National Institute of Economic and Social Research

A Simple Approach to Compute Generalized Residuals for Nonlinear Models

Arnab Bhattacharjee and Tibor Szendrei

Presented by: Arnab Bhattacharjee

September 2025

31st UK Stata Conference, London

Table of contents

Construction of generalized residuals

Generalised Linear Models

- Model: $y_i = f(x_i, \beta) + \varepsilon_i$
- Residuals: $y_i - \hat{y}_i(\hat{\beta})$
- Usually scale is not identified, but
  - Distbn of errors ε replicated under usual RE assumptions
  - Sometimes even FE/rdm coeffs, e.g., Pesaran (2006)
- But no clear extension for nonlinear models, where error ε and y are not linearly related
- Useful in several contexts, e.g.
  - quantile regression and evaluation of conditional quantiles at the tails (for example, growth at risk or welfare policy);
  - computing errors distributions (for example, binary regression and random effects models); and
  - computing network externalities in discrete choice and duration models.

- Log-likelihood

$$\ln L(\beta,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y_i - x_i'\beta)'(y_i - x_i'\beta)$$
$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\varepsilon_i'\varepsilon_i$$

- Partial derivatives wrt y and x

$$\frac{\partial \ln L_i}{\partial y_i} = \frac{1}{\sigma^2}(y_i - x_i'\beta) = \frac{1}{\sigma^2}\varepsilon_i$$

- Neat … but does not quite work with limited dependent y. So:

$$\frac{\partial \ln L_i}{\partial x_i} = \begin{pmatrix} \frac{\partial \ln L_i}{\partial x_{1i}} & \frac{\partial \ln L_i}{\partial x_{2i}} & \dots & \frac{\partial \ln L_i}{\partial x_{Ki}} \end{pmatrix} = \frac{\partial \ln L_i}{\partial y_i}\frac{\partial y_i}{\partial x_i} = \frac{1}{\sigma^2}\varepsilon_i\hat{\beta}$$
$$\Rightarrow \varepsilon_i = \frac{\sigma^2}{\hat{\beta}'\hat{\beta}}\left[\frac{\partial \ln L_i}{\partial x_i}\right]'\hat{\beta}$$

- Evaluated at $\hat{\beta}$, product of score fn and influence fn. – Neat
- Note: Scale is not identified – just the shape

Binary response – Logit model

- Here, there is a natural construction of residuals, because

$$\mathbb{E}(y_i|x_i) = \mathbb{P}(y_i = 1|x_i)$$

- Hence, LPM is applicable simply regressing binary y on x. However, several issues
  - Heteroscedasticity, plus limited dependent nature of DGP not accounted for
  - But further, distbn of latent error is identified
  - Can the new "generalized residuals" capture the error distbn
- Logit likelihood is relatively easy to work with

$$\frac{\partial \ln L_i}{\partial x_i} = \frac{\partial}{\partial x_i} \{y_i \ln [\Lambda(x_i'\beta)] + (1 - y_i) \ln [1 - \Lambda(x_i'\beta)]\}$$
$$= \sum_{k=1}^{K} \{y_i - \Lambda(x_i'\beta)\} \beta_k = \sum_{k=1}^{K} \{y_i - \Lambda(y_i^\star - \varepsilon_i)\} \beta_k,$$

where $y_i^\star$ is the (unobserved) latent variable and $\Lambda(.)$ denotes the cdf of the standard logistic distbn

Looks somewhat complicated, but really simple to code … in Stata or otherwise

```
logit y x
gen res = e(ll)
        local j = 1
        while `j' <= _N {
                qui replace x = x+1 if id==`j'
                qui logit y x
                qui replace res_`i' = (e(ll) - res_`i')/e(b)[1,1] if id==`j'
                qui replace x = x-1 if id==`j'
                        qui local j = `j' + 1
        }
egen float genres = std(res), mean(0) sd(1)
```

Figure 1: Prototype code

Can be parallelised

Monte Carlo

KL divergence (OLS)

- Really a bit of a sense check first. All models are estimated by OLS, so gen-res coincides with usual residuals
- Model: $y = 1 + x + u, \quad x \sim N(0,1) \quad u \sim \quad$ various
- But we evaluate the impact of theoretical scaling by variance against our proposed standardised gen-res.
- Generalised residuals match true error distbn better in all cases
  - particularly as the signal to noise ratio declines
  - and error distbn deviate from Gaussian

	Error Distribution				
	Norm1	Norm1.5	Norm2	Expo2	t(3.6)
Residuals (Scaled)	0.0176	0.0930	0.2410	0.0438	0.0175
Generalised Residuals	0.0130	0.0145	0.0133	0.0101	0.0107

Table 1: KL Divergence from True DGP by Error Distribution (Average of 20 MCs, each with $n = 100$)

Behaviour of Generalised residuals

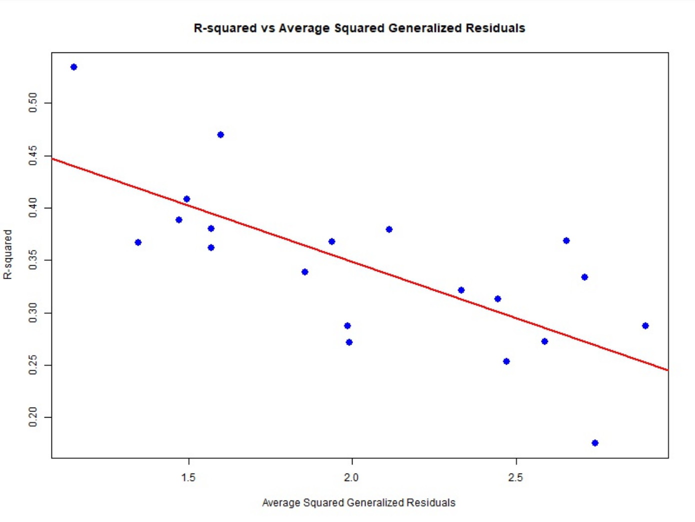

Figure 2: t-dist. error model fit vs average (squared) Generalised resid.

KL divergence (Logit)

- Model: $y = \mathbb{I}(1 + x + u > 0), \quad x \sim N(0, 1)$ as before
- Error distbn: $u \sim$ Logistic(0.55) or $N(0, 1)$, both zero mean unit variance
- Logit regression gen-res compared against Linear Probability Model (OLS) residuals

	Error Distribution	
	Normal	Logistic
LPM Residuals (Scaled)	2.8271	1.2627
(Pearson) Residuals	0.1202	0.0772
Generalised Residuals	0.0803	0.0525

Table 2: KL Divergence from True DGP by Error Distribution (Average of all 20 MCs, each with $n = 100$)

Extensions

- Extensions to other nonlinear models fairly straightforward. For example:
- Multinomial logit, using the Random Utility Model formulation of McFadden (1974)
- Cox (and beyond Cox) hazard regression models using Cox (1975) partial likelihood fn instead of the full log-likelihood
- Other things to do
    - More extensive error distbns
    - Tail probabilities
    - Network externalities
    - Applications

Thank you for your attention!

References

References I