Parametric Links for Binary Response

Roger Koenker and Jungmo Yoon

University of Illinois, Urbana-Champaign

UseR! 2006

Abstract

There is more to life than logit and probit.

Some Preliminary Market Research: A Googoloscopy

Link	GoogleHits
Logit	2,800,000
Probit	1,900,000
Cloglog	1,700
Cauchit	433

Some Preliminary Market Research: A Googoloscopy

Link	GoogleHits
Logit	2,800,000
Probit	1,900,000
Cloglog	1,700
Cauchit	433

A Meta-Analysis Proposal:

- Factors determining the use of Logit vs. Probit in binary response applications.

Some Preliminary Market Research: A Googoloscopy

Link	GoogleHits
Logit	2,800,000
Probit	1,900,000
Cloglog	1,700
Cauchit	433

A Meta-Analysis Proposal:

- Factors determining the use of Logit vs. Probit in binary response applications.
- Should we use logit or probit for the analysis?

Cauchit?

As in the Cauchy distribution, also known as the Witch of Agnesi:
Available in R since 2.1.0.

Cauchit?

As in the Cauchy distribution, also known as the Witch of Agnesi:
Available in R since 2.1.0.


Not to be confused with. . . .


Cauchit?

As in the Cauchy distribution, also known as the Witch of Agnesi:
Available in R since 2.1.0.


Not to be confused with. . . .


Cauchit is much more tolerant of a few surprising observations than is either logit or probit.

Why Do We Need Parametric Links?

The three canonical human motivations:

- Guilt: For 20 years I've been teaching Daryl Pregibon's (1980) paper "A Goodness of Link Test"

Why Do We Need Parametric Links?

The three canonical human motivations:

- Guilt: For 20 years I've been teaching Daryl Pregibon's (1980) paper "A Goodness of Link Test" – but I could never answer the obvious question: "What should we do if we reject the logistic specification?"

Why Do We Need Parametric Links?

The three canonical human motivations:

- Guilt: For 20 years I've been teaching Daryl Pregibon's (1980) paper "A Goodness of Link Test" – but I could never answer the obvious question: "What should we do if we reject the logistic specification?"
- Boredom: There must be more to life than probit or logit.

Why Do We Need Parametric Links?

The three canonical human motivations:

- Guilt: For 20 years I've been teaching Daryl Pregibon's (1980) paper "A Goodness of Link Test" – but I could never answer the obvious question: "What should we do if we reject the logistic specification?"
- Boredom: There must be more to life than probit or logit.
- Fear: Maybe we are all missing something interesting that could be revealed by more general link functions.

What is a Link Function?

Latent variable model for binary response,

$$y_i^* = x_i^\top \beta + u_i, \quad u_i \sim \text{iidF}$$

What is a Link Function?

Latent variable model for binary response,

$$y_i^* = x_i^\top \beta + u_i, \quad u_i \sim \text{iidF}$$

Observed response is:

$$y_i = \{y_i^* \geqslant 0\} = \{u_i \geqslant -x_i^\top \beta\}$$

What is a Link Function?

Latent variable model for binary response,

$$y_i^* = x_i^\top \beta + u_i, \quad u_i \sim \text{iidF}$$

Observed response is:

$$y_i = \{y_i^* \geqslant 0\} = \{u_i \geqslant -x_i^\top \beta\}$$

Probability of the event is:

$$P\{y_i = 1\} = 1 - F(-x_i^\top \beta) \equiv \pi$$

What is a Link Function?

Latent variable model for binary response,

$$y_i^* = x_i^\top \beta + u_i, \quad u_i \sim \text{iidF}$$

Observed response is:

$$y_i = \{y_i^* \geqslant 0\} = \{u_i \geqslant -x_i^\top \beta\}$$

Probability of the event is:

$$P\{y_i = 1\} = 1 - F(-x_i^\top \beta) \equiv \pi$$

Link function is just the quantile function of the error distribution,

$$g(\pi) = -F^{-1}(1 - \pi) = x_i^\top \beta$$

Two Parametric Families of Link Functions

- Gosset: The Student t family with degrees of freedom ν provides a convenient nesting of probit and Cauchit.

