

Parametric Links for Binary Response

Roger Koenker

It is a textbook cliché that probit and logit link functions for binary response models are “very similar.” While true, this remark is unfortunately often extrapolated to imply that *all* links are essentially equivalent. Like most extrapolations, this inferential leap is dangerous. Some relatively minor recent changes in the `glm` functions of R2.4.0 make it easy to explore non-traditional link functions. This note is intended to illustrate this functionality, and encourage others to explore alternative links.

Since the early eighties, I’ve been teaching applied econometrics for economics graduate students at the University of Illinois. When we come to binary response models I always introduce the Pregibon (1980) “goodness of link” test, partly as a way to reinforce the ideas underlying the Box-Cox transformation and partly because I feel it is important to convey a more skeptical attitude about the conventional choice of logit and probit links.

Pregibon considers a two parameter generalization of the logit link that we can write as,

$$g(u, a, b) = \frac{u^{a-b} - 1}{a - b} - \frac{(1 - u)^{a+b} - 1}{a + b}.$$

When $a = b = 0$ the link reduces to the logit expression $\log(u/(1 - u))$, for $b = 0$ we have a family of symmetric densities with a controlling the heaviness of the tails, and when $b \neq 0$ it controls the skewness of the distribution. The function g is the quantile function of a version of the generalized Tukey- λ distributions, for which the `gld` package of King (2005) provides “p-q-r-d” functions for R.

Another natural (one-parameter) family of link functions for binary response is the Gosset, or Student-t, link with a free degrees of freedom parameter, ν . When $\nu = 1$ we have the so-called Cauchit link corresponding to the Cauchy quantile function, while for $\nu \rightarrow \infty$ we obtain the probit link. The Cauchit link is distinctly different than probit – admitting that even extreme values of the linear predictor can occasionally be overwhelmed by an even more extreme realization of the Cauchy innovation of the latent variable form of the binary response model. See Morgan and Smith (1992) for an example of Cauchit fitting.

Implementation

Parametric links may be specified in R as a structure consisting of functions specifying the link function, the inverse of the link function, the derivative of the inverse link function, a validity check function, and

a name to be used for the link. This object should be of class `link-glm`. The Gosset link is:

```
Gosset <- function(nu) {
  qqt <- function(p, nu)
    sign(p-0.5)*sqrt(qf(1-2*pmin(p,1-p), 1, nu))
  linkfun <- function(mu) qqt(mu,nu)
  linkinv <- function(eta) {
    thresh <- -qqt(.Machine$double.eps,nu)
    eta <- pmin(thresh, pmax(eta, -thresh))
    pt(eta, nu)}
  mu.eta <- function(eta)
    pmax(dt(eta, nu), .Machine$double.eps)
  valideta <- function(eta) TRUE
  name <- "Gosset"
  structure(list(linkfun=linkfun, linkinv=linkinv,
    mu.eta=mu.eta, valideta=valideta, name=name),
    class = "link-glm")}
```

Note that `qt` has been replaced by `qqt` since the former has a restricted domain for $\nu \geq 1$ while the version based on `qf` is reliable for $\nu \geq 0.2$. (Thanks to Luke Tierney for this suggestion.)

The Pregibon link is implemented like this:

```
Pregibon <- function(a, b) {
  linkfun <- function(mu)
    - qPregibon(1 - mu, a = a, b = b)
  linkinv <- function(eta) {
    eps <- .Machine$double.eps^.5
    tlo <- qPregibon(eps, a = a, b = b)
    thi <- qPregibon(1 - eps, a = a, b = b)
    eta <- -pmin(thi, pmax(-eta, tlo))
    1 - pPregibon(-eta, a = a, b = b)}
  mu.eta <- function(eta)
    pmax(dPregibon(-eta, a = a, b = b),
    .Machine$double.eps^.5)
  valideta <- function(eta) TRUE
  name <- "Pregibon"
  structure(list(linkfun=linkfun, linkinv=linkinv,
    mu.eta=mu.eta, valideta=valideta, name=name),
    class = "link-glm")}
```

