tilde{F}, \Phi) = \int |\tilde{F}(\tilde{\epsilon}; \alpha, \delta) - \Phi(\tilde{\epsilon})| d\tilde{\epsilon}$, and in the PDF, $L(\tilde{f}, \phi) = \int |\tilde{f}(\tilde{\epsilon}; \alpha, \delta) - \phi(\tilde{\epsilon})| d\tilde{\epsilon}$, where Φ and ϕ refer to the CDF and PDF of the standard normal distribution. We find $L(\tilde{F}, \Phi) = 0.0028$, which means that as Φ rises from 0 to 1, $\tilde{F}$ follows it with a cumulative error of only 0.28 percentage points. Also, $L(\tilde{f}, \phi) = 0.0067$, which implies that the total difference in the area underneath the PDFs (itself equal to 1) equals only 0.0067. Thus, GTL(0.1349,0) closely resembles the standard normal distribution.[11,12] In this way, the pregibit(0.1349,0) model closely approximates the probit model, which therefore permits a test of the probit model against a more general member of the pregibit family that permits skewness and different tail behavior.

Third, when δ is restricted to 0 while α is estimated as a free parameter, skewness κ_3 equals 0. This class of special cases of the pregibit model is therefore referred to as *symmetric pregibit*. Now, as α declines from 0.1349 to -0.25 while holding δ at 0, κ_4 rises from 3 to ∞. Kurtosis of the t distribution varies over the same range: κ_4 of a $t(\nu)$ distribution rises from 3 to ∞ as ν decreases from ∞ to 4, and of course the t distribution is symmetric. Thus, any $t(\nu)$ distribution for $\nu > 4$ may be matched with a GTL$(\alpha, 0)$ distribution with an α in the range of $(-0.25, 0.1349)$. For example, GTL$(-0.1359, 0)$ matches the moments of $t(5)$, and GTL$(0.0571, 0)$ matches the moments of $t(15)$. The absolute difference indices are small, though not as small as for the standard normal distribution: $L(\tilde{F}, \tilde{F}_{t(5)}) = 0.0326$ and $L(\tilde{f}, \tilde{f}_{t(5)}) = 0.0650$—but we will return this shortly. Thus, by close approximation, the pregibit family encompasses the gosset model of Liu (2006) and Koenker and Yoon (2009).[13]

Fourth, let us consider the loglog and cloglog models. Disturbances ϵ that follow an extreme (maximum) value distribution (which may also be referred to as the Gumbel or loglog distribution) yield a cloglog link function and thus underlie the cloglog discrete choice model that permits asymmetrically distributed disturbance. Similarly, if ϵ follows an extreme minimum value (or cloglog, the complement of Gumbel) distribution, the loglog discrete choice model results. These two models are mirror images of each other, with the same kurtosis but opposite skewness. Now, for the Gumbel distribution, $\kappa_3 = 1.14$ and $\kappa_4 = 5.4$, which is matched by GTL$(0.2107, -0.2419)$: pregibit can also closely approximate the loglog and cloglog models.

---

[11] The value of the standard normal density at $\tilde{\epsilon} = 5.06$ equals 0.000001, close to the value of 0 that GTL(0.1349,0) yields. The peaks also are close: 0.3989 for the standard normal and 0.4018 for GTL(0.1349,0).
[12] Skewness and kurtosis also happen to equal 0 and 3 when $(\alpha, \delta) = (5.2, 0)$, but this GTL density is radically different. For example, the endpoints are found at -2.39 and 2.39, the density function does not go down to 0 at these endpoints but to 0.08, and the PDF peaks at $\tilde{\epsilon} = 0$ at a value of 0.7391 rather than 0.3989. The absolute difference indices equal $L(\tilde{F}, \Phi) = 0.1081$ and $L(\tilde{f}, \phi) = 0.3368$.
[13] Matching by moments is not feasible for $\nu \le 4$. We address the approximation of $t(\nu)$ for $\nu \le 4$ later on.

Figure 6: Comparing GTL and Gumbel distributions

A: CDF, distribution matched by moments

B: PDF, distribution matched by moments

C: CDF, distribution matched by $L(\tilde{F}, \tilde{F}_G)$

D: PDF, distribution matched by $L(\tilde{f}, \tilde{f}_G)$
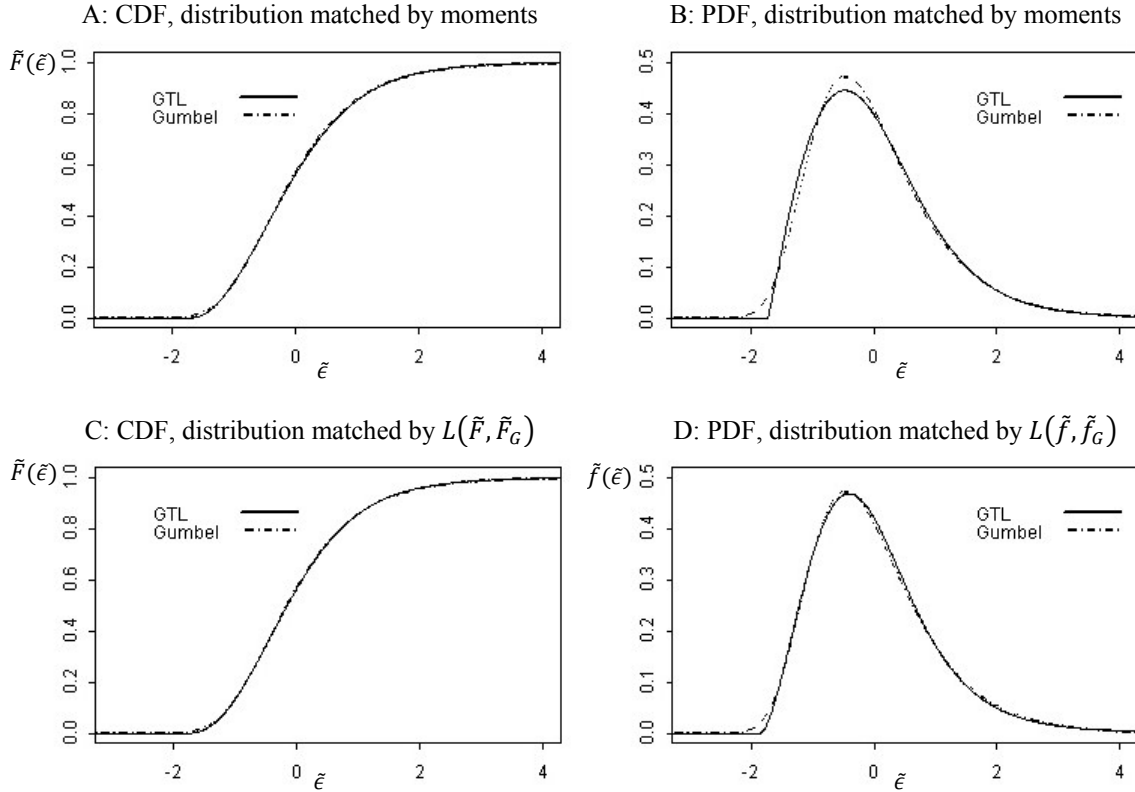
How close is this approximation? Figure 6 visualizes the Gumbel and its GTL match graphically. Panel A and B show the CDF and PDF respectively. In Panel A, a small gap is visible, which is quantified in the absolute difference index $L(\tilde{F}, \tilde{F}_G) = 0.0190$; Panel B shows a substantial discrepancy, quantified with $L(\tilde{f}, \tilde{f}_G) = 0.0571$. This raises the question whether matching by moments is actually an optimal strategy. In Panel C, the CDFs are depicted when (α, δ) are matched by minimizing $L(\tilde{F}, \tilde{F}_G)$ instead: the minimum equals 0.0120 at $(\alpha, \delta) = (0.1671, -0.2334)$, and the gap is visibly smaller indeed. In Panel D, the distribution is matched by minimizing $L(\tilde{f}, \tilde{f}_G)$; at the minimum of 0.0307, we have $(\alpha, \delta) = (0.1422, -0.2290)$, with a remarkable improvement compared to Panel B.

The same kind of improvement is possible for the matching with the standard normal and the $t(v)$ distributions. Table 2 provides the numbers: the L-indices can be cut in half with a better choice of (α, δ).

Then, should distributions be matched by their CDF or their PDF? On the one hand, we are interested in the relationship between the probability of success and its determinants, which favors matching by the CDF. On the other hand, the discrete choice framework starts with positing a distribution for the disturbance, which favors matching by the PDF. This dilemma is resolved by a Monte Carlo study with data of one of the six applications that will be used to illustrate the pregibit model in Section 4; a second Monte Carlo study with another dataset

yielded the same results.[14] Let us consider an analysis of the probit match. As a first step, we determine β by estimating a probit model on the full dataset. Then, for 100 runs, we select 2000 observations at random, compute $X'\beta$, add a standard normal draw of $\tilde{\epsilon}$, determine y, and then estimate the probit and the GTL-matched pregibit models. Across the 100 runs, Table 2 reports the average, minimum (largest negative) and maximum (largest positive) difference in the log-likelihood value, as well as the average of the largest difference in the estimated probability of success among the 2000 observations. Both the likelihood function and the estimated probabilities of success are closer when the pregibit model is based on a GTL that is matched by the PDFs. The same is true for the $t(5)$ and cloglog models.[15] Thus, in the following, we will consider the GTL(0.1436,0) to match the probit model, GTL(0.1422,−0.2290) to match the cloglog model, and GTL(0.1422,0.2290) to match the loglog model.

Lastly, the pregibit model yields one more special case that warrants our interest.[16] Figure 4A.2 already shows that the GTL distribution becomes a uniform distribution when $\alpha = 1$ and $\delta = 0$. Indeed, the link function may be written as $\epsilon = 2q - 1$, implying $f(\epsilon) = 0.5$ for $-1 \leq \epsilon \leq 1$ and $\sigma_\epsilon^2 = 1/3$. In standardized form, we have

$$
\begin{aligned}
p_i &= 0 && \text{for} \quad X_i'\tilde{\beta} < -\sigma_\epsilon^{-1} \\
&= 0.5\,(\sigma_\epsilon X_i'\tilde{\beta} + 1) && \text{for} \quad -\sigma_\epsilon^{-1} \leq X_i'\tilde{\beta} \leq \sigma_\epsilon^{-1} \qquad (8) \\
&= 1 && \text{for} \quad X_i'\tilde{\beta} > \sigma_\epsilon^{-1}
\end{aligned}
$$

Now, in the familiar linear probability model that is estimated with OLS, y is regressed against X. One characteristic of that model is that $E[y_i] = P[y_i = 1] = X_i'\beta^*$, from which the model gets its name: the probability of success is a linear function of $X'\beta^*$. One of the complaints about that model is that predicted probabilities may be less than 0 or greater than 1. Through equation (8) based on the GTL(1,0) distribution, we arrive at a model where, within feasible limits, p is a linear function of $X'\beta$ and yet the structure of the model is consistent with the common framework of dichotomous choice models. In this way, the linear probability model is approximately nested within the pregibit family and may be tested against more general members of this family.

Unfortunately, the GTL distribution turns into a uniform distribution not only when $(\alpha, \delta) = (1,0)$ but also when $(\alpha, \delta) = (2,0)$ and when $(\alpha, \delta) = (\alpha, -\alpha + 1)$ for $\alpha \to \infty$ and when $(\alpha, \delta) = (\alpha, \alpha - 1)$ for $\alpha \to \infty$ (Freimer et al., 1988, p.3550).[17] Thus, the log-likelihood function of the pregibit model takes on the same value for these four parameter pairs. This has two consequences. First, this creates an identification issue. Even so, GTL density functions for values of (α, δ) in the neighborhood of these four points are distinctly different (Appendix B),

---

[14] The data used for Table 2 pertain to Application 4 (post-secondary enrollment) in Section 4, selected because successes and failures are roughly balanced. A similar Monte Carlo analysis with data from Application 5 (mortgage denial) yields similar results, even though the sample proportion of successes is only 12 percent.

[15] With $\alpha = -0.8416$, even $t(1)$ is matched with high accuracy ($L(f_t, f_G) = 0.0106$), but since $t(v)$ cannot be standardized for $v \leq 2$, a scaling factor must be computed numerically along with α to help match the PDFs.

[16] Ramberg et al. (1979) mention that, with suitably selected (α, δ), the GTL distribution can also closely approximate the lognormal, gamma, and Weibull distributions. It is not likely that these distributions (shifted so as to have a mean of 0) have ever been used within a discrete choice framework, but pregibit is a comprehensive model indeed.

[17] The latter conditions can be restated as $(\alpha - \delta, \alpha + \delta) = (\infty, 1)$ and $(\alpha - \delta, \alpha + \delta) = (1, \infty)$.

which suggests that the identification issue is limited to only these four points. Second, the configuration of these four points makes it likely that the loglikelihood function is poorly behaved for values of (α, δ) with $\alpha \geq 1$ and $-\alpha + 1 \leq \delta \leq \alpha - 1$, which in Figure 1 corresponds to an area to the right of a kinked line that intersects the horizontal axis at $\alpha = 1$. Local maxima may occur. In our applications, we indeed found some evidence of this.

2.4. Comparing members of the pregibit family

The pregibit family has many members that are popular in the empirical literature but are special cases of the more general pregibit model. The parameter restrictions that these popular models impose ought to be tested. Four tests are available. First, a Wald test compares $(\hat{\alpha}, \hat{\delta})$ of the unrestricted model with the implied (α_0, δ_0) of the restricted model. Second, a likelihood ratio test compares the log-likelihood values of the restricted and unrestricted models, which therefore must both be estimated, making this test computationally intensive. The third and fourth tests, the Lagrange Multiplier (LM) and Goodness of Link (GL) tests, require estimation of only the restricted model. These two tests are related and are now discussed in more detail.

