Universität Innsbruck

Beyond Binary GLMs: Exploring Heteroskedastic Probit and Parametric Link Models

Achim Zeileis

http://eeecon.uibk.ac.at/~zeileis/

Overview

- Generalized linear models (GLMs) with extra parameters
- R package *glmx*
- Heteroskedastic probit models (and “heteroskedastic” GLMs)
- (Binary) GLMs with parametric links
- Discussion
  - Nonlinearity
  - Identifiability
  - Parameter recovery

Generalized linear models

The conditional distribution of the dependent variable is from an exponential family (including Gaussian, binomial, Poisson, gamma, ...).

$$y_i \mid x_i \;\sim\; F(\mu_i, \phi).$$

Conditional expectation μ_i depends on linear predictor through known link function $g(\cdot)$:

$$g(\mu_i) \;=\; x_i^\top \beta.$$

Dispersion parameter ϕ may be known and fixed (e.g., for binomial or Poisson) or treated as a nuisance parameter.

GLMs with extra parameters

Possible extensions:

- Distribution $F(\mu_i, \phi, \theta)$ may have additional parameters θ but is exponential family for given θ (e.g., negative binomial).
- The link function $g(\cdot, \theta)$ only known up to parameter θ, typically with standard link functions as special cases (e.g., Gossett, Pregibon, ...).
- Additional "heteroskedasticity" by scaling the linear predictor with σ_i (e.g., heteroskedastic probit).

R package *glmx*

- Testbed for GLMs with extra parameters. Development on R-Forge. First CRAN release this week.
- Joint work with Roger Koenker and Philipp Doebler.
- Function `glmx()` for parametric `family` objects.
  - Optimization via `optim(method = "BFGS")`.
  - By default using profile likelihood.
  - Analytical gradients may be supplied optionally.
  - Several S3 methods available but more desirable.
- Function `hetglm()` for heteroskedastic GLMs.
  - Optimization via `nlminb()` (default) or `optim()`.
  - Both analytical gradients and analytical expected Hessian available.
  - Large set of S3 methods available.
- Various new parametric link generators: Gossett, Pregibon, Aranda-Ordaz, Guerrero-Johnson, Rocke, folded exponential, t_α, zero-censored negative binomial, …

Heteroskedastic probit models

Standard probit motivation: Latent variable $y_i^* = x_i^\top \beta + \varepsilon_i$ that captures "propensity" for "success" but only $y_i = I(y_i^* > 0)$ is observed.

For Gaussian errors $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$:

$$\begin{aligned}
\text{Prob}(y_i = 1 \mid x_i) &= \text{Prob}(y_i^* > 0 \mid x_i) \\
&= \Phi\left(\frac{x_i^\top \beta - 0}{\sigma}\right)
\end{aligned}$$

Scale parameter σ not identified and hence usually $\sigma = 1$.

Extension: Scale itself is not identified but scale *differences* via

$$\log(\sigma_i) = z_i^\top \gamma$$

where z_i must not include a constant term for identifiability.

Interpretation: "Ambivalence" in discrete choices.

Heteroskedastic probit models

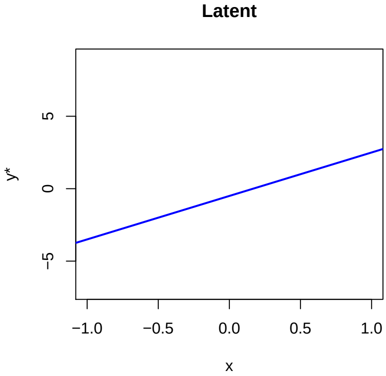
Questions:

- Heteroskedasticity vs. nonlinearity.
- Heteroskedasticity vs. interaction effects.
- Identification.
- Parameter recovery.

Heteroskedastic probit models

Example: Nonlinearity for $x = z \in [-1, 1]$ with

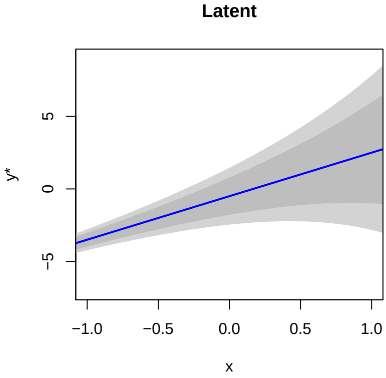
$$\Phi^{-1}(\mu) \;=\; \frac{-0.5 + 3 \cdot x}{\exp(1 \cdot x)}.$$



Heteroskedastic probit models

Example: Nonlinearity for $x = z \in [-1, 1]$ with

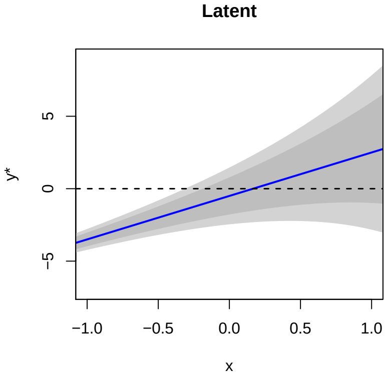
$$\Phi^{-1}(\mu) \; = \; \frac{-0.5 + 3 \cdot x}{\exp(1 \cdot x)}.$$