Two Parametric Families of Link Functions

- Gosset: The Student t family with degrees of freedom ν provides a convenient nesting of probit and Cauchit.
- Pregibon: The (generalized) Tukey λ family

$$g(\pi) = \frac{\pi^{\alpha+\delta}}{\alpha+\delta} - \frac{(1-\pi)^{\alpha-\delta}}{\alpha-\delta}$$

provides a nice nesting of logit: $(\alpha, \delta) = (0, 0)$, the parameters α and δ can be interpreted as kurtosis and skewness, respectively.

The Pregibon Family

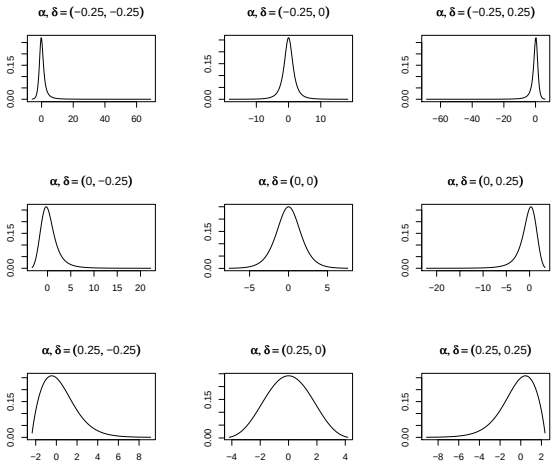

Figure: Pregibon Densities for various (α, δ)'s. All densities scaled to have the same interquartile range.

Implementation in R

- Crucial Change is to permit "..." in `glm` families:

```
family = binomial('Gosset', ...)
```


Implementation in R

- Crucial Change is to permit "..." in `glm` families:

```
family = binomial('Gosset', ...)
```

- Provide p-d-q functions for the new link.
  - Thanks to Luke Tierney for a R-devel suggestion to expand the range of `qt()`.
  - Thanks to Robert King for the gld package for the generalized Tukey λ family.

Implementation in R

- Crucial Change is to permit "..." in `glm` families:

```
family = binomial('Gosset', ...)
```

- Provide p-d-q functions for the new link.
  - Thanks to Luke Tierney for a R-devel suggestion to expand the range of `qt()`.
  - Thanks to Robert King for the gld package for the generalized Tukey λ family.
- Choose optimizer for the profiled likelihood:
  - Gosset: `optimize()` for $\nu \in (0.15, 30)$
  - Pregibon: `optim()` for $(\alpha, \delta) \in [-0.5, 0.5]^2$

Implementation in R

- Crucial Change is to permit "..." in `glm` families:

```
family = binomial('Gosset', ...)
```

- Provide p-d-q functions for the new link.
  - Thanks to Luke Tierney for a R-devel suggestion to expand the range of `qt()`.
  - Thanks to Robert King for the gld package for the generalized Tukey λ family.
- Choose optimizer for the profiled likelihood:
  - Gosset: `optimize()` for $\nu \in (0.15, 30)$
  - Pregibon: `optim()` for $(\alpha, \delta) \in [-0.5, 0.5]^2$
- Plea to R-core: Quite minor changes in `glm()` and friends would be sufficient to allow users to (more easily) "roll their own links."