Consider the model in standardized form:[18] the CDF $\tilde{F}(X_i'\tilde{\beta}; \alpha, \delta) \equiv p_i$ is derived from the expression $\mu_\epsilon - \sigma_\epsilon X_i'\tilde{\beta} = G(1 - p_i; \alpha, \delta)$, where $\mu_\epsilon = \mu(\alpha, \delta)$ and $\sigma_\epsilon = \sigma(\alpha, \delta)$. Define ζ to be the vector that combines all parameters of the pregibit model: $\zeta = (\tilde{\beta}' \ \alpha \ \delta)'$. Define $d_i = \partial F / \partial \zeta$. Specifically,

$$d_i = \begin{pmatrix} \dfrac{-\sigma_\epsilon X_i}{G_{qi}} \\ \dfrac{1}{G_{qi}}\left(-G_{\alpha i} + \dfrac{\partial \mu_\epsilon}{\partial \alpha} - X_i'\tilde{\beta}\dfrac{\partial \sigma_\epsilon}{\partial \alpha}\right) \\ \dfrac{1}{G_{qi}}\left(-G_{\delta i} + \dfrac{\partial \mu_\epsilon}{\partial \delta} - X_i'\tilde{\beta}\dfrac{\partial \sigma_\epsilon}{\partial \delta}\right) \end{pmatrix}, \tag{9}$$

where $q = 1 - p$ as the first argument of the function G, and where subscripts of G denote derivatives with respect to the indicated argument.[19]

For the LM test, the first order derivative of the likelihood function is

$$\frac{\partial \ln L}{\partial \zeta} = \sum_{i=1}^{n} \frac{1 - y_i - \tilde{F}_i}{\tilde{F}_i(1 - \tilde{F}_i)} d_i \tag{10}$$

and the expected value of the second order derivative is[20]

$$E\frac{\partial^2 \ln L}{\partial \zeta \partial \zeta'} = -\sum_{i=1}^{n} \frac{1}{\tilde{F}_i(1 - \tilde{F}_i)} d_i d_i' \tag{11}$$

---

[18] The LM and GL tests are shown for the standardized pregibit model. For tests on the unstandardized model, just set μ_ϵ equal to 0 and σ_ϵ equal to 1. When iqr-standardization is applied, replace μ_ϵ with m_ϵ and σ_ϵ with s_ϵ.

[19] For an arbitrary (α, δ), we have $G_\alpha = \frac{1}{\alpha-\delta}\left(q^{\alpha-\delta}\ln q - \frac{q^{\alpha-\delta}-1}{\alpha-\delta}\right) - \frac{1}{\alpha+\delta}\left((1-q)^{\alpha+\delta}\ln(1-q) - \frac{(1-q)^{\alpha+\delta}-1}{\alpha+\delta}\right)$. For $\alpha - \delta = 0$, the first term of this expression equals $\frac{1}{2}(\ln q)^2$ by L'Hôpital's Rule. Similarly, for $\alpha + \delta = 0$, the first term of this expression equals $\frac{1}{2}(\ln(1-q))^2$. We also have $G_\delta = -\frac{1}{\alpha-\delta}\left(q^{\alpha-\delta}\ln q - \frac{q^{\alpha-\delta}-1}{\alpha-\delta}\right) - \frac{1}{\alpha+\delta}\left((1-q)^{\alpha+\delta}\ln(1-q) - \frac{(1-q)^{\alpha+\delta}-1}{\alpha+\delta}\right)$. G_q is more straightforward: $G_q = q^{\alpha-\delta-1} + (1-q)^{\alpha+\delta-1}$.

[20] Since $E[y_i] = 1 - \tilde{F}_i$, many terms drop out of the expression for the second order derivatives when the expectation is taken.

The LM test statistic is then computed in the familiar way: $LM = \frac{\partial \ln L}{\partial \zeta'} \left(-E \frac{\partial^2 \ln L}{\partial \zeta \partial \zeta'} \right)^{-1} \frac{\partial \ln L}{\partial \zeta}$. The expressions are evaluated at the estimate of $\hat{\zeta}' = \left(\hat{\beta}', \alpha_0, \delta_0 \right)$ of the restricted pregibit model (e.g., probit) that is to be tested. LM is distributed asymptotically as $\chi^2(2)$ for every restricted model.[21] To compute LM, the restricted model is estimated only once.

As for the GL test, Pregibon (1980) proposed a goodness of link test as a way to test for the logistic specialty case of the generalized Tukey lambda link function in the context of *grouped* date. Within the context of individual data and discrete choice models other than logit, this test is straightforwardly to implement as well. Rewrite the link function in standardized form as[22]

$$X'\tilde{\beta} = -\frac{G(q;\alpha,\delta) - \mu_\epsilon}{\sigma_\epsilon} = -\tilde{G}(q;\alpha,\delta) \tag{12}$$

As before, let α_0 and δ_0 be the hypothesized population parameters. The link function in equation (12) with correctly specified parameters $\alpha \neq \alpha_0$ and $\delta \neq \delta_0$ may be approximated with a first order Taylor expansion around (α_0, δ_0):

$$\tilde{G}(q;\alpha,\delta) = \tilde{G}(q;\alpha_0,\delta_0) + \tilde{G}_{\alpha 0}(\alpha - \alpha_0) + \tilde{G}_{\delta 0}(\delta - \delta_0) \tag{13}$$

where the derivatives $\tilde{G}_{\alpha 0} = \partial \tilde{G}(q;\alpha_0,\delta_0)/\partial \alpha$ and $\tilde{G}_{\delta 0} = \partial \tilde{G}(q;\alpha_0,\delta_0)/\partial \delta$ are both functions of q, α_0 and δ_0. Inserting (13) into (12) and rearranging terms yields

$$X'\tilde{\beta} + \tilde{G}_{\alpha 0}(\alpha - \alpha_0) + \tilde{G}_{\delta 0}(\delta - \delta_0) = -\tilde{G}(q;\alpha_0,\delta_0) \tag{12}$$

In other words, the deviation between the correct and the hypothesized link function will show up through significant parameter estimates on the auxiliary variables $\tilde{G}_{\alpha 0}$ and $\tilde{G}_{\delta 0}$ that are added to the hypothesized model along with the regular explanatory variables X. The GL test is therefore a Wald test on the parameters of $\tilde{G}_{\alpha 0}$ and $\tilde{G}_{\delta 0}$, distributed asymptotically as $\chi^2(2)$.[23] Note that the GL test requires estimation of two restricted pregibit models since q in $\tilde{G}_{\alpha 0}$ and $\tilde{G}_{\delta 0}$ depends on $(\beta' \ \alpha_0 \ \delta_0)$.

As discussed in Section 2.2, for some values of (α, δ), the link function links q to $X'\tilde{\beta}$ only over a finite range of $X'\tilde{\beta}$. In other words, $\tilde{G}$ becomes vertical for $q = 0$ and $q = 1$. For such values of q, α and δ have no impact on the relationship between q and $X'\tilde{\beta}$. Accordingly, we define $\tilde{G}_\alpha(0;\alpha_0,\delta_0) = \tilde{G}_\alpha(1;\alpha_0,\delta_0) = 0$ and similarly $\tilde{G}_\delta(0;\alpha_0,\delta_0) = \tilde{G}_\delta(1;\alpha_0,\delta_0) = 0$.

Apart from these formal statistical tests, a less formal comparison of pregibit models may be desirable as well. In the case that (α, δ) is such that standardization by μ_ϵ and σ_ϵ is feasible, we have $Var(y^*) = Var(X'\beta) + 1$ since ϵ and X are independent by assumption. Thus, the fit of the model may be expressed by a pseudo-R^2 measure equal to $R^2_{var} = Var\left(X'\hat{\tilde{\beta}}\right)/\left(Var\left(X'\hat{\tilde{\beta}}\right) + 1\right)$, which parallels the traditional R^2 statistics of linear models estimated by ordinary least squares (McKelvey and Zavoina, 1975; Laitila, 1993;

---

[21] For the symmetric pregibit model, which only restricts δ, evaluate the expressions at $\hat{\zeta}' = \left(\hat{\beta}', \hat{\alpha}, \delta_0 \right)$ with $\delta_0 = 0$; the LM test is distributed as $\chi^2(1)$.

[22] Note that μ_ϵ and σ_ϵ are functions of α and δ as well, as highlighted in the derivatives in equation (9).

[23] For a goodness of link test on the symmetric pregibit model, only $\tilde{G}_{\delta 0} = \tilde{G}_\delta(q;\hat{\alpha},0)$ is added to the set of regressors. The asymptotic distribution of GL is $\chi^2(1)$.

Windmeier, 1995). In this similarity, it has an advantage over other measures of fit that describe the match of the probability of success to a successful outcome. In other words, R^2_{var} measures what the explanatory factors in the data say about the variation of y^* rather than y. One example where this matters is in the measurement of the willingness to pay for a particular service in the context of a discrete outcome: this relies on determining the variation in y^*. Note, however, that computation of R^2_{var} is not feasible when σ_ϵ is not defined.

3. ASYMPTOTIC BIAS IN PROBIT AND LOGIT UNDER PREGIBIT

The gain offered by the pregibit model lies in the accurate quantification of the relationship between explanatory variables X and the dichotomous variable Y. This relationship is expressed as $P[Y = 1|X] = 1 - \tilde{F}(-X'\tilde{\beta}; \theta)$ if standardization of the disturbance term of the model is feasible, or it is expressed through the marginal impact formula as

$$\frac{dP[Y=1|X]}{dX} = \tilde{f}(-X'\tilde{\beta})\tilde{\beta} \tag{19}$$

where $\tilde{f}$ is the standardized version of the GTL density. If an erroneous distribution is assigned to ϵ, the discrete choice model suffers from a specification error. In this section, we examine the asymptotic bias in $\tilde{\beta}$ and in the marginal effect when $\tilde{f}$ is misspecified as either standard normal or standardized logistic.

Let us consider of values of (α, δ) that permit standardization of the GTL distribution and therefore are not in any way "extreme." α ranges from -0.4 to 1, corresponding to distributions with infinite tails ($\alpha \leq 0$) and with shorter finite tails ($\alpha > 0$). For $\alpha = -0.2$, δ varies from -0.27 to 0.27, and for $\alpha = 0.5$, δ varies from -0.5 to 0.5. We employ a simple model that contains only one explanatory variable. The data generating process uses $\tilde{\beta}_0 = 1$ and $\tilde{\beta}_1 = 1$ with the GTL-distributed disturbance term for several combinations of α and δ. We consider three scenarios that differ in the likelihood of success; more particularly, for each scenario we select the range of the explanatory variable $[X_L, X_U]$ such that, for the given $\tilde{\beta}$, α and δ, $P[Y = 1|X]$ falls within a specified interval. The intervals that we select are $(0.01, 0.99)$, $(0.5, 0.99)$ and $(0.333, 0.667)$, such that the data of the first probability interval cover two tails; the second covers one tail; and the third observes neither tail. There are 500 equally spaced X values within each range.

Table 3 summarizes logit and probit estimators of $\tilde{\beta}_1$ and the difference between the estimated marginal logit and probit impacts and the true pregibit impact, expressed as a ratio for ease of comparison; the table reports the lowest value of this ratio, the median value, and the highest value. As can be seen, the asymptotic bias in the logit and probit estimators of $\tilde{\beta}_1$ differs according to the data range and the parameter values of (α, δ). When data are characterized with a higher probability in success, i.e., $P[Y = 1|X]$ ranges between $(0.5, 0.99)$, the estimate of $\tilde{\beta}_1$ at $(\alpha = -0.2, \ \delta = -0.27)$ is extremely large, e.g., 4.716 or 4.660, even though the difference of the estimated marginal logit and probit impacts and the true pregibit impact is not large. But once again, the relationship between X and Y is more subtle, depending on the CDF of the assigned distribution. Thus, depending on the values of α and δ and the data range, the marginal impact may be underestimated by as much as 97 percent or overestimated by more than 270 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 40 percent.

4. APPLICATIONS

A model that is more elaborate than its popular predecessors adds value only if it improves research outcomes. Thus, we have looked for examples in the literature for which data have been made publicly available. Table 4 gives an overview of the results; detailed estimates are provided in Appendix C.[24] The first four lines of the table describe each dataset in the simplest terms: the size of the sample and the range of the probability of a success ($y = 1$) according to the probit model. Rather than getting carried away by potential outliers, we characterize the probability range by the 1st and 99th percentile of the estimated probability values, and the mean probability corresponds closely to the proportion of the sample reporting a success. As stated before, the pregibit model distinguishes itself from simpler models such as logit or probit in the behavior of the probability of success in the tails and, possibly, in the variation in the slope of the CDF over common middle ranges of the probability of success. Research applications with large samples where the probability of success ranges from nearly nil to nearly certain hold the greatest promise of evidence in support of pregibit. Thus, the application in the sixth column is actually not particularly promising: even though the sample is quite large, the probit probability of success ranges only from 47 to 72 percent. The third and fifth applications lack evidence in the lower and upper tails, respectively. The first and second applications use smaller samples that may prevent an unambiguous diagnosis of pregibit behavior.

In the bottom portion of the table, results of the LR, SC and LM tests are reported; to conserve on space, the Wald test is omitted. A quick glance suggests that the popular models are adequate for some applications and must be rejected in other cases. Let us now consider each application in turn for evidence in regard to the pregibit model.