Heteroskedastic probit models

Example: Nonlinearity for $x = z \in [-1, 1]$ with

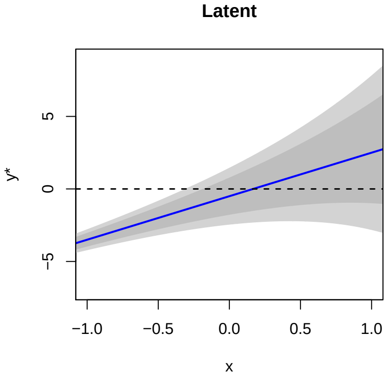
$$\Phi^{-1}(\mu) \;=\; \frac{-0.5 + 3 \cdot x}{\exp(1 \cdot x)}.$$

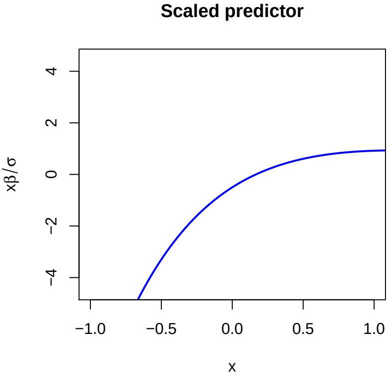


Heteroskedastic probit models

Example: Nonlinearity for $x = z \in [-1, 1]$ with

$$\Phi^{-1}(\mu) \;=\; \frac{-0.5 + 3 \cdot x}{\exp(1 \cdot x)}.$$

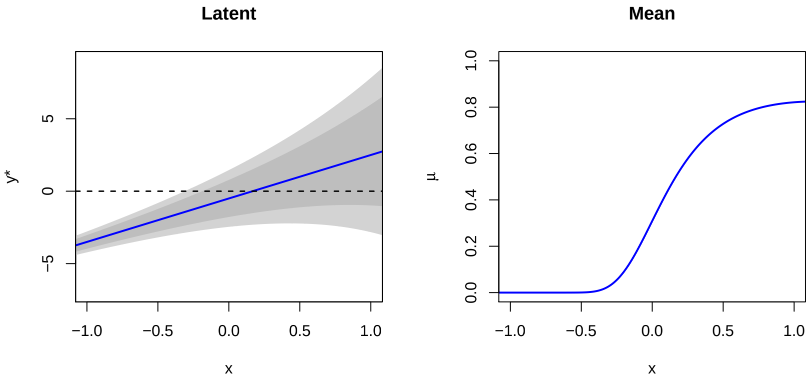




Heteroskedastic probit models

Example: Nonlinearity for $x = z \in [-1, 1]$ with

$$\Phi^{-1}(\mu) \;=\; \frac{-0.5 + 3 \cdot x}{\exp(1 \cdot x)}.$$

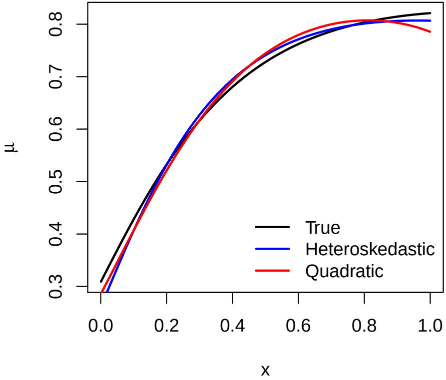


Heteroskedastic probit models

Hence: Hard to distinguish heteroskedasticity from nonlinear mean.

Example: Heteroskedastic probit model with $\beta = (-0.5, 3)^\top$, $\gamma = 1$, and $x = z \in [0, 1]$ for $n = 1000$ observations.

Models: Correctly specified model vs. probit with quadratic trend.



Heteroskedastic probit models

Example: Interactions with a numeric regressor x and two groups $z \in \{0, 1\}$.

First group ($z = 0$):

$$\Phi^{-1}(\mu) \;=\; \frac{\beta_1 + \beta_2 \cdot x}{\exp(0)} \;=\; \beta_1 + \beta_2 \cdot x$$

Second group ($z = 1$):

$$\Phi^{-1}(\mu) \;=\; \frac{\beta_1 + \beta_2 \cdot x}{\exp(\gamma)} \;=\; \tilde{\beta}_1 + \tilde{\beta}_2 \cdot x$$

Interpretation: Groupwise intercepts and slopes but with proportionality constraint,

$$\beta_1/\tilde{\beta}_1 \;=\; \beta_2/\tilde{\beta}_2 \;=\; \exp(\gamma).$$

Heteroskedastic probit models

Example: Lack of identification for two groups with $x = z \in \{0, 1\}$.

