Estimating the Tobit-1 model with the charitable data set

July 9, 2025

We'll reproduce here some results obtained by Wilhelm (2008) using a data set which deals with charitable giving. The `charitable` data set is shipped with the **micsr** package.

```
library(micsr)
```

```
head(charitable, 5)
```

```
  donation donparents        education   religion   income married south
1      335       5210 less_high_school      other 21955.13       0     0
2       75      13225      high_school protestant 22103.82       0     0
3     6150       3375     some_college   catholic 50298.87       0     0
4       25         50     some_college   catholic 28666.14       1     0
5       25         25 less_high_school       none 13670.03       0     1
```

The response is called `donation`, it measures annual charitable givings in \$US. This variable is left-censored for the value of 25, as this value corresponds to the item "less than 25 \$US donation". Therefore, for this value, we have households who didn't make any charitable giving and some which made a small giving (from 1 to 24 \$US).

The covariates used are the donation made by the parents (`donparents`), two factors indicating the educational level and religious beliefs (respectively `education` and `religion`), annual income (`income`) and two dummies for living in the south (`south`) and for married couples (`married`).

Wilhelm (2008) considers the value of the donation in logs and subtract $\ln 25$, so that the response is 0 for households who gave no donation or a small donation.

```
charitable$logdon <- with(charitable, log(donation) - log(25))
```

The tobit model can be estimated by maximum likelihood using `AER::tobit`, `censReg::censReg` or with the `tobit1` package.

```
char_form <- logdon ~ log(donparents) + log(income) +
    education + religion + married + south
if (requireNamespace("AER")){
    library("AER")
    ml_aer <- tobit(char_form, data = charitable)
}
if (requireNamespace("censReg")){
    library("censReg")
    ml_creg <- censReg(char_form, data = charitable)
}
ml <- tobit1(char_form, data = charitable)
```

tobit1 provides a rich set of estimation methods, especially the trimmed or **SCLS** (symmetrically censored least squares) estimator proposed by Powell (1986). We also, for pedagogical purposes, estimate the OLS estimator although it is known to be inconsistent.

```
scls <- update(ml, method = "trimmed")
ols <- update(ml, method = "lm")
```

The results of the three models are presented in Table 1

The results match exactly the first two columns of (Wilhelm 2008, table 3 page 577).

Note that the OLS estimators are all lower in absolute values than those of the two other estimators, which illustrate the fact that OLS estimators are biased toward zero when the response is censored. The maximum likelihood is consistent and asymptotically efficient if the conditional distribution of y^* (the latent variable) is homoskedastic and normal.

Specification tests for the maximum likelihood estimator can be conducted using conditional moments tests. This can easily be done using the `micsr::cmtest` function, which can take as input a model fitted by either `AER::tobit`, `censReg::censReg` or `tobit1::tobit1`:

```
cmtest(ml) |> gaze()
## chisq = 116.351, df: 2, pval = 0.000
```

`cmtest` has a `test` argument with default value equal to `normality`. To get a heteroskedasticity test, we would use:

```
cmtest(ml, test = "heterosc") |> gaze()
## chisq = 103.592, df: 12, pval = 0.000
```

Table 1: Estimation of charitable giving models

	OLS	maximum likehihood	SCLS
(Intercept)	−10.071	−17.618	−15.388
	(31.854)	(0.898)	(20.052)
log(donparents)	0.135	0.200	0.167
	(209.215)	(0.025)	(142.544)
log(income)	0.941	1.453	1.320
	(344.236)	(0.087)	(228.142)
educationhigh_school	0.151	0.622	0.655
	(18.987)	(0.188)	(8.682)
educationsome_college	0.470	1.100	1.042
	(16.887)	(0.194)	(9.887)
educationcollege	0.761	1.325	1.284
	(12.983)	(0.215)	(11.135)
educationpost_college	1.121	1.727	1.588
	(9.873)	(0.236)	(9.319)
religioncatholic	0.298	0.639	0.433
	(15.118)	(0.171)	(8.917)
religionprotestant	0.731	1.257	0.983
	(22.354)	(0.154)	(14.130)
religionjewish	0.629	1.001	0.768
	(5.300)	(0.307)	(6.648)
religionother	0.430	0.837	0.596
	(11.014)	(0.194)	(6.085)
married	0.562	0.767	0.702
	(25.393)	(0.117)	(18.411)
south	0.111	0.113	0.064
	(17.615)	(0.105)	(9.483)
sigma		2.114	
		(0.041)	
Num.Obs.	2384	2384	2384
AIC		8038.5	
BIC		8119.4	
Log.Lik.		−4005.274	