The first application is drawn from Bliss (1935) and was used by Prentice (1976) and Pregibon (1980) in their analysis of discrete choice models. The data describe mortality among beetles when they are exposed to varying dosages of a pesticide. Pregibon illustrated with a graph that his model, which is referred to as the unstandardized pregibit model in this paper, performs better than logit, but he reported neither the estimates of α and δ nor any of the test statistics—and he did not consider any other model as a special case of pregibit. The pregibit model is preferred over any of the symmetric models: the underlying distribution is decidedly asymmetric.[25] However, cloglog is an acceptable model as well.

The labor force participation example uses well-known data of Mroz (1987), used in textbooks by Berndt (1991) and Wooldridge (2002). The estimated pregibit model finds $\hat{\alpha} = 0.500$ and $\hat{\delta} = -0.057$. The distribution of the disturbance term is therefore nearly symmetric ($\kappa_3 = 0.12$) but less peaked ($\kappa_4 = 2.12$) than probit or logit. These estimates imply that standardized disturbance ranges from -1.99 to 2.37. However, the difference with either

---

[24] In addition to the six illustrations reported in Table 4, we applied the pregibit model to a seventh dataset that measured the incidence of prostate cancer spreading to lymph nodes on the basis of indicators that are more easily obtainable than a biopsy of the lymph nodes themselves (data from Brown (1980), made available through Davison and Hinkley (1997)). With only 53 observations, there is no clear preference for any model. The iterative search to maximize the likelihood function of the pregibit model failed to formally converge but indicate no improvement over logit or probit.

[25] Skewness κ_3 cannot be computed as $(\hat{\alpha}, \hat{\delta})$ falls left of the kinked line in Figure 1 that intersects the horizontal axis at 0.333.

logit or probit is small and statistically insignificant, and the difference with the linear probability model is smaller yet.

The third illustration is drawn from Vijverberg and Zeager (1994) concerning sector choice in Tanzania. The model estimated here is a simplified version of the one published originally, permitting a larger sample size. With $\hat{\alpha} = 1.472$ and $\hat{\delta} = 0.682$, the distribution of the disturbance term is distinctly left-skewed and rather "flat" ($\kappa_3 = -0.36$, $\kappa_4 = 1.97$), and the range of the standardized disturbance equals $[-2.12, 1.46]$. Figure 7 depicts the estimated CDF and PDF of the disturbances of the probit and pregibit models: although the CDFs appear to overlap, the densities are clearly very different. However, the general pregibit model is not statistically preferred over probit or any of the popular special cases (with the possible exception of the loglog model).

Figure 7: Difference between pregibit and probit: Tanzanian data on sector choice
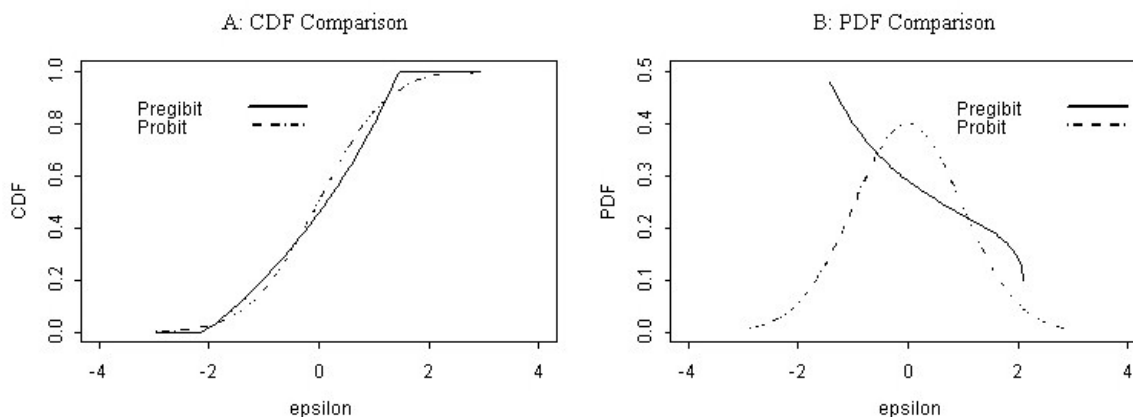
The fourth application employs data gathered by Kelchtermans and Verboven (2010) for a study of the demand for post-secondary education among high school students in Flanders, Belgium. Their nested multinomial logit model includes every major at every institution of higher education as possible outcomes as well as the option of not pursuing a post-secondary degree. We simplify the model by aggregating all higher education options into a single aggregate outcome that is contrasted with the option of not pursuing a post-secondary degree. We condense one of their datasets, which contains a still-large subsample of students with all of their options. In addition to personal characteristics and the nature of the high school curriculum, we include both the minimum and the mean cost of travel to the various institutions as determinants of the choice to pursue a post-secondary degree; both variables matter in the discrete choice models we estimate (see Appendix C). The pregibit model resoundingly rejects the logit, probit and linear probability models (Table 3), opting instead for an unstandardized model with ($\hat{\alpha} = -1.467$, $\hat{\delta} = -0.016$) with a distribution that is nearly symmetric and has long tails.

Figure 8 shows the implications of the distributional shift: whereas the probit model detects some types of individuals who appear to be certain to pursue a higher education (probability near 1) and others who apparently are certain not to continue their education (probability near 0), the pregibit model tempers such predictions. Panel A shows this through a sample scatterplot that contrasts the probit and pregibit probabilities; Panel B illustrates this by

Figure 8: Difference between pregibit and probit: enrollment in post-secondary education in Belgium
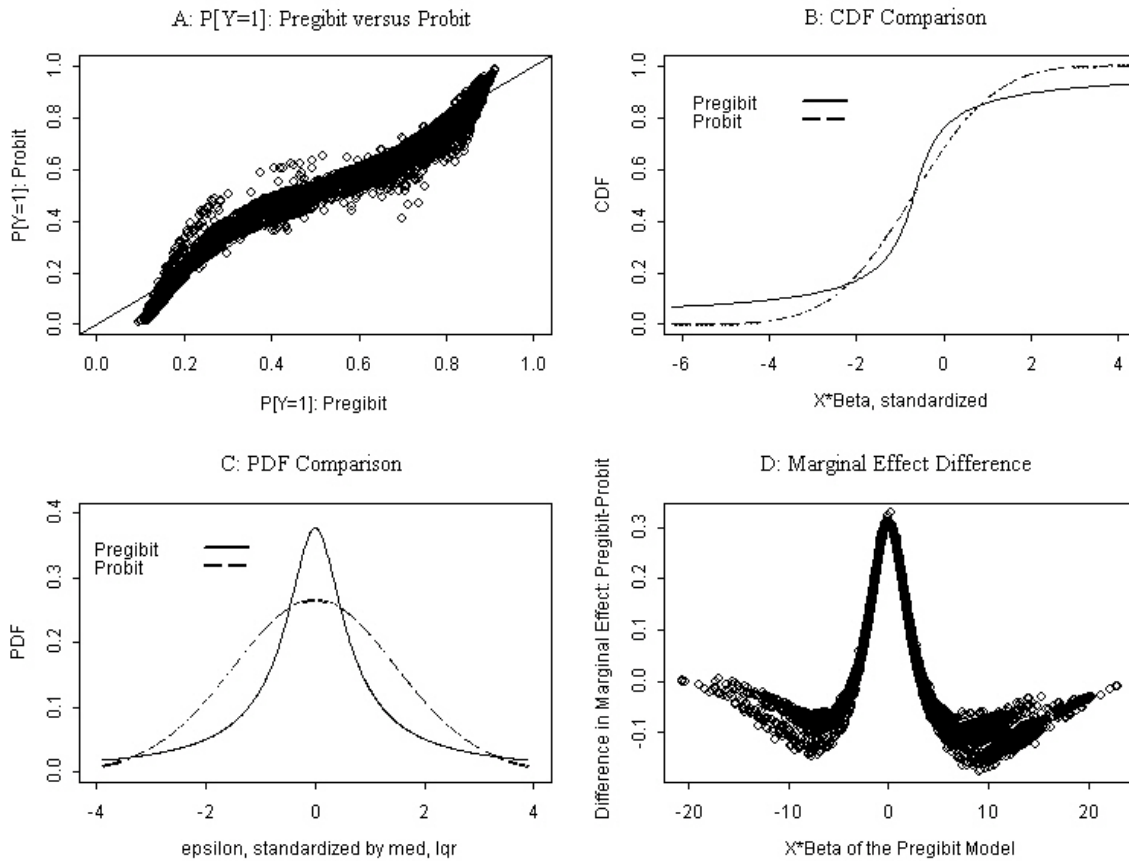
both CDFs as functions of the respective standardized values of $X'\hat{\beta}$.[26]   Over the middle probability range, a 1-standard deviation change in $X'\hat{\beta}$ produces a greater impact in the pregibit model, but closer to the tails, the impact is more muted. Panel C compares the PDFs of the disturbances of the two models again, scaled by the median as the measure of centrality and one half of the interquartile range as the measure of spread since the mean and variance of the disturbance are not defined (see Section 2.4). The higher peak and the thicker tails of the pregibit distribution are offset by lower density values at intermediate-sized disturbances.

Thus, the effect of $X'\hat{\beta}$ on the probability of success consists of three phases: starting from the mean, an increase in $X'\hat{\beta}$ initially induces large numbers of students to seek a higher education; a subsequent increase in $X'\hat{\beta}$ adds smaller numbers according to the pregibit than according to the probit model; and a continued increase still finds students to induce to higher education under pregibit, whereas probit asserts that there are hardly any students left to induce. Decreases in $X\hat{\beta}$ from the mean encounter the same three phases. Thus, the thick tails of the distribution of ϵ indicate that, regardless of the deterministic component of the choice model,

---

[26] The values of $X\hat{\beta}$ differ greatly between the two models, as is evident from the range of pregibit's $X\hat{\beta}$ in panel D. The standardization rescales $X\hat{\beta}$ in the traditional way such that, for both models, its mean is 0 and its standard deviation equals 1. In this way, the CDFs are comparable.

success (or failure) will never be virtually certain. All of this implies that the marginal effect under the probit model must differ substantially from that under the pregibit model. Panel D expresses this for the Belgian sample of high school students in the form of the difference in the marginal effect, arranged by the value of pregibit's $X'\hat{\beta}$.[27]

The fifth application concerns the outcome of mortgage applications (Munnell et al., 1996; see also Stock and Watson, 2007). In our model, success is defined as a mortgage application for which funding was denied. We estimate the same specification as Munnell et al.'s baseline model, with three differences. First, we are not able to include fixed effects for census tract and for lender identity as these variables are not available in the publicly available data. Second, we omit a dummy indicating whether the applicant's application for mortgage insurance had been denied. We consider this dummy variable to be potentially endogenous: officers evaluating mortgage insurance applications look at similar information as those evaluating mortgage applications, such that this dummy variable may well be correlated with the disturbances that drive the mortgage application outcome.[28] Third, we do not pool African-American and Hispanic applicants into a single minority dummy variable.

The (unstandardized) pregibit model is strongly preferred over all symmetric models, with $\hat{\alpha} = -1.626$ and $\hat{\delta} = 0.919$: the disturbances are distributed with a left skew and long tails. Even the asymmetric models (loglog and cloglog) fail because while they are able to capture skewness, they cannot generate long tails.

Figure 9 examines the implications of the difference between probit and pregibit. Panel A contrasts the pregibit and probit probabilities of denial for members of the sample; Panel B compares the probabilities of denial by overlaying the CDFs of the two models such that $X\hat{\beta}$ is standardized to have mean 0 and variance 1. In Panel A, for the large group of observations with a pregibit probability ranging of 0.0 to 0.2, which is around the average denial rate of 0.14, the probit probability is more likely to be higher than the pregibit probability. In other words, according to the probit model, these applicants are more at risk of being denied their mortgage request than the preferred pregibit model indicates. On the other hand, when the pregibit probability is between 0.25 and 0.65, the risk of being denied is actually higher relative to the probit model, by 5 percentage points on average and by more than 8 percentage points for one fourth of the applicants in this range. Finally, for 1.3 percent of the sample, the pregibit risk of being denied exceeds 70 percent but none ever exceeds 0.87. For this group, the probit model is on average 9 percentage points more definite that the mortgage is denied than the pregibit model and indicates virtual certainty of denial for a few persons. The more clinical comparison in Panel B highlights the difference in the relationship between determinants and the probability of denial, especially in the upper tail.

---

[27] For any continuous variable, the marginal effect may be computed with equation (19). Since marginal effects are difficult to fruitfully compare when there are many continuous and discrete variable, we employ a more aggregative measure instead, both here and in the discussion of other applications below. Since X is the same for the two models that we compare, the standard deviation of $X\hat{\beta}$ (denoted as $\sigma_{X\hat{\beta}}$) measures the scale of $\hat{\beta}$ for a given model. Thus, we compare marginal effects across models on the basis of $f(-X_i\hat{\beta})\sigma_{X\hat{\beta}}$.

[28] Indeed, 93 percent of the applicants who failed to obtain mortgage insurance had their mortgage application denied as compared with about 12 percent in the overall sample.