First group ($x = z = 0$):

$$\Phi^{-1}(\mu) \ = \ \frac{\beta_1 + \beta_2 \cdot 0}{\exp(0)} \ = \ \beta_1$$

Second group ($x = z = 1$):

$$\Phi^{-1}(\mu) \ = \ \frac{\beta_1 + \beta_2 \cdot 1}{\exp(\gamma)} \ = \ \tilde{\beta}_1$$

Interpretation: Three parameters but only two identified, i.e., grouping in latent mean must not coincide with grouping in latent scale.

Heteroskedastic probit models

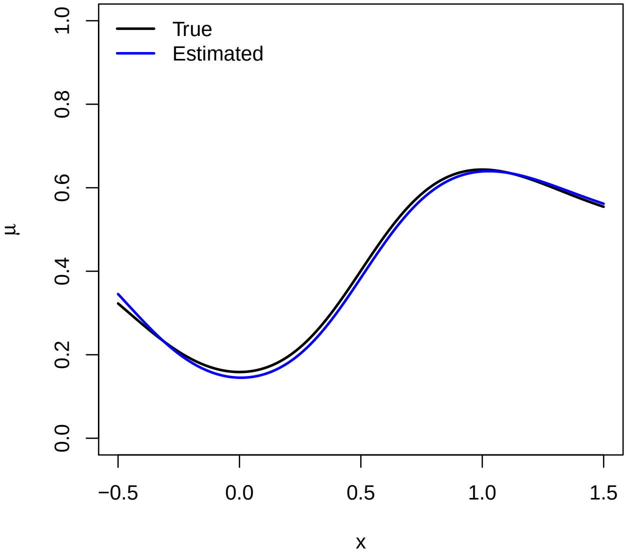
Parameter recovery:

- Estimates can be very biased even in moderately large samples.
- Parameters are often only “weakly” identified.
- Mean function can typically be recovered much better.

Example: Quadratic polynomial for both $x = z \in [-0.5, 1.5]$ and $\beta = (-1, 1, 1)^\top$, $\gamma = (-1, 2)^\top$, $n = 1000$.

Parameter		True	Estimate	Lower	Upper
β	Intercept	-1.00	-1.06	-1.26	-0.86
	x	1.00	0.38	-1.77	2.52
	x^2	1.00	2.00	-1.56	5.55
γ	x	-1.00	-0.41	-3.05	2.24
	x^2	2.00	1.71	0.08	3.35

Heteroskedastic probit models



Heteroskedastic GLMs

More generally: GLMs with heteroskedastic linear predictor scale.

$$
\begin{aligned}
y_i \mid x_i &\sim F(\mu_i, \phi) \\
g(\mu_i) &= \frac{x_i^\top \beta}{\sigma_i} \\
h(\sigma_i) &= h(1) + z_i^\top \gamma
\end{aligned}
$$

where

- Distribution $F(\mu, \phi)$ and link function $g(\cdot)$ are "as usual".
- $h(\cdot)$ is an additional link function, e.g., log, square root, identity.
- $h(1)$ assures $\sigma_i = 1$ if $z_i^\top \gamma = 0$.

(Binary) GLMs with parametric links

Idea: Additional flexibility by extra parameter(s) in the link function.

$$g(\mu_i, \theta) \; = \; x_i^\top \beta$$

Example: Gossett link for binary responses.

- $g(\cdot, \theta)$ is the quantile function of the Student t distribution with θ degrees of freedom.
- Contains the probit link ($\theta = \infty$) and the cauchit link ($\theta = 1$) as special cases.
- Can be employed for "goodness-of-link" tests.
- Numerically challenging for $\theta < 0.5$.

Similarly: Other families of links that generalize/nest standard link functions.

(Binary) GLMs with parametric links

Properties: Similar as for heteroskedastic GLMs.

- Additional model flexibility.
- Tension between using extra parameters θ vs. adding regressors/interactions in $x_i^\top \beta$.
- Parameter recovery may be difficult even for moderately large samples.

Summary

- Binary choice models with heteroskedasticity or parametric links can now easily be estimated in R using the *glmx* package.
- These models can capture nonlinearity in different (and sometimes more parsimonious) ways than nonlinear terms in the predictor.
- Hard to distinguish from other forms of nonlinearity.
- Parameters are harder to recover than the conditional mean function.

References

Zeileis A, Koenker R, Doebler P (2013). *glmx: Generalized Linear Models Extended.* R package version 0.1-0.
URL `http://CRAN.R-project.org/package=glmx`

Koenker R, Yoon J (2009). "Parametric Links for Binary Choice Models: A Fisherian-Bayesian Colloquy." *Journal of Econometrics*, **152**, 120–130.

Alvarez RM, Brehm J (1995). "American Ambivalence towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values." *American Journal of Political Science*, **39**(4), 1055–1082.

Keele LJ, Park DK (2006). *Ambivalent about Ambivalence: A Re-Examination of Heteroskedastic Probit Models.* Unpublished manuscript. `http://www.polisci.ohio-state.edu/faculty/lkeele/hetprob.pdf`