Using the tobit1 package with the charitable data set

We'll reproduce here some results obtained by Wilhelm (2008) using a data set which deals with charitable giving. The data set is shiped with the `tobit1` package and can be accessed as soon as this package is attached.

```
library("tobit1")
library("tidyverse")
```

```
charitable %>% print(n = 5)
```

```
## # A tibble: 2,384 x 7
##   donation donparents education        religion   income married south
##      <dbl>      <dbl> <fct>            <fct>       <dbl>   <dbl> <dbl>
## 1     335        5210 less_high_school other      21955.       0     0
## 2      75       13225 high_school      protestant 22104.       0     0
## 3    6150.       3375 some_college     catholic   50299.       0     0
## 4      25          50 some_college     catholic   28666.       1     0
## 5      25          25 less_high_school none       13670.       0     1
## # ... with 2,379 more rows
```

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 <- charitable %>% mutate(logdon = log(donation) - log(25))
```

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

```
library("AER")
library("censReg")
char_form <- logdon ~ log(donparents) + log(income) +
    education + religion + married + south
ml_aer <- tobit(char_form, data = charitable)
ml_creg <- censReg(char_form, data = charitable)
ml_tb1 <- tobit1(char_form, data = charitable)
```

`tobit1` provide 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_tb1, method = "trimmed")
ols <- update(ml_tb1, method = "lm")
```

	OLS	maximum likehihood	SCLS
(Intercept)	-10.071 $(0.556)^{***}$	-17.618 $(0.898)^{***}$	-15.388 $(1.472)^{***}$
log(donparents)	0.135 $(0.017)^{***}$	0.200 $(0.025)^{***}$	0.167 $(0.035)^{***}$
log(income)	0.941 $(0.056)^{***}$	1.453 $(0.087)^{***}$	1.320 $(0.120)^{***}$
educationhigh_school	0.151 (0.115)	0.622 $(0.188)^{***}$	0.655 (0.815)
educationsome_college	0.470 $(0.121)^{***}$	1.100 $(0.194)^{***}$	1.042 (0.813)
educationcollege	0.761 $(0.138)^{***}$	1.325 $(0.215)^{***}$	1.284 (0.814)
educationpost_college	1.121 $(0.155)^{***}$	1.727 $(0.236)^{***}$	1.588 (0.819)
religioncatholic	0.298 $(0.111)^{**}$	0.639 $(0.171)^{***}$	0.433 (0.236)
religionprotestant	0.731 $(0.098)^{***}$	1.257 $(0.154)^{***}$	0.983 $(0.216)^{***}$
religionjewish	0.629 $(0.214)^{**}$	1.001 $(0.307)^{**}$	0.768 $(0.261)^{**}$
religionother	0.430 $(0.125)^{***}$	0.837 $(0.194)^{***}$	0.596 $(0.264)^{*}$
married	0.562 $(0.079)^{***}$	0.767 $(0.117)^{***}$	0.702 $(0.169)^{***}$
south	0.111 (0.071)	0.113 (0.105)	0.064 (0.130)
sigma		2.114 $(0.041)^{***}$	
logLik		-4005.274	
N	2384	2384	2384
left_cens			828
neg_linpred			58
right_trimmed			296

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Estimation of charitable giving models

The results of the three models are presented in table @ref(tab:models).

```
texreg::texreg(list("OLS" = ols, "maximum likehihood" = ml_tb1, "SCLS" = scls),
               caption = "Estimation of charitable giving models",
               label = "tab:models", single.row = TRUE, digits = 3)
```

The last two columns of table @ref(tab:models) 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. The **SCLS** estimator consistency relies only the hypothesis that the errors are symetrical around 0. However, if they are also normal and homoscedastic, it is less efficient than the maximum likelihood estimator. Therefore, the strong distributional hypothesis of the maximum likelihood estimator can be adressed using a Hausman test:

```
coefs_scls <- coef(scls)[-1]
nms_coefs <- names(coefs_scls)
coefs_ml <- coef(ml_tb1)[nms_coefs]
delta <- coefs <- coefs_scls - coefs_ml
V <- vcov(scls)[nms_coefs, nms_coefs] - vcov(ml_tb1)[nms_coefs, nms_coefs]
stat_hausman <- as.numeric(crossprod(solve(V, delta), delta))
pval_hausman <- pchisq(stat_hausman, df = length(delta), lower.tail = FALSE)
c(stat = stat_hausman, pval = pval_hausman)
```

```
##       stat       pval
## 11.0283057  0.5264945
```

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

```r
library("cmtest")
cmtest(ml_tb1)
```

```
##
##  Conditional Expectation Test for Normality
##
## data:  char_form
## 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:

```r
cmtest(ml_tb1, test = "heterosc")
```

```
##
##  Heteroscedasticity Test
##
## data:  char_form
## 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`.

```r
cmtest(ml_tb1, test = "normality", OPG = TRUE)
```

```
##
##  Conditional Expectation Test for Normality
##
## data:  char_form
## chisq = 200.12, df = 2, p-value < 2.2e-16
```

```r
cmtest(ml_tb1, test = "heterosc", OPG = TRUE)
```

```
##
##  Heteroscedasticity Test
##
## data:  char_form
## 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.