Figure 9: Differences between pregibit and probit: denial of mortgage applications in the US
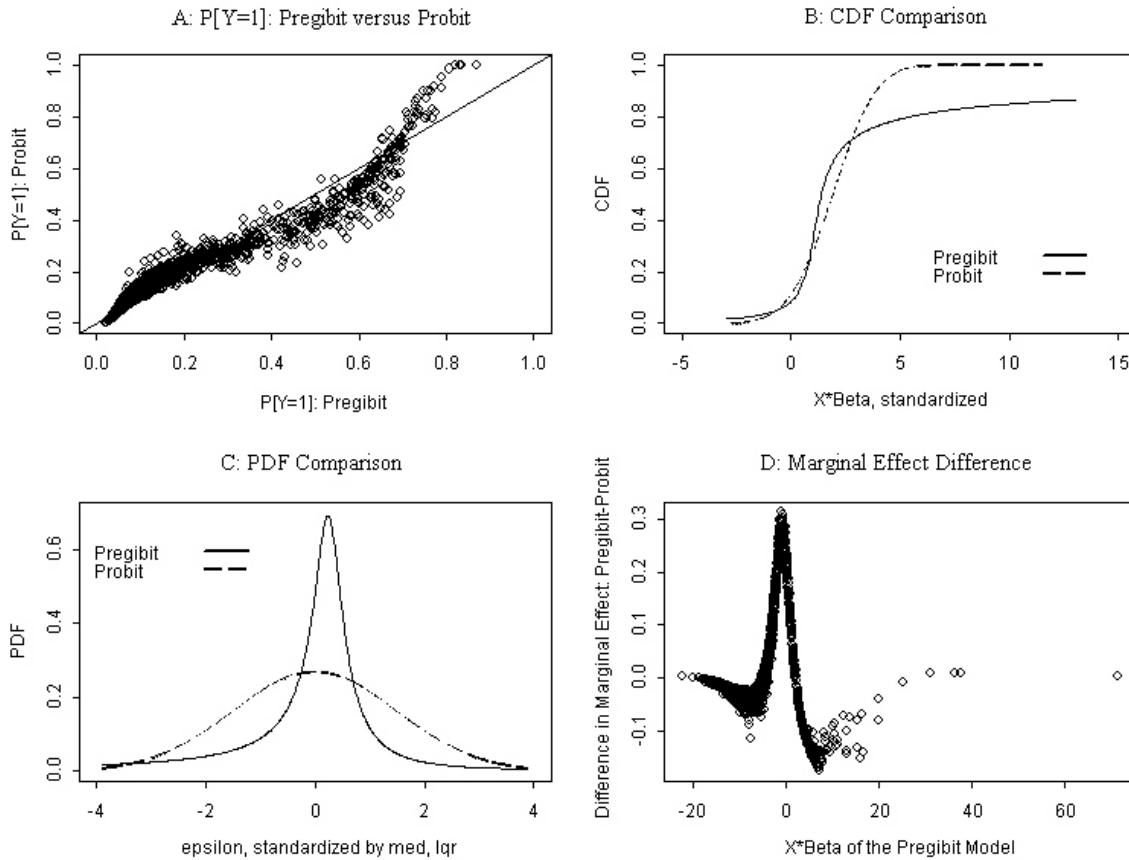
The difference in the disturbance distribution is even starker when viewed by their PDFs (Panel C). The densities are both scaled by the median and half-IQR. The GTL density is sharply peaked and has much longer tails: the tails that Panel C cuts off of the normal density contain only 0.45 percent probability each, but the long tails of the GTL density hold 18.2 percent on the left and 1.4 percent on the right. Thus, the left tail is noticeably thicker than the right tail: GTL density is also badly skewed to the left.

For the evaluation of mortgage applications, this means the following. Recall that the index function is stated as $y^* = X'\beta + \epsilon$ and that the mortgage is denied whenever $y^* \geq 0$: a negative ϵ helps to get the application accepted. Thus, holding observable factors constant, the array of unobservables that cause the mortgage to be accepted is larger than the set of unobservable factors that cause the mortgage to be denied. In other words, mortgage officers appear to be more inclined to uncover reasons to make a deal than to send clients away empty-handed.

The sharply different distributions imply different marginal effects as well. According to the pregibit model, variations in $X'\hat{\beta}$ matter much more near the center of the distribution than out in the tail.

The sixth and last illustration derives from data used by Riphahn, Wambach and Million (2003); see also Greene (2012, Ch.17). The study examines determinants of doctor's visits in

Germany. The database is a panel of households; to simplify the example, we define the dependent variable as a 0/1 indicator of whether the household visited the doctor's office during the survey year (rather than the number of visits it made), and to avoid an obvious violation of the iid assumption of the discrete choice models studied here, we retain households in the first year of the panel only.

This application is not as successful as the previous five. The iterative search for the maximum of the log-likelihood function takes off to large negative estimates of $\hat{\alpha}$ and $\hat{\delta}$; in the unstandardized version of the model, $\hat{\beta}$ becomes exceedingly large, whereas when the model is standardized with the median and IQR, $\tilde{\beta}$ becomes very small. As the search progresses, the link function $G(q)$ produces values in the order of 10^{20}, at which point one must worry about numerical overflow and resulting numerical inaccuracies that diminish the validity of the estimation outcome. Thus, we stopped the search at a point where the obtained parameters, though not optimized, are already able to illustrate implications of the pregibit model for these data. Table 4 reports estimates $\hat{\alpha} = -10.037$ and $\hat{\delta} = -4.760$: for these parameters, the mean and standard deviation of the GTL distribution are not defined, and the model is estimated with the standardization on basis of the median and the IQR. The range of the unstandardized disturbances is $(-\infty, \infty)$. All restricted pregibit models have significantly lower likelihood values.[29] But are the results plausible? Figure 10.A plots probit probabilities of visiting a doctor against pregibit probabilities: the probit model underpredicts the probability of a doctor's visit both for households that are least likely to visit a doctor and for households that are most likely to visit a doctor: the 1-to-99 percentile range of probabilities changes from (0.47,0.73) for probit to (0.53,0.76) for pregibit, while maintaining the same sample-averaged probability of 0.59. The scaled CDFs in Panel B highlight the divergence of the two models: in the pregibit model, a change in $X\beta$ has little impact at the lower end but makes a large difference in the upper middle half. The scatterplot of the marginal effect (panel D) indicates this succinctly. Panel C focuses our attention on the distribution that is driving these results. Scaled by median and IQR, the density is so sharply peaked that it can be shown together with the normal density only if the vertical axis is in logarithmic scale. Both GTL tails are long but especially the right tail is heavy: the distribution is strongly right-skewed. The range of the standardized $X\beta$ shown in panel D is tiny compared to the tails that are illustrated in panel C, but variations in $X\beta$ do matter much near the narrow peak of the distribution.

This application exemplifies the caution for the use of the pregibit model. Among the six applications that we have considered, only this sixth application does not allow measurement of tail events: there are no households for which, on the basis of observable characteristics, visits to the doctor are nearly certain or exceedingly rare. The pregibit model is less well suited for such an application.[30]

---

[29] The log-likelihood function actually becomes quite flat: for $\hat{\alpha} = -18$ and $\hat{\delta} = -8$, $\ln L$ rises by less than 2.

[30] One might argue that the shape of the estimated GTL distribution indicates model misspecification in that the true model might be a mixture of (*i*) an exogenous common risk that a given household must visit the doctor's office in a given year but is unrelated to household resources, and (*ii*) a discretionary decision to see a doctor for less urgent needs that is related to the explanatory variables included in our model. The fat right tail of the GTL density (panel C of Figure 9) sets a high floor on the probability of a doctor's visit (panel B), but ultimately the GTL distribution is not designed to mimic such a mixture.

Figure 10: between pregibit and probit: doctor's visits in Germany
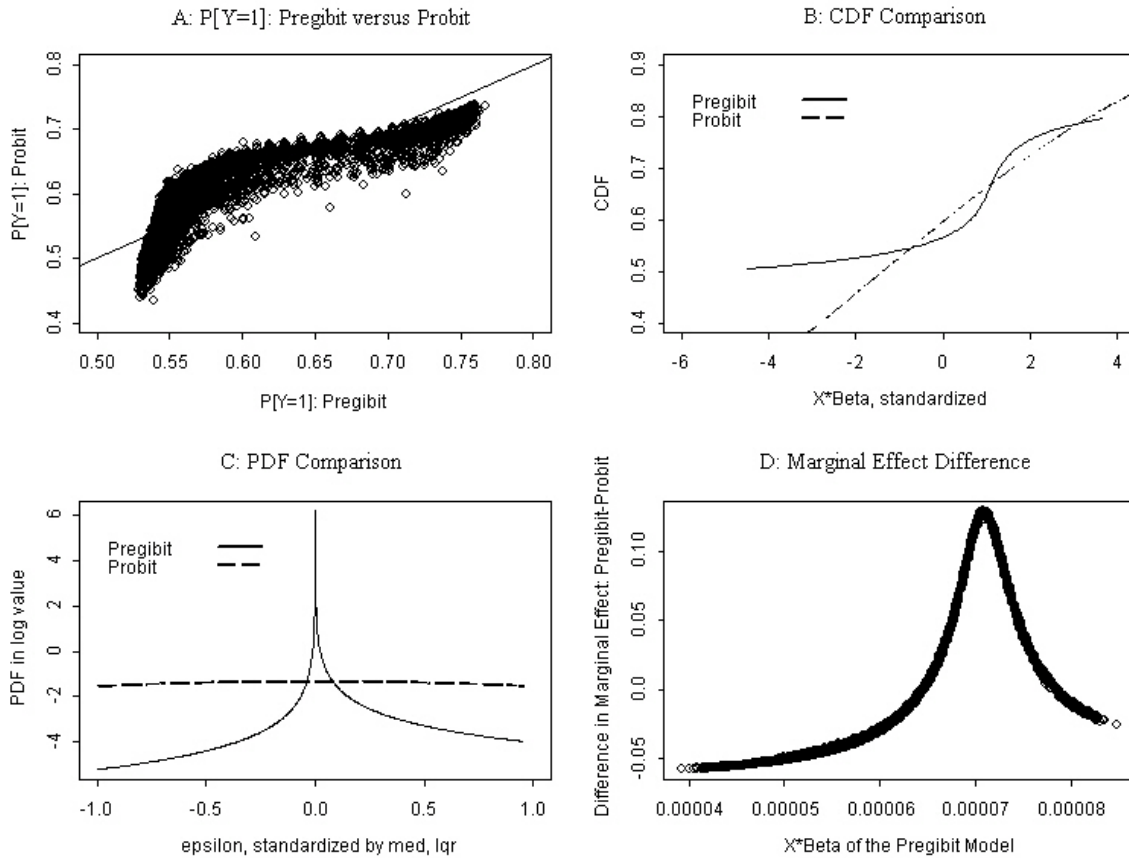
The six applications yield a range of experiences with the pregibit model. For three, the estimates of α and δ permit a traditional standardization with μ_ϵ and σ_ϵ, the mean and standard deviation of the disturbance, in the way that is done with probit implicitly—and we have done this also for each of the special cases of pregibit in Appendix C such that the slope estimates $\hat{\beta}$ become more easily comparable. But for the other three applications, μ_ϵ and σ_ϵ are not defined for the estimated values of α and δ. Section 2.2 suggested standardization by means of m_ϵ and s_ϵ, the median and half-IQR. Table 5, panel A, shows the benefit of doing so for the fifth application which deals with denial of mortgages. The probit estimates in the first column are traditional and therefore standardized, but the pregibit in the third column are unstandardized because standardization is not feasible. The slopes take on a different order of magnitude and simply are not comparable. However, the estimates of probit in column 2 and pregibit in column 4 are standardized with m_ϵ and s_ϵ and are comparable.[31] Pregibit parameter estimates are all somewhat smaller, but less so for 'PI ratio' (ratio of debt payments over income) and 'Credit score' (evidence of problematic credit history of the applicant) and more so for African-American and self-employed applicants. Thus, the pregibit model assigns more weight to

---

[31] In its unstandardized (or raw) form, the pregibit approximation of probit with $\alpha = 0.1436$ and $\delta = 0$ has a mean and median equal to 0, a standard deviation of 1.445, and an IQR of 1.950. The unrestricted, unstandardized pregibit model has a median of -1.006 and an IQR of 14.596. Thus, the distributions have a very different scale, but standardization makes them comparable (see also Figure 8, panel C).

objective financial variables and less to personal characteristics; this conclusion is supported also by the marginal effects that are computed for each variable separately in panel B.[32] Even so, there are some applicants for whom a small variation in $X'\hat{\beta}$ generates a large rise in the probability of success (Figure 8, panel B): for them, any variation in any variable generates a large marginal effect. Thus, the larger pregibit marginal effects are all substantially greater than the larger probit marginal effects, for every variable.

The pseudo-R^2 values that describe what the data reveal about the importance of observable determinants in explaining the variation in the latent variable y^* that underlies the observed outcomes y. For the probit model, R^2_{var} ranges from a high of 0.611 in the study of beetle mortality to a low of 0.030 in the study of visits to the doctor's office: pesticides inflict mortality in a highly predictive manner, but a visit to the doctor's office depends on many unobserved factors. The other studies found intermediate values of 0.301 (mortgage applications), 0.314 (post-secondary education), 0.373 (labor force participation), and 0.190 (sector choice). Within any given study, a model with higher log-likelihood value usually but not always yields a higher R^2_{var}. In the labor force participation study, R^2_{var} rises from 0.373 (probit) to 0.411 (unrestricted pregibit) and further to 0.424 for the linear probability model that is a restricted pregibit model. But indeed, the R^2_{var} focuses not on y but on y^*. In the mortgage application, post-secondary education and doctor's visits studies, R^2_{var} cannot be computed for the pregibit model because the variance of ϵ is undefined for the estimated values of α and δ. The design of a pseudo-R^2 that overcomes this weakness while summarizing explanatory evidence on y^* remains a topic for future research.