Normality and heterosledasticity are strongly rejected. The values are different from Wilhelm (2008) as he used the "outer product of the gradient" form of the test. These versions of the test can be obtained by setting the `OPG` argument to `TRUE`.

```
cmtest(ml, test = "normality", opg = TRUE) |> gaze()
## chisq = 200.117, df: 2, pval = 0.000
cmtest(ml, test = "heterosc", opg = TRUE) |> gaze()
## chisq = 127.308, df: 12, pval = 0.000
```

Non-normality can be further investigate by testing separately the fact that the skewness and kurtosis indicators are respectively different from 0 and 3.

```
cmtest(ml, test = "skewness") |> gaze()
## z = 10.393, pval = 0.000
cmtest(ml, test = "kurtosis") |> gaze()
## z = 2.329, pval = 0.020
```

The hypothesis that the conditional distribution of the response is mesokurtic is not rejected at the 1% level and the main problem seems to be the asymmetry of the distribution, even after taking the logarithm of the response.

This can be illustrated (see Figure 1) by plotting the (unconditional) distribution of the response (for positive values) and adding to the histogram the normal density curve.

References

Powell, J. 1986. "Symmetrically Trimed Least Squares Estimators for Tobit Models." *Econometrica* 54: 1435–60.

Wilhelm, Mark Ottoni. 2008. "Practical Considerations for Choosing Between Tobit and SCLS or CLAD Estimators for Censored Regression Models with an Application to Charitable Giving." *Oxford Bulletin of Economics and Statistics* 70 (4): 559–82.

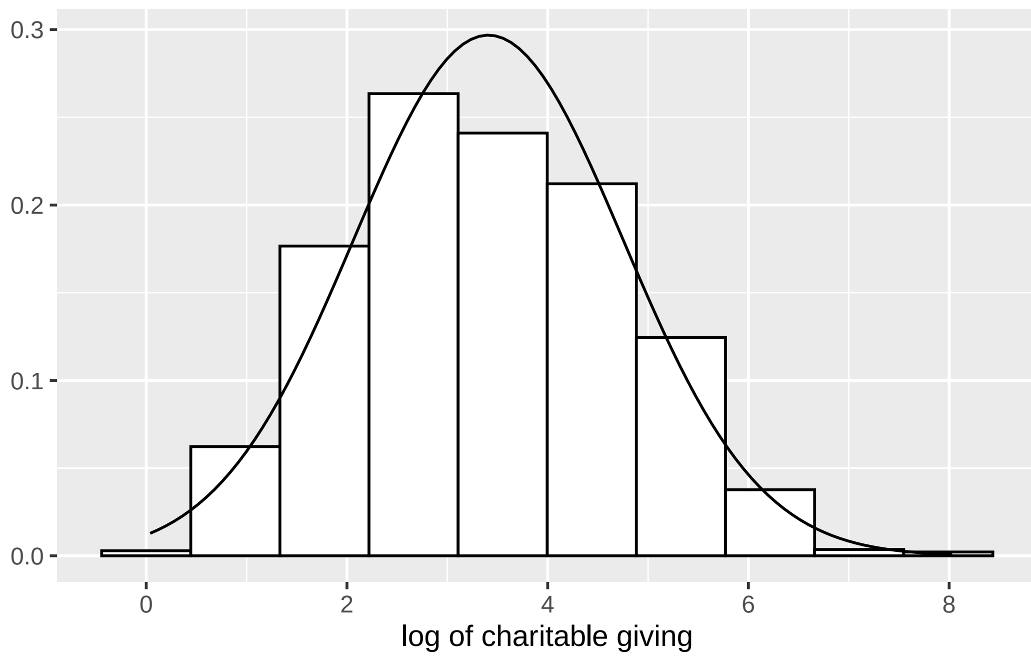

Figure 1: Empirical distribution of the response and normal approximation