Estimating the Tobit-1 model with the charitable data set

June 8, 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 shiped with the `micsr`.

```
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) consider the value of the donation in logs and substract $\ln 25$, 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 **SCLS** (symetrically censored least squares) estimator proposed by Powell (1986). We also, for pedagogical purposes, estimate the OLS estimator although it is known to be unconsistent.

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

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 asymtotically efficient if the conditional distribution of y^* (the latent variable) is homoscedastic 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)
```

```
    Conditional Expectation Test for Normality

data:  logdon ~ log(donparents) + log(income) + education + religion +  ...
chisq = 116.35, df = 2, p-value < 2.2e-16
```

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

```
cmtest(ml, test = "heterosc")
```

      Heteroscedasticity Test

data: logdon ~ log(donparents) + log(income) + education + religion + ...
chisq = 103.59, df = 12, p-value < 2.2e-16

Normality and heteroscedasticity 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)
```

      Conditional Expectation Test for Normality

data: logdon ~ log(donparents) + log(income) + education + religion + ...
chisq = 200.12, df = 2, p-value < 2.2e-16

```
cmtest(ml, test = "heterosc", opg = TRUE)
```

      Heteroscedasticity Test

data: logdon ~ log(donparents) + log(income) + education + religion + ...
chisq = 127.31, df = 12, p-value < 2.2e-16

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")
```

      Conditional Expectation Test for Skewness

data: logdon ~ log(donparents) + log(income) + education + religion + ...
z = 10.393, p-value < 2.2e-16

```
cmtest(ml, test = "kurtosis")
```

```
	Conditional Expectation Test for Kurtosis

data:  logdon ~ log(donparents) + log(income) + education + religion +  ...
z = 2.3294, p-value = 0.01984
```

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 asymetry 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.
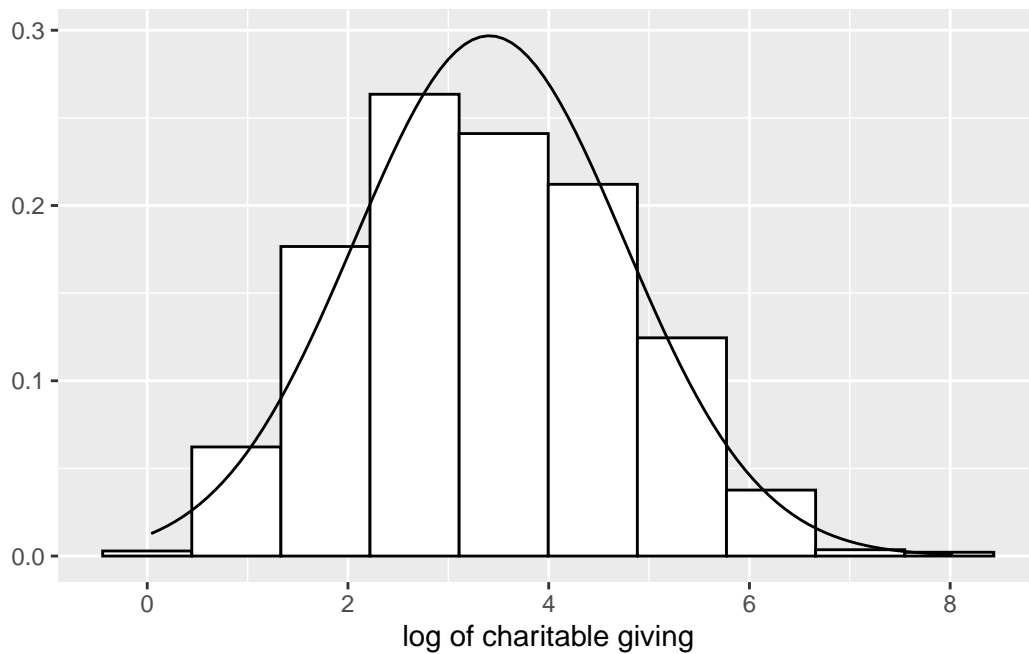

Figure 1: Empirical distribution of the response and normal approximation

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.