5. CONCLUDING REMARKS

The pregibit model adds considerable flexibility to the estimation of discrete choice models. It does not impose a rigid distribution on the disturbances that help generate successes and failures. Rather, it allows for symmetric or asymmetric disturbances, and it allows for distributions with fat tails, thin tails, or no tails. It nests, exactly or approximately, all of the models that researchers turn to when they study discrete choices: logit, probit, loglog, and cloglog; the gosset model that is built on the *t*-distribution; and a linear probability model that is reformulated within the context of the discrete choice framework.

The six applications that we explore illustrate the benefits of the pregibit model. For example, an analysis under a maintained assumption of normally distributed disturbances leads to biases in the parameter estimates and the marginal effects; as in the mortgage application, the pregibit model may place a greater emphasis on different variables than the probit model. The pregibit model may show thicker tails, which implies that observations with a predicted probability of success (or failure) near 1 under probit are not such a certain thing anymore, as the post-secondary enrollment application illustrates well.

---

[32] The African-American indicator matters more for African-Americans than for non-African-Americans in both the probit and pregibit models, but whereas the pregibit marginal effects are substantially smaller for non-African-Americans, they are generally only a little smaller for African-Americans (and on average even slightly larger). The same is true for Hispanics, but no such discrepancy is found for the self-employment effect by self-employment status. Thus, the flexibility of the pregibit model also yields differences in the way that explanatory variables determine outcomes across relevant subgroups in the sample.

The pregibit model is best able to distinguish itself from the traditional discrete choice models if the sample is large enough and contains near-certain successes and/or failures, provided naturally that the traditional models are a misspecification of the underlying data generating process. Since estimation of the pregibit model relies on more intensive numerical computations than the traditional models, it takes a little longer to obtain pregibit estimates than, say, logit estimates. Moreover, the likelihood function can become rather flat for larger negative values of α and δ and may exhibit local maxima for larger positive values of α and δ, such that the iterative search may have difficulty to converge. Thus, the gains that pregibit can bring about come at some cost—but as the six applications show, the model can yield significant new insights into the determinants of success.

The pregibit model is designed for a single-equation discrete choice context. The exploration in Section 2 yields useful insights also for the case where the dependent variable is continuous: the GTL distribution can generalize standard regression models as well. Furthermore, through the copula method, bivariate distributions can be constructed that permit flexible marginal GTL distributions and non-zero correlations between the equations. This will permit a generalization of the Heckman selection model.

REFERENCES

Amemiya, Takeshi (1981). "Qualitative response models: a survey." *Journal of Economic Literature*, 19:4, 481-536.

Bazán, Jorge L., Helena Bolfarine, and Márcia Branco (2010). "A framework for skewed-probit links in binary regression." *Communications in Statistics—Theory and Methods*, 39, 678-697.

Berndt, Ernst (1991). *The Practice of Econometrics: Classic and Contemporary*. Reading, MA : Addison-Wesley

Bliss, C.I. (1935). "The calculation of the dosage mortality curve." *Annals of Applied Biology*, 22, 134-167.

Brown, B.W. (1980). "Prediction analysis for binary data." In *Biostatistics Casebook*. R.G. Miller, B. Efron, B.W. Brown and L.E. Moses (editors), 3–18. John Wiley.

Davison, A.C. and D.V. Hinkley (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.

Freimer, M., G. S. Mudholkar, G. Kollia, and C. T. Lin (1988). "A Study of the Generalized Tukey Lambda Family." *Computational Statistics – Theory and Method*, 17:10, 3547-3567.

Greene, William H. (2012). *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

Joiner, Brian L., and Joan R. Rosenblatt (1971). "Some properties of the range in samples from Tukey's symmetric lambda distributions." *Journal of the American Statistical Association*, 66:334, 394-399.

Kelchtermans, Stijn, and Frank Verboven (2010). "Participation and study decisions in a public system of higher education." *Journal of Applied Econometrics*, 25, 355-391.

Koenker, Roger, and Jungmo Yoon (2009). "Parametric links for binary choice models: A Fisherian_Bayesian colloquy." *Journal of Econometrics*, 152, 120-130.

Laitila, T. (1993). "A pseudo-R2 measure for limited and qualitative dependent variable models." *Journal of Econometrics*, 56, 341-356.

McKelvey, R.D., and W. Zavoina (1975). "A statistical model for the analysis of ordinal level dependent variables." *Journal of Mathematical Sociology*, 4, 103-120.

Liu, Chuanhai (2006). "Robit regression: a simple robust alternative to logistic and probit regression." Working Paper, Bell Laboratories.

Mroz, Thomas A. (1987). "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions." *Econometrica*, 55:4, 765-799.

Munnell, A.H., G.M.B. Tootell, and L.E. Browne (1996). "Mortgage lending in Boston: Interpreting HMDA data." *American Economic Review*, 86:1, 25-53.

Nagler, Jonathan (1994). "Scobit: An alternative estimator to logit and probit." *American Journal of Political Science*, 38:1, 230–255.

Pregibon, Daryl (1980). "Goodness of link tests for generalized linear models." *Journal of the Royal Statistics Society, Series C (Applied Statistics)*. 29:1, 15-23.

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

Ramalho, Esmeralda, and Joaquim Ramalho (2011). "Alternative estimating and testing empirical strategies for fractional regression models." *Journal of Economic Surveys*, 25:1, 19-69.

Ramberg, John S., and Bruce W. Schmeiser (1974). An approximate method for generating asymmetric random variables." *Communications of the ACM*, February, 17:2, 78-82.

Ramberg, John S., Pandu R. Tadikamalla, Edward J. Dudewicz, and Edward F. Mykytka (1979). "A probability distribution and its uses in fitting data." *Technometrics*, 21:2, 201-214.

Riphahn, Regina., A. Wambach, and A. Million (2003). "Incentive effects in the demand for health care: a bivariate panel count data estimation." *Journal of Applied Econometrics*, 18:4, 387-405.

Stock, James H., and Mark W. Watson (2007). *Introduction to Econometrics*. Boston, MA: Pearson.

Stukel, Thérèse A. (1998). "Generalized logistic models." *Journal of the American Statistical Association*. 83:402, 425-431.

Vijverberg, Wim P.M. (2000). "Betit: A Class That Nests Logit and Probit," IZA Discussion Paper 222, IZA, Bonn.

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

Windmeier, Frans (1995). "Goodness of Fit Measures in Binary Choice Models." *Econometrics Reviews*, 14, 101-116.

Wooldridge, Jeffrey M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Table 1: Characteristics of generalized Tukey lambda distributions

	$\alpha - \delta$						
$\alpha + \delta$	−0.4	−0.2	−0.1	0	1	2	4
Mean							
−0.4	0.00	0.42	0.56	0.67	1.17	1.33	1.47
−0.2	−0.42	0.00	0.14	0.25	0.75	0.92	1.05
−0.1	−0.56	−0.14	0.00	0.11	0.61	0.78	0.91
0	−0.67	−0.25	−0.11	0.00	0.50	0.67	0.80
1	−1.17	−0.75	−0.61	−0.50	0.00	0.17	0.30
2	−1.33	−0.92	−0.78	−0.67	−0.17	0.00	0.13
4	−1.47	−1.05	−0.91	−0.80	−0.30	−0.13	0.00
Standard deviation							
−0.4	5.68	4.44	4.27	4.17	3.87	3.81	3.77
−0.2	4.44	2.74	2.47	2.31	1.84	1.74	1.67
−0.1	4.27	2.47	2.19	2.01	1.49	1.38	1.30
0	4.17	2.31	2.01	1.81	1.26	1.14	1.06
1	3.87	1.84	1.49	1.26	0.58	0.43	0.35
2	3.81	1.74	1.38	1.14	0.43	0.29	0.20
4	3.77	1.67	1.30	1.06	0.35	0.20	0.12
Skewness							
−0.4	…	…	…	…	…	…	…
−0.2	…	0.00	1.10	1.70	3.58	4.07	4.40
−0.1	…	−1.10	0.00	0.55	2.08	2.45	2.69
0	…	−1.70	−0.55	0.00	1.42	1.74	1.94
1	…	−3.58	−2.08	−1.42	0.00	0.23	0.26
2	…	−4.07	−2.45	−1.74	−0.23	0.00	−0.08
4	…	−4.40	−2.69	−1.94	−0.26	0.08	0.00
Kurtosis							
−0.4	…	…	…	…	…	…	…
−0.2	…	22.21	18.45	21.92	47.56	57.79	65.90
−0.1	…	18.45	6.79	6.05	11.44	14.10	16.20
0	…	21.92	6.05	4.20	5.90	7.29	8.38
1	…	47.56	11.44	5.90	1.80	1.84	1.94
2	…	57.79	14.10	7.29	1.84	1.80	1.96
4	…	65.90	16.20	8.38	1.94	1.96	2.45

Table 2: Matching standard normal, $t(5)$, and Gumbel distributions to GTL

	Standard normal matched by[a]			$t(5)$ matched by[b]			Gumbel matched by[c]		
	Moments	CDF	PDF	Moments	CDF	PDF	Moments	CDF	PDF
α	0.1349	0.1421	0.1436	-0.1359	-0.0783	-0.0710	0.2107	0.1671	0.1422
δ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.2419	-0.2334	-0.2290
GTL: κ_3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.1412	1.2797	1.3796
GTL: κ_4	3.0001	2.9620	2.9542	9.0032	5.9352	5.7005	5.4001	6.4868	7.3851
Absolute difference indices									
CDF	0.0028	0.0015	0.0016	0.0326	0.0076	0.0087	0.0190	0.0120	0.0152
PDF	0.0067	0.0033	0.0031	0.0650	0.0147	0.0128	0.0571	0.0350	0.0307
Difference in log-likelihood value across 100 runs[d]									
Average	-0.0048	-0.0005	-0.0003	0.0269	0.0019	0.0029	-0.7686	-0.4108	-0.2070
Minimum	-0.1768	-0.0645	-0.0334	-0.6501	-0.2942	-0.2332	-1.4619	-0.8167	-0.5515
Maximum	0.0908	0.0417	0.0309	0.5977	0.2388	0.1780	-0.0394	-0.0145	0.1746
Largest absolute difference in estimated probability of success across 100 runs[d]									
Average	0.0019	0.0009	0.0007	0.0133	0.0050	0.0040	0.0104	0.0089	0.0086

Notes: [a] For standard normal, $\kappa_3 = 0$ and $\kappa_4 = 3$.
 [b] For $t(5)$, $\kappa_3 = 0$ and $\kappa_4 = 9$.
 [c] For Gumbel, $\kappa_3 = 1.1396$ and $\kappa_4 = 5.4$. The match by moments is the result of minimizing the sum of the absolute differences in skewness and kurtosis of the Gumbel and GTL distributions.
 [d] Results of a Monte Carlo analysis of a model based on data of Application 4 in Section 4.

Table 3: Asymptotic Comparison of Logit, Probit, and Pregibit

Scenario		logit asy.value	$\Delta P_{logit}/\Delta P_{pregibit}$			probit asy.value	$\Delta P_{probit}/\Delta P_{pregibit}$		
α	δ	of β_1	min	med	max	of β_1	min	med	max
Range of P[Y=1] = (0.01–0.99)									
-0.4	0	1.538	0.460	1.200	1.772	1.452	0.207	1.324	2.361
-0.2	-0.27	2.273	0.124	0.944	2.992	2.048	0.024	1.036	3.707
-0.2	0	1.098	0.749	1.098	1.299	1.080	0.469	1.175	1.644
-0.2	0.27	2.273	0.124	0.944	2.992	2.048	0.024	1.036	3.707
0	0	1.000	1.000	1.000	1.000	1.018	0.814	1.062	1.206
.1349	0	0.964	0.861	0.940	1.088	0.999	0.976	1.002	1.012
0.5	-0.5	0.854	0.337	0.953	1.241	0.856	0.359	1.002	1.366
0.5	0	0.887	0.610	0.838	1.228	0.950	0.685	0.900	1.157
0.5	0.5	0.854	0.337	0.953	1.241	0.856	0.359	1.002	1.366
1	0	0.817	0.355	0.884	1.284	0.891	0.392	0.924	1.232
Range of P[Y=1] = (0.5–0.99)									
-0.4	0	1.255	0.498	1.216	1.578	1.076	0.367	1.313	1.91
-0.2	-0.27	4.716	0.907	0.968	1.259	4.660	0.964	1.023	1.097
-0.2	0	1.008	0.757	1.087	1.252	0.920	0.600	1.142	1.506
-0.2	0.27	1.415	0.346	1.283	1.706	1.182	0.246	1.402	2.052
0	0	1.000	1.000	1.000	1.000	0.964	0.843	1.041	1.178
.1349	0	1.004	0.869	0.957	1.136	0.996	0.981	0.999	1.009
0.5	-0.5	1.608	0.331	1.023	1.715	1.742	0.366	1.048	1.642
0.5	0	0.981	0.615	0.936	1.361	1.025	0.688	0.981	1.250
0.5	0.5	0.725	0.783	1.007	1.046	0.690	0.651	1.060	1.208
1	0	0.926	0.356	1.017	1.455	0.993	0.393	1.045	1.373
Range of P[Y=1] = (0.333–0.667)									
-0.4	0	2.337	0.984	1.005	1.061	2.642	0.979	1.007	1.083
-0.2	-0.27	3.676	0.870	0.999	1.196	4.156	0.885	0.994	1.223
-0.2	0	1.303	0.992	1.002	1.029	1.474	0.987	1.004	1.049
-0.2	0.27	3.676	0.870	0.999	1.196	4.156	0.885	0.994	1.223
0	0	1.000	1.000	1.000	1.000	1.131	0.995	1.002	1.020
.1349	0	0.890	0.983	0.999	1.005	1.007	0.999	1.000	1.002
0.5	-0.5	0.922	0.753	1.023	1.165	1.043	0.765	1.018	1.191
0.5	0	0.733	0.945	0.996	1.015	0.829	0.963	0.997	1.010
0.5	0.5	0.922	0.753	1.023	1.165	1.043	0.765	1.018	1.191
1	0	0.651	0.912	0.994	1.023	0.737	0.930	0.995	1.018