Since the parameterization of the Pregibon link differs slightly from the parameterization used in the `gld` package we use the following code to define the required functions:

```
qPregibon <- function(x,a = 0,b = 0){
  s <- (qgl(3/4,c(0,1,a-b,a+b)) -
    qgl(1/4,c(0,1,a-b,a+b)))/2.197224
  qgl(x,c(0,1,a-b,a+b))/s}
pPregibon <- function(x,a = 0,b = 0,tol=1e-12){
  s <- (qgl(3/4,c(0,1,a-b,a+b)) -
    qgl(1/4,c(0,1,a-b,a+b)))/2.197224
  pgl(x*s, c(0,1,a-b,a+b),inverse.eps=tol)}
dPregibon <- function(x,a = 0,b = 0,tol=1e-12){
  s <- (qgl(3/4,c(0,1,a-b,a+b)) -
    qgl(1/4,c(0,1,a-b,a+b)))/2.197224
  dgl(x*s, c(0,1,a-b,a+b),inverse.eps=tol)*s}
rPregibon <- function(n,a = 0,b = 0){
  qPregibon(runif(n),a=a,b=b)}
```

Note that we have fixed scale so that all members of the family have the same interquartile range; recall that scale is unidentified in this class of models.

An Example

To illustrate the fitting procedure we will consider a model of quit behavior for Western Electric workers. The transformed probability of quitting within six months of starting a new job is modeled as a linear function of a gender indicator, the score on a pre-employment dexterity exam, and a quadratic function of years of education, denoted by *sex*, *dex*, and *lex*, respectively. The data come originally from the study of Klein et al. (1991), but have been modified over the years to heighten the pedagogical impact.

To fit a Gosset model with a fixed value of the degrees of freedom parameter one can simply write:

```
u <- "http://www.econ.uiuc.edu/~roger/courses/
471/data/weco.dat"
d <- read.table(u, header=TRUE)
f <- glm(kwit ~ sex + dex + poly(lex, 2),
data=d, family=binomial(link=Gosset(1.0)))
```

This is equivalent to:

```
f <- glm(kwit ~ sex + dex + poly(lex, 2),
data=d, family=binomial(link="cauchit"))
```

but of course the value 1.0 can be replaced with something else.

Figure 1 plots twice the log-likelihood as a function of ν for this model and indicates a 95% confidence interval for ν based on classical χ^2 limiting theory. This interval falls strictly below one, so the Cauchit model is rejected at this conventional level of significance in favor of even heavier tailed alternative links. In light of this evidence one can, and should, still ask the question: How different are the predicted probabilities from the estimated Gosset model when compared to those from the probit specification. Figure 2 shows a PP plot comparing these fitted probabilities evaluated at the sample observations. Clearly, the two models deliver dramatically different estimates of the quit probabilities.

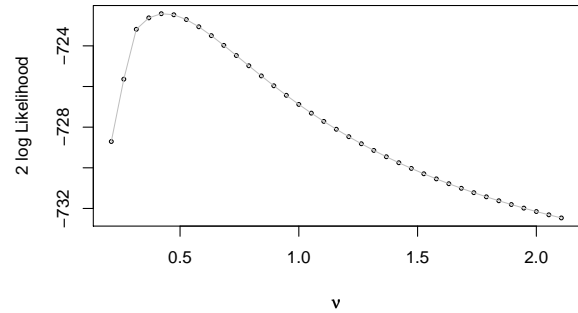


Figure 1: Profile likelihood for the Gosset link parameter ν for a model of quit behavior. The vertical lines indicate a 95% confidence interval for ν .

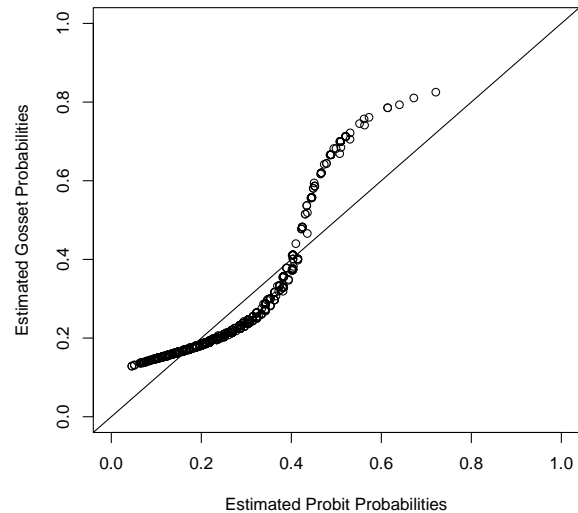


Figure 2: PP Plot of Fitted Probabilities of the Probit and Gosset Models for the WECO data: The solid line is the 45 degree line.