Koenker and Yoon (UIUC) Parametric Links UseR! 2006 8 / 14

Performance of the Gosset Link

A model of job tenure at Western Electric (R.I.P.), the probability π_i of quiting within 6 months of initial employment is given by,

$$g_\nu(\pi_i) = \beta_0 + \beta_1 SEX_i + \beta_2 DEX_i + \beta_3 LEX_i + \beta_4 LEX_i^2$$

Performance of the Gosset Link

A model of job tenure at Western Electric (R.I.P.), the probability π_i of quiting within 6 months of initial employment is given by,

$$g_\nu(\pi_i) = \beta_0 + \beta_1 SEX_i + \beta_2 DEX_i + \beta_3 LEX_i + \beta_4 LEX_i^2$$

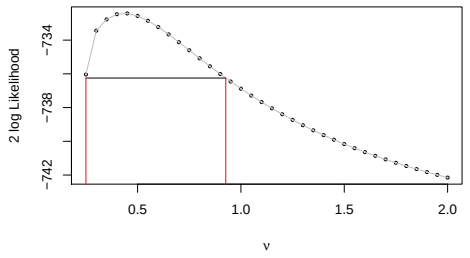

Figure: Profile likelihood for the Gosset link parameter ν

Koenker and Yoon (UIUC) Parametric Links UseR! 2006 9 / 14

Does the Link Really Matter?

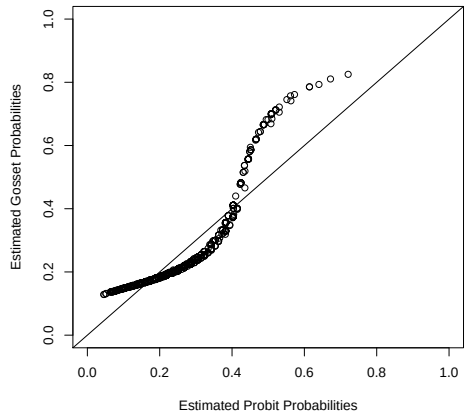

Figure: PP Plot of Fitted Probabilities: Probit vs MLE Gosset Models

Can We Distinguish Gosset Links?

Frequency	$n = 500$			$n = 1000$		
	$\nu_0 = 1$	$\nu_0 = 2$	$\nu_0 = 6$	$\nu_0 = 1$	$\nu_0 = 2$	$\nu_0 = 6$
$H_0 : \nu_0 = 1$	0.062	0.530	0.988	0.056	0.842	1.000
$H_0 : \nu_0 = 2$	0.458	0.056	0.516	0.776	0.070	0.808
$H_0 : \nu_0 = 6$	0.930	0.522	0.010	1.000	0.814	0.042

Table: Rejection frequencies of the likelihood ratio test. Column entries represent fixed values of the true ν parameter, while row entries represent fixed values of the hypothesized parameter. Thus, diagonal table entries indicate size of the test, off-diagonal entries report power.Results are based on 500 replications for each sample size.

A More Direct Measure of Performance?

$$d_p(\hat{F}, F) = \left(\int |\hat{F}(x^\top \hat{\beta}) - F(x^\top \beta)|^p \, dG(x)\right)^{1/p}$$

Estimator	d_1			d_2			d_∞		
	$\nu = 1$	$\nu = 2$	$\nu = 6$	$\nu = 1$	$\nu = 2$	$\nu = 6$	$\nu = 1$	$\nu = 2$	$\nu = 6$
Probit	0.065	0.038	0.013	0.133	0.119	0.092	0.186	0.171	0.136
Cauchit	0.016	0.024	0.033	0.022	0.034	0.048	0.055	0.107	0.167
MLE	0.020	0.016	0.012	0.027	0.024	0.021	0.070	0.065	0.058
Bayes	0.020	0.018	0.013	0.028	0.027	0.024	0.071	0.077	0.069

Table: Performance of Several Binary Response Estimators: The Gosset MLE and Bayes (posterior coordinatewise median) perform well in all three settings.

Pregibon Link?

Pregibon link is computationally more challenging than the Gosset link:

- But profile likelihood is still well-behaved,
- GLM method of scoring with step halving works well,
- Standardizing the interquartile range is helpful,
- Complements influence robust methods in `glmrob`,
- Bayesian MCMC offers a complementary approach to MLE,
- More details, simulation results, etc available from

`/http://www.econ.uiuc.edu/~roger`

Binary Response

- Can be more than a choice between probit and logit.
- One, two, many links!