Table 4: Evidence of Pregibit in Six Datasets

	Beetle mortality	Labor force participation	Sector of employment	Demand for post-sec. education	Denial of mortgage	Visit to doctor's office
			Vijverberg and Zeager	Kelchtermans and Verboven	Munnell et al	Riphahn et al
	Bliss	Mroz (1987)	(1994)	(2010)	(1996)	(2003)
Sample size	477	753	1820	14464	2925	7293
Range of $P[Y = 1]$ (Probit)						
1-pctile	0.059	0.038	0.239	0.102	0.012	0.472
average	0.614	0.569	0.609	0.656	0.144	0.594
99-pctile	0.988	0.975	0.958	0.956	0.789	0.727
Parameter estimates[a]						
α	−0.102	0.500	1.472	−1.467	−1.626	−10.037
	(0.352)	(0.477)	(1.061)	(0.210)	(0.563)	(0.440)
δ	−0.284	−0.057	0.682	−0.016	0.919	−4.760
	(0.152)	(0.139)	(0.851)	(0.066)	(0.364)	(0.241)
Tests of restricted models						
Probit[b]						
LR	6.04	1.01	3.54	106.81	24.02	17.18
GL	5.33	0.41	2.68	79.52	9.24	12.98
LM	5.65	0.37	2.65	101.11	16.11	12.89
Logit[b]						
LR	7.07	1.43	4.24	83.62	19.27	17.65
GL	5.80	0.48	3.25	69.46	9.62	13.36
LM	6.46	0.48	3.30	71.70	15.90	13.39
Linear probability model[b]						
LR	7.99	0.72	1.78	307.47	85.64	15.31
GL	7.46	0.76	n.av.	128.05	n.av.	11.36
LM	6.53	0.21	0.54	199.28	65.90	10.88
Cloglog[b]						
LR	0.69	1.14	3.17	140.06	33.88	15.27
GL	0.75	0.77	1.03	71.86	18.60	11.20
LM	0.78	0.76	1.00	84.16	35.58	10.90
Loglog[b]						
LR	22.35	5.34	3.65	152.03	41.31	19.21
GL	14.53	3.32	3.52	79.62	25.28	14.81
LM	18.91	3.50	2.66	307.17	40.03	15.09
Symmetric Pregibit[c]						
LR	4.78	0.17	1.59	0.06	15.89	15.02
GL	4.07	0.14	n.av.	0.04	9.19	9.19
LM	4.42	0.07	0.49	0.04	17.74	10.39

Notes: [a] Estimates of the full models are given in the Appendix. Standard errors in parentheses. Estimates in the last column are provided subject to a caution about convergence, as discussed in the text.
[b] Critical values: 3.22 (20%), 4.61 (10%), 5.99 (5%), 9.21 (1%), 13.82 (0.1%).
[c] Critical values: 1.64 (20%), 2.71 (10%), 3.84 (5%), 6.63 (1%), 10.83 (0.1%).

Table 5: Comparing probit and pregibit estimation results: Denial of mortgage

	Probit[a]			Probit[c]			Pregibit[b]			Pregibit[c]		
A: Model estimates												
	Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.	
African-American	0.509	0.078	***	0.754	0.115	***	3.367	1.569	**	0.461	0.248	*
Hispanic	0.354	0.111	***	0.525	0.164	***	2.521	1.262	**	0.345	0.213	
PI ratio	3.413	0.434	***	5.058	0.644	***	33.497	17.925	*	4.590	2.252	**
HEI ratio	-0.744	0.517		-1.103	0.766		-4.881	4.685		-0.669	0.663	
LV medium	0.340	0.068	***	0.504	0.100	***	2.646	1.409	*	0.363	0.194	*
LV high	1.009	0.136	***	1.495	0.201	***	7.097	3.468	**	0.973	0.490	**
Credit score	0.253	0.061	***	0.374	0.090	***	2.480	1.432	*	0.340	0.168	**
Public record	0.907	0.094	***	1.343	0.139	***	6.184	2.945	**	0.847	0.411	**
Self-employed	0.280	0.095	***	0.415	0.141	***	1.683	1.041		0.231	0.154	
Intercept	-3.010	0.166	***	-4.460	0.246	***	-23.449	11.604	**	-3.351	1.611	**
α	0.1436			0.1436			-1.626	0.563	***	-1.626	0.563	***
δ	0			0			0.919	0.364	**	0.919	0.364	**
lnL	-974.68			-974.68			-962.48			-962.48		

B: Marginal effects[d]

	Probit				Pregibit			
	Average	5th pct	50th pct	95th pct	Average	5th pct	50th pct	95th pct
African-American	0.101	0.066	0.102	0.137	0.088	0.026	0.063	0.252
Hispanic	0.069	0.043	0.069	0.096	0.064	0.018	0.044	0.199
PI ratio	0.071	0.020	0.063	0.149	0.084	0.013	0.047	0.319
HEI ratio[e]	-0.013	-0.004	-0.012	-0.028	-0.011	-0.002	-0.006	-0.040
LV medium	0.063	0.040	0.061	0.092	0.058	0.019	0.040	0.193
LV high	0.223	0.158	0.228	0.270	0.251	0.066	0.228	0.456
Credit score	0.025	0.007	0.022	0.052	0.030	0.004	0.016	0.112
Public record	0.197	0.137	0.201	0.243	0.211	0.056	0.175	0.420
Self-employed	0.053	0.032	0.054	0.076	0.041	0.012	0.028	0.133
$X\hat{\beta}$	0.117	0.033	0.105	0.246	0.120	0.018	0.066	0.453

Notes: [a] The model is estimated in standardized form on the basis of mean and standard deviation.
[b] The model is estimated in unstandardized form.
[c] The model is estimated in standardized form on the basis of median and half-interquartile range.
[d] For dummy variables, marginal effects reflect the change in probability of mortgage denial when the variable changes from 0 to 1. For continuous variables, marginal effects are computed as the product of the density function, the parameter estimate, and the standard deviation of the variable.
[e] Percentile values are ordered according to their absolute value, i.e., from least negative to most negative.
***, **, * denote asymptotic significance levels of the *t*-statistic at 1, 5 and 10 percent level.

A.1 Lower and Upper Limit of the GTL Distribution

The link function is given by the following formula

$$\epsilon = G(q;\theta) = \frac{q^{\alpha-\delta}-1}{\alpha-\delta} - \frac{(1-q)^{\alpha+\delta}-1}{\alpha+\delta} \tag{A.1}$$

The lower bound ϵ_L is found as follows:

$$\epsilon_L = \lim_{q\downarrow0} G(\epsilon) = \lim_{q\downarrow0} \frac{q^{\alpha-\delta}-1}{\alpha-\delta} - \lim_{q\downarrow0} \frac{(1-q)^{\alpha+\delta}-1}{\alpha+\delta}$$

$$= \lim_{q\downarrow0} \frac{q^{\alpha-\delta}-1}{\alpha-\delta} - 0$$

$$= -\infty \ \text{if} \ \alpha-\delta \leq 0 \ \text{or} \ \delta \geq \alpha$$

$$= -\frac{1}{\alpha-\delta} \ \text{if} \ \alpha-\delta > 0 \ \text{or} \ \delta < \alpha \tag{A.2}$$

The upper bound ϵ_H is found in a similar way:

$$\epsilon_H = \lim_{q\uparrow1} G(\epsilon) = \lim_{q\uparrow1} \frac{q^{\alpha-\delta}-1}{\alpha-\delta} - \lim_{q\uparrow1} \frac{(1-q)^{\alpha+\delta}-1}{\alpha+\delta}$$

$$= 0 - \lim_{q\uparrow1} \frac{(1-q)^{\alpha+\delta}-1}{\alpha+\delta}$$

$$= \infty \ \text{if} \ \alpha+\delta \leq 0 \ \text{or} \ \delta \leq -\alpha$$

$$= \frac{1}{\alpha+\delta} \ \text{if} \ \alpha+\delta > 0 \ \text{or} \ \delta > -\alpha \tag{A.3}$$

Thus, in particular, the range is finite on both ends if $-\alpha < \delta < \alpha$.

A.2 The Moments of the GTL Distribution

When computing moments, it is always easier to make a change of variable from ϵ to $u = g^{-1}(\epsilon)$. Thus, $\epsilon = g(u)$ and the jacobian $\frac{d\epsilon}{du} = g'(u)$ happens to be equal the ratio of 1 over the density function. Accordingly, the mean is found to be

$$E[\epsilon] = \int_{\epsilon_L}^{\epsilon_U} \epsilon f(\epsilon)d\epsilon = \int_0^1 g(u)du = \frac{1}{\alpha-\delta}\int_0^1 (u^{\alpha-\delta}-1)du - \frac{1}{\alpha+\delta}\int_0^1 ((1-u)^{\alpha+\delta}-1)du$$

$$= -\frac{1}{\alpha-\delta+1} + \frac{1}{\alpha+\delta+1} = \frac{-2\delta}{(\alpha-\delta+1)(\alpha+\delta+1)} \tag{A.4}$$

The integrals may be solved provided that $\alpha-\delta+1 > 0$ and $\alpha+\delta+1 > 0$, or that $-(\alpha+1) < \delta < \alpha+1$.

The solution of the second, third and fourth moments depends on several intermediate results. The first are properties of symmetry in particular expressions, where λ may denote either $\alpha - \delta$ or $\alpha + \delta$, $\lambda_1 = \alpha - \delta$ and $\lambda_2 = \alpha + \delta$.

$$\int_0^1 \frac{1}{\lambda^k}\left((1-u)^\lambda - 1\right)^k du = \int_0^1 \frac{1}{\lambda^k}\left(v^\lambda - 1\right)^k dv \equiv J_1(\lambda, k) \tag{A.5}$$

$$\int_0^1 \frac{1}{\lambda_1^{k_1}}\left(u^{\lambda_1} - 1\right)^{k_1} \frac{1}{\lambda_2^{k_2}}\left((1-u)^{\lambda_2} - 1\right)^{k_2} du \equiv J(\lambda_1, k_1, \lambda_2, k_2) = J_2(\lambda_2, k_2, \lambda_1, k_1) \tag{A.6}$$

Generally,

$$E[\epsilon^k] = \int_0^1 g(u)^k du \equiv \mu_k \tag{A.7}$$

Therefore, $\mu_\epsilon = \mu_1$, $\sigma_\epsilon^2 = \mu_2 - \mu_1^2$, $\kappa_3 = (\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3)/\sigma_\epsilon^3$, and $\kappa_4 = (\mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4)/\sigma_\epsilon^4$. From the definition of g, J_1 and J_2 follows:

$$E[\epsilon] = J_1(\lambda_1, 1) - J_1(\lambda_2, 1) \tag{A.8}$$

$$E[\epsilon^2] = J_1(\lambda_1, 2) - 2J_2(\lambda_1, 1, \lambda_2, 1) + J_1(\lambda_2, 2) \tag{A.9}$$

$$E[\epsilon^3] = J_1(\lambda_1, 3) - 3J_2(\lambda_1, 2, \lambda_2, 1) + 3J_2(\lambda_1, 1, \lambda_2, 2) - J_1(\lambda_2, 3) \tag{A.10}$$

$$E[\epsilon^4] = J_1(\lambda_1, 4) - 4J_2(\lambda_1, 2, \lambda_2, 1) + 6J_2(\lambda_1, 2, \lambda_2, 2) - 4J_2(\lambda_1, 1, \lambda_2, 2) + J_1(\lambda_2, 4) \tag{A.11}$$

The solution of J_1 is:

$$J_1(\lambda, k) = \frac{(-1)^k k!}{(k\lambda+1)((k-1)\lambda+1)\dots(\lambda+1)} = \frac{(-1)^k k!}{\prod_{j=1}^k (j\lambda+1)} \tag{A.12}$$

provided that $k\lambda + 1 > 0$, which, when applied to λ_1 and λ_2, implies that the kth moment exists if $-\alpha - \frac{1}{k} < \delta < \alpha + \frac{1}{k}$.