The Pregibon link poses somewhat more of a challenge for fitting our quit model. In Figure 3 we plot likelihood contours for the $\theta = (a, b)$ parameters of the Pregibon link. The contours are labeled according to the asymptotic relation

$$2(\ell(\hat{\theta}) - \ell(\theta)) \sim \chi_2^2,$$

where $\hat{\theta}$ is the maximum likelihood estimate. Thus, if $dAIC$ is a difference in AIC values at the points θ and $\hat{\theta}$ then $pchisq(dAIC, 2)$ is the asymptotic p -value of a test, and confidence regions can be labeled accordingly. A formal test of the logistic hypothesis against Pregibon alternatives gives a test statistic of 14.85 with an asymptotic p -value of 0.0006, strongly repudiating the logit specification. The maximum-

mum likelihood estimates from the Pregibon model are $(\hat{\alpha}, \hat{\delta}) = (-3.58, 0.57)$ suggesting a much longer tailed and somewhat skewed innovation density relative to the logistic. Figure 4 plots the fitted density.

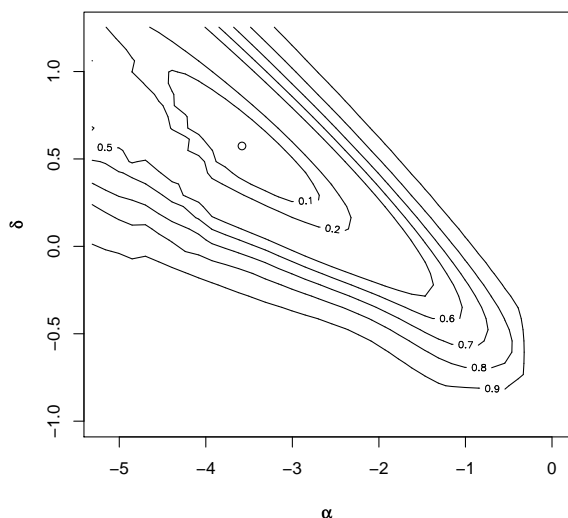


Figure 3: Profile likelihood contours for the Pregibon link parameters for a model of quit behavior. Contours are labeled according to asymptotic probability content.

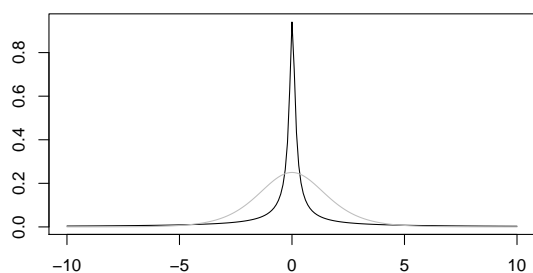


Figure 4: Fitted Pregibon innovation density for the quit model compared to the standardized logistic density (in grey).

Evaluation of the likelihood in the region depicted in the figure requires some care to assure con-

vergence of the `glm` algorithm. This is facilitated by specifying reasonable initial values of the linear predictor coefficients using `start` argument. We begin the evaluation of the likelihood on the grid required for the contour plot with a central point of the grid near the maximum likelihood estimate. It is prudent to organize the remaining evaluations by moving away from the central grid point in a rectangular spiral, updating the `start` argument as one travels around the grid.

The code used to produce the figures is available at: <http://www.econ.uiuc.edu/~roger/research/links/links.html>

Conclusions

Given the opportunity to estimate `glm` models with a wider class of parametric link functions it is obviously attractive to embed such estimation into R optimization of profiled likelihoods. This step, implemented in a somewhat perfunctory fashion in the code referred to above, finally answers my students' annual query: "What do we do if the Pregibon goodness-of-link test rejects?"

Acknowledgments

I would like to thank John Fox and Jungmo Yoon for comments that led to substantial improvements in the exposition, and I would like to thank Brian Ripley for suggestions that led to a dramatic simplification of the code.

Bibliography

- R. King. *gld: Basic functions for the generalised (Tukey) lambda distribution, R package version 1.7.2*, 2005. <http://www.r-project.org>.
- R. Klein, R. Spady, and A. Weiss. Factors affecting the output and quit propensities of production workers. *The Review of Economic Studies*, 58:929–954, 1991.
- B. J. T. Morgan and D. M. Smith. A note on Wadley's problem with overdispersion. *Applied Statistics*, 41: 349–354, 1992.
- D. Pregibon. Goodness of link tests for generalized linear models. *Applied Statistics*, 29:15–24, 1980.