```r
cmtest(ml_tb1, test = "skewness")
```

```
##
##  Conditional Expectation Test for Skewness
##
## data:  char_form
## z = 10.393, p-value < 2.2e-16
```

```r
cmtest(ml_tb1, test = "kurtosis")
```

```
##
##  Conditional Expectation Test for Kurtosis
##
## data:  char_form
## 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@ref(fig:histnorm)) by plotting the (unconditional) distribution of the response (for positive values) and adding to the histogram the normal density curve.

```
moments <- charitable %>% filter(logdon > 0) %>% summarise(mu = mean(logdon), sigma = sd(logdon))
ggplot(filter(charitable, logdon > 0), aes(logdon)) +
    geom_histogram(aes(y = ..density..), color = "black", fill = "white", bins = 10) +
    geom_function(fun = dnorm, args = list(mean = moments$mu, sd = moments$sigma)) +
    labs(x = "log of charitable giving", y = NULL)
```

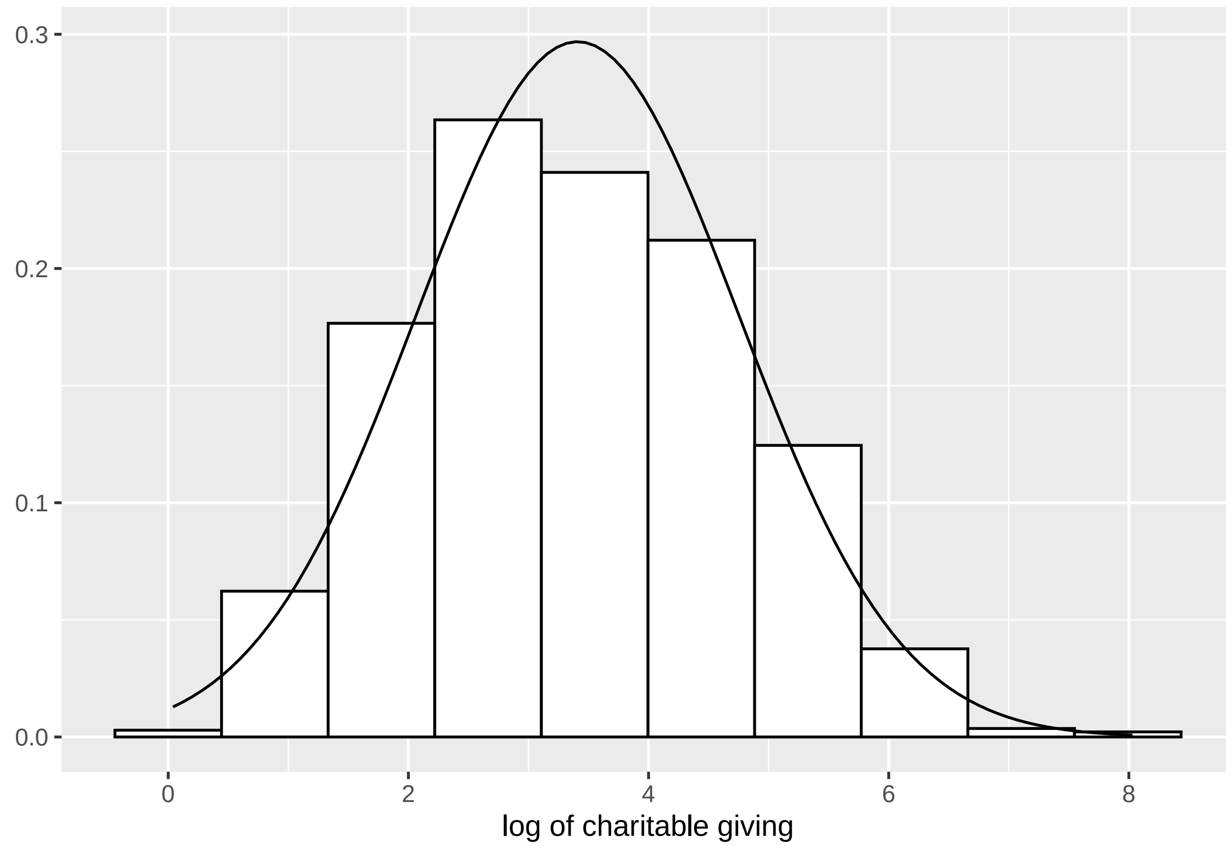

Figure 1: Empirical distribution of the response and normal approximation

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. https://doi.org/https://doi.org/10.1111/j.1468-0084.2008.00506.x.