The solution of J_2 is stated for each case separately. We start with $k_1 = k_2 = 1$. For $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$, we have

$$J_2(\lambda_1, 1, \lambda_2, 1) = \frac{1}{\lambda_1\lambda_2}\left(B(\lambda_1 + 1, \lambda_2 + 1) - \frac{1}{\lambda_1+1} - \frac{1}{\lambda_1+1} + 1\right) \tag{A.13}$$

$$= \frac{1}{\lambda_2(\lambda_1+1)} - \frac{1}{\lambda_2(\lambda_2+1)\Gamma(\lambda_1+\lambda_2+2)}\frac{\Gamma(\lambda_1+\lambda_2+2)-\Gamma(\lambda_1+1)\Gamma(\lambda_2+1)}{\lambda_1} \tag{A.13'}$$

where $B(\cdot,\cdot)$ and $\Gamma(\cdot)$ are the beta and gamma functions, respectively. With $\lambda_1 \to 0$, the numerator and denominator of the last ratio in (A.13') both go to 0. L'Hôpital's Rule implies that this ratio goes to $\Gamma(\lambda_2 + 2)\big(\psi_0(\lambda_2 + 2) - \psi_0(1)\big)$, where $\psi_0(\cdot)$ is the digamma function (the derivative of $\ln\Gamma(\cdot)$). Thus,

$$J_2(0, 1, \lambda_2, 1) = \frac{1}{\lambda_2} - \frac{1}{\lambda_2(\lambda_2+1)}\big(\psi_0(\lambda_2 + 2) - \psi_0(1)\big) \tag{A.14}$$

The solution for $J_2(\lambda_1, 1, 0, 1)$ follows from (A.6) and (A.14). (For $\lambda_1 = \lambda_2 = 0$, the GTL density becomes the logistic density, in which case the moments are well-known.) Other cases are derived in the same way, where $\psi_1(\cdot)$ is the derivative of $\psi_0(\cdot)$ and $\psi_2(\cdot)$ is the derivative of $\psi_1(\cdot)$:

$$J_2(\lambda_1, 2, \lambda_2, 1) = \frac{1}{\lambda_1^2 \lambda_2} \left(B(2\lambda_1 + 1, \lambda_2 + 1) - 2B(\lambda_1 + 1, \lambda_2 + 1) + \frac{1}{\lambda_2 + 1} - \frac{2\lambda_1^2}{(\lambda_1 + 1)(2\lambda_1 + 1)} \right) \quad (A.15)$$

$$J_2(0, 2, \lambda_2, 1) = \frac{-2}{\lambda_2} + \frac{1}{\lambda_2(\lambda_2 + 1)} \left\{ \left(\psi_0(\lambda_2 + 2) - \psi_0(1) \right)^2 - \left(\psi_1(\lambda_2 + 2) - \psi_1(1) \right) \right\} \quad (A.16)$$

$$J_2(\lambda_1, 2, 0, 1) = \frac{1}{\lambda_1^2} \left\{ 1 + \frac{1}{2\lambda_1 + 1} \left(\psi_0(2\lambda_1 + 2) - \psi_0(1) \right) - \frac{2}{\lambda_1 + 1} \left(\psi_0(\lambda_1 + 2) - \psi_0(1) \right) \right\} \quad (A.17)$$

$$J_2(\lambda_1, 3, \lambda_2, 1) = \frac{1}{\lambda_1^3 \lambda_2} \left(B(3\lambda_1 + 1, \lambda_2 + 1) - 3B(2\lambda_1 + 1, \lambda_2 + 1) + 3B(\lambda_1 + 1, \lambda_2 + 1) + c \right)$$

$$(A.18)$$

where $c = -\frac{1}{\lambda_2 + 1} - \frac{1}{3\lambda_1 + 1} + \frac{3}{2\lambda_1 + 1} - \frac{3}{\lambda_1 + 1} + 1$.

$$J_2(0, 3, \lambda_2, 1) = \frac{6}{\lambda_2} - \frac{1}{\lambda_2(\lambda_2 + 1)} \left\{ \left(\psi_0(\lambda_2 + 2) - \psi_0(1) \right)^3 + c \right\} \quad (A.19)$$

where $c = -3(\psi_1(\lambda_2 + 2) - \psi_1(1))(\psi_0(\lambda_2 + 2) - \psi_0(1)) + (\psi_2(\lambda_2 + 2) - \psi_2(1))$.

$$J_2(\lambda_1, 3, 0, 1) = \frac{1}{\lambda_1^3} \left\{ 1 - \frac{1}{3\lambda_1 + 1} \left(\psi_0(3\lambda_1 + 2) - \psi_0(1) \right) + c \right\} \quad (A.20)$$

where $c = \frac{3}{2\lambda_1 + 1} \left(\psi_0(2\lambda_1 + 2) - \psi_0(1) \right) - \frac{3}{\lambda_1 + 1} \left(\psi_0(\lambda_1 + 2) - \psi_0(1) \right)$.

$$J_2(\lambda_1, 2, \lambda_2, 2) = \frac{1}{\lambda_1^2 \lambda_2^2} \left(B(2\lambda_1 + 1, 2\lambda_2 + 1) - 2B(2\lambda_1 + 1, \lambda_2 + 1) - 2B(\lambda_1 + 1, 2\lambda_2 + 1) + c \right)$$

$$(A.21)$$

where $c = 4B(\lambda_1 + 1, \lambda_2 + 1) + \frac{1}{2\lambda_1 + 1} - \frac{2}{\lambda_1 + 1} + \frac{1}{2\lambda_2 + 1} - \frac{2}{\lambda_2 + 1} + 1$.

$$J_2(0, 2, \lambda_2, 2) = \frac{1}{\lambda_2^2} \left\{ 2 + \frac{1}{2\lambda_2 + 1} \left(\left(\psi_0(2\lambda_2 + 2) - \psi_0(1) \right)^2 - \left(\psi_1(2\lambda_2 + 2) - \psi_1(1) \right) \right) + c \right\} \quad (A.22)$$

where $c = -\frac{2}{\lambda_2 + 1} \left(\left(\psi_0(\lambda_2 + 2) - \psi_0(1) \right)^2 - \left(\psi_1(\lambda_2 + 2) - \psi_1(1) \right) \right)$.

APPENDIX B: GTL DISTRIBUTIONS AND THE UNIFORM DISTRIBUTION

As noted in Section 2.4, four pairs of (α, δ) values simplify the GTL distribution to a uniform distribution. The density functions for values neighboring these four sets of values exhibit very different shapes, however. Figure B.1 shows four densities for values around $(\alpha, \delta) = (1,0)$; Figure B.2 does the same for values around $(\alpha, \delta) = (2,0)$; and Figures B.3 aand B.4 show densities for $(\alpha, \delta) = (10, -9)$ and $(\alpha, \delta) = (10,9)$, respectively, in representation of the case where $(\alpha, \delta) = (\lambda, -\lambda + 1)$ for $\lambda \to \infty$ and $(\alpha, \delta) = (\lambda, \lambda - 1)$ for $\lambda \to \infty$. The vertical axes are curtailed in order to zoom in on the curvature of the densities.

Figure B.1: GTL densities near $(\alpha, \delta) = (1,0)$ **Figure B.2: GTL densities near $(\alpha, \delta) = (2,0)$**
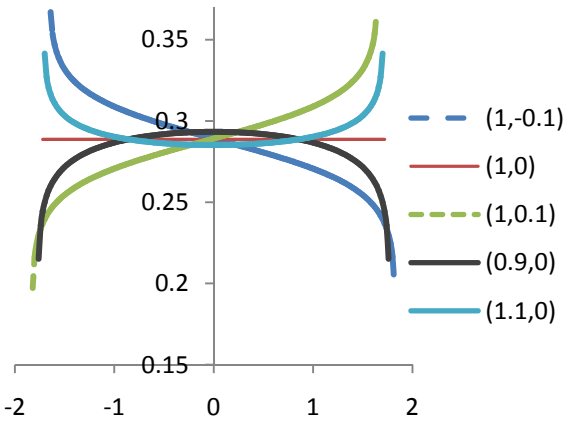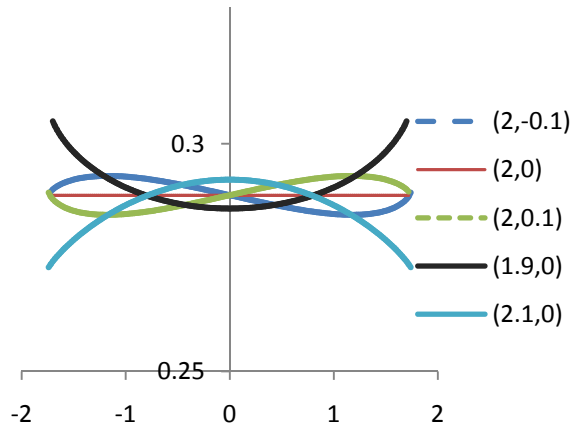
Figure B.3: GTL densities for $(\alpha, \delta) = (10, -9)$ **Figure B.4: GTL densities for $(\alpha, \delta) = (10,9)$**
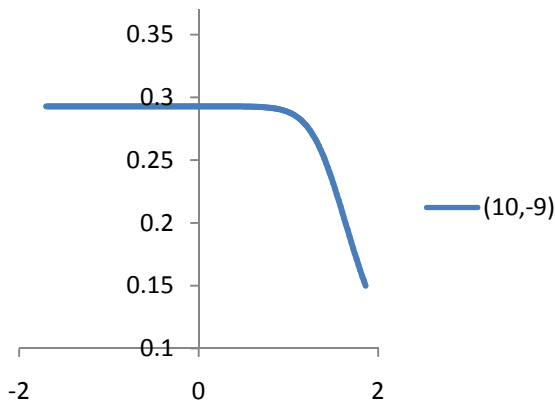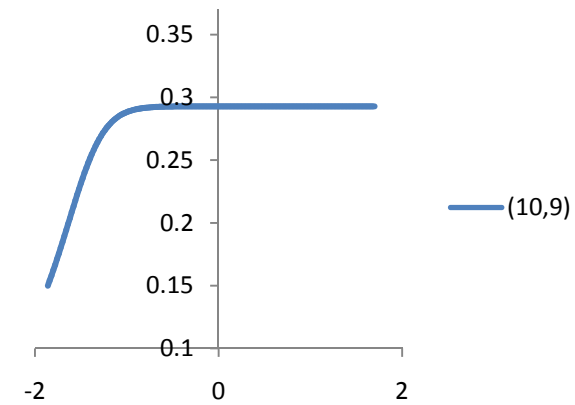
In each of the following applications, four specifications are shown. These were selected because (i) they were the specifications yielding the lowest log-likelihood value, (ii) they showed the performance under an *a priori* asymmetric distributional assumption, or (iii) they demonstrate the performance under various symmetry assumptions.

Table C.1: Beetle Mortality

A: Variable definition and descriptive statistics

Variable	Definition	Mean	Std.
Mortality[a]	Dummy: 1=beetle died, 0=beetle survived	0.610	0.488
lndosage	Log of dosage	1.794	0.063

Note: [a] Dependent variable
Number of observations: 477

B: Parameter estimates

	Probit[a]			Linear Probability[a]			Cloglog[a]			Pregibit[a]	
	Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.
lndosage	19.839	1.496	***	18.552	0.600	***	16.446	1.268	***	9.807	20.198
Intercept	-35.110	2.659	***	-32.836	1.099	***	-29.071	2.288	***	-17.275	35.751
α	0.1436			1			0.1422			-0.102	0.352
δ	0			0			-0.2290			-0.284	0.152
lnL	-181.92			-182.92			-179.27			-178.93	
R^2_{var}	0.611			0.580			0.521			0.279	
Mean[b]	0			0			0.366			0.783	
St.Dev.[b]	1.464			0.577			1.648			3.764	
Skewness[b]	0			0			1.380			n.a.	
Kurtosis[b]	3			1.800			7.385			n.a.	

Notes: [a] The model is estimated in standardized form. Other models: logit (lnL = -182.46), loglog (lnL = -190.11), and symmetric pregibit (lnL= -181.32, $\hat{\alpha} = 0.410$ (0.175)).
[b] Characteristics of the Generalized Tukey Lambda distribution in its unstandardized form.
***, **, * denote asymptoticsignificance levels of the t-statistic at 1, 5 and 10 percent level.

Table C.2: Female labor force participation

A: Variable definition and descriptive statistics

Variable	Definition	Mean	Std.
Lfp[a]	Dummy, =1 if person is employed, =0 if not	0.568	0.496
lnwage	Imputed logwage	1.095	0.315
kl6	Number of children less than 6 years of age	0.238	0.524
k618	Number of children between 6 and 18 years of age	1.353	1.320
age	Age (years)	42.538	8.073
educ	Schooling (years)	12.287	2.280
urate	Unemployment rate in the county of residence	8.624	3.115
city	Dummy, =1 if person lives in an SMSA	0.643	0.480
othinc	Other household income ($000s)	20.129	11.635

Note: [a] Dependent variable
Number of observations: 753

B: Parameter estimates

	Probit[a]			Linear Probability[a]			Symmetric Pregibit[a]			Pregibit[a]		
	Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.	
lnwage	2.324	0.273	***	2.562	0.252	***	2.528	0.274	***	2.474	0.333	***
kl6	-0.904	0.118	***	-1.074	0.121	***	-1.033	0.140	***	-1.023	0.153	***
k618	0.003	0.042		0.016	0.043		0.006	0.045		0.006	0.045	
age	-0.044	0.008	***	-0.050	0.008	***	-0.049	0.008	***	-0.048	0.009	***
educ	-0.075	0.036	**	-0.097	0.037	***	-0.088	0.040	**	-0.087	0.041	**
urate	-0.014	0.017		-0.015	0.020		-0.015	0.018		-0.014	0.018	
city	-0.150	0.114		-0.188	0.126		-0.171	0.123		-0.155	0.125	
othinc	-0.013	0.005	***	-0.013	0.005	***	-0.014	0.005	***	-0.014	0.005	***
Intercept	1.149	0.503	**	1.493	0.547	***	1.394	0.565	**	1.392	0.555	**
α	0.1436			1			0.568	0.430		0.500	0.477	
δ	0			0			0			-0.057	0.139	
lnL	-414.86			-414.74			-414.46			-414.37		
R^2_{var}	0.373			0.424			0.418			0.411		
Mean[b]	0			0			0			0.051		
St.Dev.[b]	1.464			0.577			0.862			0.930		
Skewness[b]	0			0			0			0.115		
Kurtosis[b]	3			1.800			2.014			2.105		

Notes: [a] The model is estimated in standardized form. Other models: logit (lnL = -415.09), cloglog (lnL = -414.94), and loglog (lnL = -417.04)
[b] Characteristics of the Generalized Tukey Lambda distribution in its unstandardized form.
***, **, * denote asymptotic significance levels of the t-statistic at 1, 5 and 10 percent level.

Table C.3: Sector choice in Tanzania

A: Variable definition and descriptive statistics

Variable	Definition	Mean	Std.
Public[a]	Dummy, =1 if person is employed, =0 if person is employed in private sector	0.609	0.488
edst1	Education: Dummy, =1 if person completed Standard 1-4 or more, =0 if not	0.814	0.389
edst5	Education: Dummy, =1 if person completed Standard 5-8 or more, =0 if not	0.601	0.490
edfm1	Education: Dummy, =1 if person completed Form I-IV or more, =0 if not	0.198	0.399
edfm5	Education: Dummy, =1 if person completed Form V, VI or more, =0 if not	0.024	0.152
eduni	Education: Dummy, =1 if person has a university degree, =0 if not	0.011	0.104
age	Age in years	29.827	9.783
sex	Dummy, =1 if person is female, =0 if male	0.132	0.338
married	Dummy, =1 if person is married, =0 if not	0.579	0.494
relig	Dummy, =1 if person is Christian, =0 if not	0.459	0.498
skilled	Dummy, =1 if person is in a skilled occupation, =0 if not	0.640	0.480
citizen	Dummy, =1 if person is citizen of Tanzania, =0 if not	0.960	0.195
salaam	Dummy, =1 if person resides in Dar Es Salaam, =0 if not	0.661	0.474

Note:　[a] Dependent variable
Number of observations: 1820

B: Parameter estimates

	Probit[a]			Linear Probability[a]			Cloglog[a]			Pregibit[a]		
	Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.	
edst1	-0.241	0.098	**	-0.341	0.127	***	-0.236	0.087	***	-0.263	0.144	*
edst5	0.435	0.085	***	0.550	0.108	***	0.367	0.073	***	0.545	0.112	***
edfm1	0.410	0.096	***	0.416	0.096	***	0.299	0.069	***	0.536	0.147	***
edfm5	0.975	0.509	*	0.809	0.617		0.660	0.279	**	1.078	0.482	**
eduni	-0.587	0.615		-0.285	0.664		-0.329	0.354		-0.527	0.649	
age	0.015	0.004	***	0.016	0.004	***	0.012	0.003	***	0.020	0.005	***
sex	-0.042	0.095		-0.078	0.111		-0.041	0.076		-0.074	0.116	
married	0.152	0.073	**	0.164	0.090	*	0.102	0.059	*	0.203	0.094	**
relig	0.165	0.067	**	0.221	0.078	***	0.155	0.054	***	0.194	0.098	**
skilled	0.403	0.070	***	0.545	0.091	***	0.351	0.060	***	0.530	0.106	***
citizen	0.762	0.167	***	0.945	0.199	***	0.664	0.150	***	0.912	0.187	***
salaam	0.126	0.067	*	0.158	0.078	**	0.098	0.054	*	0.171	0.084	**
Intercept	-1.540	0.234	***	-1.797	0.269	***	-1.098	0.200	***	-2.068	0.293	***
α	0.1436			1			0.1422			1.472	1.061	
δ	0			0			-0.2290			0.682	0.851	
lnL	-1107.16			-1106.30			-1107.00			-1105.41		
R^2_{var}	0.190			0.242			0.173			0.278		
Mean[b]	0			0			0.366			-0.242		
St.Dev.[b]	1.464			0.577			1.648			0.483		
Skewness[b]	0			0			1.380			-0.359		
Kurtosis[b]	3.000			1.800			7.385			1.968		

Notes:　[a] The model is estimated in standardized form. Other models: logit (lnL = -1107.53), loglog (lnL = -1107.24), and symmetric pregibit (lnL = -1106.21, $\hat{\alpha}$ = 0.744 (0.467)).
　[b] Characteristics of the Generalized Tukey Lambda distribution in its unstandardized form.
　***, **, * denote asymptotic significance levels of the t-statistic at 1, 5 and 10 percent level.

Table C.4: Demand for post-secondary education, Belgium

A: Variable definition and descriptive statistics

Variable	Definition	Mean	Std.
Tertiary[a]	Dummy, =1 if person enrolled at an institution of post-sec education, =0 if not	0.657	0.475
Male	Dummy, =1 if male, =0 if female	0.486	0.500
Foreign	Dummy, =1 if foreigner, =0 if Belgian national	0.019	0.135
Catholic	Dummy, =1 if catholic high school, =0 if not	0.778	0.416
Classical	Dummy, =1 if high school field in classical languages, =0 if not	0.105	0.307
Lang	Dummy, =1 if high school field in modern languages, =0 if not	0.184	0.387
Econ	Dummy, =1 if high school field in economics, =0 if not	0.147	0.354
Science	Dummy, =1 if high school field in sciences, =0 if not	0.160	0.366
Math	Dummy, =1 if high school field in mathematics, =0 if not	0.229	0.420
TSOH	Dummy, =1 if `product-focused' technical high school, =0 if not	0.142	0.349
TSO	Dummy, =1 if technical high school, =0 if not	0.356	0.479
Years rep	Years of repetition (= age - 18)	1.603	0.908
Meantrav	Average travel cost to institutions of post-secondary education	0.836	0.201
Mintrav	Minimum travel cost to institutions of post-secondary education	0.163	0.103

Note: [a] Dependent variable
Number of observations: 14464

B: Parameter estimates

	Probit[a]			Logit[a]			Symmetric Pregibit[b]			Cloglog[a]		
	Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.	
Male	-0.166	0.026	***	-0.162	0.024	***	-1.529	0.372	***	-0.110	0.019	***
Foreign	-0.512	0.087	***	-0.511	0.083	***	-6.382	2.057	***	-0.496	0.086	***
Catholic	0.148	0.028	***	0.128	0.026	***	0.395	0.225	*	0.153	0.022	***
Classical	0.574	0.054	***	0.563	0.055	***	8.445	2.871	***	0.351	0.032	***
Lang	0.351	0.046	***	0.310	0.045	***	2.725	1.028	***	0.282	0.031	***
Econ	0.563	0.053	***	0.536	0.052	***	5.244	1.467	***	0.378	0.035	***
Science	0.451	0.050	***	0.433	0.049	***	4.509	1.309	***	0.288	0.032	***
Math	0.410	0.046	***	0.372	0.045	***	3.954	1.268	***	0.310	0.032	***
TSOH	-0.225	0.040	***	-0.200	0.036	***	-1.768	0.508	***	-0.189	0.032	***
TSO	0.507	0.034	***	0.457	0.031	***	3.415	0.751	***	0.389	0.029	***
Years rep	-0.387	0.015	***	-0.361	0.015	***	-3.016	0.641	***	-0.350	0.014	***
Meantrav	-0.211	0.064	***	-0.212	0.060	***	-2.456	0.717	***	-0.109	0.048	**
Mintrav	-0.344	0.125	***	-0.303	0.116	***	-1.952	1.046	*	-0.290	0.095	***
Intercept	0.774	0.069	***	0.743	0.064	***	7.419	1.682	***	0.765	0.055	***
α	0.1436			0			-1.450	0.197	***	0.1422		
δ	0			0			0			-0.2290		
lnL	-7627.6			-7616.01			-7574.23			-7644.23		
R^2_{var}	0.314			0.286			n.a.			0.229		
Mean[c]	0			0			n.a.			0.366		
St.dev.[c]	1.464			1.814			n.a.			1.648		
Skewness[c]	0			0			n.a.			1.380		
Kurtosis[c]	3.000			4.200			n.a.			7.385		

Notes: [a] The model is estimated in standardized form. Other models: linear probability (lnL = -7727.94) and loglog (lnL = -7650.21).
[b] The model is estimated in unstandardized form. Other model: unrestricted pregibit (lnL = -7574.20, $\hat{\alpha} = -1.467$ (0.210) and $\hat{\delta} = -0.016$ (0.066)).
[c] Characteristics of the Generalized Tukey Lambda distribution in its unstandardized form.
***, **, * denote asymptotic significance levels of the t-statistic at 1, 5 and 10 percent level.

Table C.5: Result of mortgage application

A: Variable definition and descriptive statistics

Variable	Obs	Mean	Std.
Deny[a]	Dummy, =1 if mortgage application was denied, =0 if not	0.144	0.351
Black	Dummy, =1 if applicant was African-American, =0 if not	0.160	0.367
Hisp	Dummy, =1 if applicant was Hispanic, =0 if not	0.073	0.260
PIratio	Ratio of total monthly debt payment to total monthly income	0.333	0.113
HEIratio	Ratio of housing expense to income	0.255	0.098
LVmed	Loan/Appraised value: medium (between 0.80 and 0.95)	0.409	0.492
LVhi	Loan/Appraised value: high (above 0.95)	0.039	0.193
Creditscore	Consumer credit score (index between 1 and 6)	1.742	0.533
Publicrecord	Dummy, =1 if public record of credit problems, =0 if not	0.079	0.270
Selfemp	Dummy, =1 if applicant was self-employed, =0 if not	0.117	0.322

Note: [a] Dependent variable
Number of observations: 2925

B: Parameter estimates

	Logit[a]			Symmetric Pregibit[a]			Cloglog[a]			Pregibit[b]		
	Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.	
Black	0.501	0.076	***	1.319	0.337	***	0.572	0.086	***	3.367	1.569	**
Hisp	0.364	0.108	***	1.029	0.379	***	0.434	0.125	***	2.521	1.262	**
PIratio	3.653	0.459	***	11.007	3.204	***	3.415	0.392	***	33.497	17.925	*
HEIratio	-0.724	0.519		-2.024	1.469		-0.532	0.603		-4.881	4.685	
LVmed	0.364	0.070	***	1.072	0.342	***	0.511	0.087	***	2.646	1.409	*
LVhi	0.986	0.129	***	2.607	0.649	***	1.123	0.134	***	7.097	3.468	**
Creditscore	0.266	0.063	***	0.765	0.259	***	0.305	0.071	***	2.480	1.432	*
Publicrecord	0.862	0.089	***	2.176	0.467	***	0.899	0.091	***	6.184	2.945	**
Selfemp	0.264	0.097	***	0.640	0.290	**	0.257	0.111	**	1.683	1.041	
Intercept	-3.079	0.184	***	-8.638	2.107	***	-3.244	0.192	***	-23.449	11.604	**
α	0			-0.257	0.142	*	0.142			-1.626	0.563	***
δ	0			0			-0.229			0.919	0.364	**
lnL	-972.11			-970.43			-979.42			-962.48		
R^2_{var}	0.301			0.252			0.338			n.a.		
Mean[c]	0			0			0.366			n.a.		
St.dev.[c]	1.814			3.192			1.648			n.a.		
Skewness[c]	0			0			1.380			n.a.		
Kurtosis[c]	4.200			na			7.385			n.a.		

Notes: [a] The model is estimated in standardized form. Other models: probit (lnL = -974.68), linear probability
(lnL = -1005.30, failing to converge close to a presumed maximum), loglog (lnL = -983.14).
[b] The model is estimated in unstandardized form.
[c] Characteristics of the Generalized Tukey Lambda distribution in its unstandardized form.
***, **, * denote asymptotic significance levels of the t-statistic at 1, 5 and 10 percent level.

Table C.6: Visit to the doctor's office, Germany

A: Variable definition and descriptive statistics

Variable	Definition	Mean	Std.
Visit[a]	Dummy, =1 if household visited doctor's office during survey year, =0 if not	0.594	0.491
Age	Age in years, divided by 10	3.982	1.206
Hhninc	Household income, divided by 10000	0.324	0.164
Hhkids	Number of children in the household	0.414	0.493
Educ	Years of education of household head	11.377	2.306
Married	Dummy, =1 if household head is married, =0 if not	0.691	0.462

Note:　[a] Dependent variable
Number of observations: 7293

B: Parameter estimates

	Probit[a]			Symmetric Pregibit[a]			Cloglog[a]			Pregibit[b]		
	Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.	
Age	0.095	0.014	***	0.131	0.019	***	0.080	0.011	***	8.10E-06	3.62E-06	**
Hhninc	-0.113	0.094		-0.160	0.129		-0.096	0.080		6.10E-07	4.43E-06	
Hhkids	-0.186	0.034	***	-0.251	0.047	***	-0.153	0.028	***	-6.61E-06	3.80E-06	*
Educ	-0.022	0.007	***	-0.030	0.009	***	-0.018	0.006	***	-6.75E-07	5.46E-07	
Married	0.020	0.038		0.024	0.051		0.014	0.031		-7.61E-07	2.11E-06	
Intercept	0.219	0.101	**	0.294	0.141	**	0.338	0.085	***	3.76E-05	1.62E-05	**
α	0.1436			1.493	0.931		0.1422			-10.037	0.440	***
δ	0			0			-0.2290			-4.760	0.241	***
lnL	-4856.81			-4855.73			-4855.86			-4848.22		
R^2_{var}	0.030			0.055			0.021			n.a.		
Mean[c]	0			0			0.366			n.a.		
St.dev.[c]	1.464			0.398			1.648			n.a.		
Skewness[c]	0			0			1.380			n.a.		
Kurtosis[c]	3.000			1.753			7.385			n.a.		

Notes:　[a] The model is estimated in standardized form. Other models: logit (lnL = -4857.05), linear probability
(lnL = -4855.88) and loglog (lnL = -4857.83).
[b] The model is estimated in standardized form based on the median and half of the IQR. Estimates are
provided subject to a caution about convergence, as discussed in the text.
[c] Characteristics of the Generalized Tukey Lambda distribution in its unstandardized form.
***, **, * denote asymptotic significance levels of the t-statistic at 1, 5 and 10 percent level.