Applied Econometrics with R

Christian Kleiber
Universität Basel
Switzerland

Achim Zeileis
Wirtschaftsuniversität Wien
Austria

Outline

- R and econometrics

- Robust standard errors

  Example: Sandwich variance estimators for a tobit model

- Gaps

- AER: book and package

---

R and econometrics

- Language and terminology in econometrics is somewhat distinct from the terminology used in mainstream statistics.

  Two examples:

Statistics	Econometrics
factor	dummy variables
generalized linear model	probit, logit, ...

- Generally, not much awareness of statistical GLM literature among econometricians.

- Visualization not very common.

Spinograms and GLMs
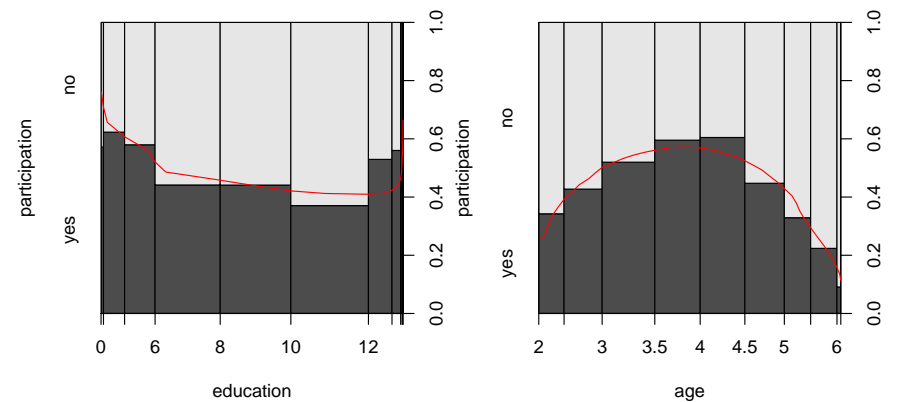
R and econometrics

Traditional econometric software

Applied econometrics:

      EViews, TSP, PcGIVE, SAS, Stata, ...

Theory and methodology:

      GAUSS, Ox, Matlab, S-PLUS, ...

R and econometrics

Why R?

- superior graphics

- object orientation

- reproducibility

R and econometrics

- **Linear models and extensions:** OLS, nonlinear regression, systems of equations

- **Diagnostics and validation:** Robust regression, sandwich covariance matrices, diagnostic tests

- **Microeconometrics:** Logit, Probit, Poisson regression (via `glm()`), Tobit, modified count data models (ZIP, hurdle), duration models (package `survival`)

- **Time series:** (S)ARIMA(X), unit roots and cointegration (packages `tseries`, `urca`), structural change, ARCH models (see `Rmetrics`), structural time series models

See also CRAN task view: econometrics at

    `http://CRAN.R-project.org/src/contrib/Views/Econometrics.html`

Robust standard errors

In the linear regression model

$$y_i = x_i^\top \beta + \varepsilon_i,$$

we have for OLS, under technical assumptions,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2 Q_{XX}^{-1})$$

provided model is **correctly specified**.

If **only conditional mean is correctly specified**, we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\ d\ } \mathcal{N}(0, Q_{XX}^{-1}\Sigma_v Q_{XX}^{-1})$$

a sandwich variance formula. Robustness considerations suggest to estimate the latter.

In econometrics usually called "White standard errors" or "heteroskedasticity-consistent (HC) standard errors".

Depending on the context, this is also known as Eicker-White, Huber-White, Eicker-Huber-White ...

The matrix $\Sigma_v = \text{Cov}(v_i) = \text{Cov}(x_i\varepsilon_i) = \text{E}(\varepsilon_i^2 x_i x_i^\top)$ comes from an estimating equation. Idea generalizes to GLMs and many other models.

R provides infrastructure for HC (and also HAC) covariances in the sandwich package.

Robust standard errors

Example: "Fair's affairs" (Fair, *J. Political Economy* 1978)

- Cross-section data on frequency of extramarital affairs from a survey conducted by *Psychology Today* in 1969.

- n = 601, dependent variable is number of extramarital affairs, covariates are gender, age, years married, children, religiousness, education, occupation, rating of marriage.

- 75.04% of the respondents do not report any extramarital affairs.

Data will be available in package AER.

Robust standard errors

Classical Tobit model (Tobin 1958) is model for left-censored (at zero) data. Standard approach employs Gaussian MLE.

In R, this can be fitted (easily) using survreg() from the survival package, see example("tobin") there.

New function tobit() in package AER provides convenience interface to survreg() (and a bit more).

Task: standard errors under weaker assumptions.

Robust standard errors

```
R> fm_tobit <- tobit(affairs ~ age + yearsmarried + religiousness +
+     occupation + rating, data = FairAffair)
R> coeftest(fm_tobit)

z test of coefficients:

             Estimate Std. Error z value Pr(>|z|)
(Intercept)    8.1742     2.7414    2.98   0.0029
age           -0.1793     0.0791   -2.27   0.0234
yearsmarried   0.5541     0.1345    4.12  3.8e-05
religiousness -1.6862     0.4038   -4.18  3.0e-05
occupation     0.3261     0.2544    1.28   0.2000
rating        -2.2850     0.4078   -5.60  2.1e-08
Log(scale)     2.1099     0.0671   31.44  < 2e-16
```

```
R> coeftest(fm_tobit, vcov = sandwich)
```

```
z test of coefficients:

              Estimate Std. Error z value Pr(>|z|)
(Intercept)     8.1742     3.0779    2.66   0.0079
age            -0.1793     0.0889   -2.02   0.0437
yearsmarried    0.5541     0.1372    4.04  5.3e-05
religiousness  -1.6862     0.3999   -4.22  2.5e-05
occupation      0.3261     0.2460    1.33   0.1850
rating         -2.2850     0.3935   -5.81  6.4e-09
Log(scale)      2.1099     0.0548   38.48  < 2e-16
```

Robust standard errors

Can also do

```
R> linear.hypothesis(fm_tobit, "age = 0", vcov = sandwich)
```

```
Linear hypothesis test

Hypothesis: age = 0

[...]

  Res.Df  Df Chisq Pr(>Chisq)
1    594
2    595  -1  4.07      0.044
```

Gaps

- dynamic regressions (but see packages dyn and dynlm)

- multiple time series models (structural VARs, ...)

- nonlinear time series models (TAR, smooth transition models, ...)

- panel data methods, in particular

  - least-squares methods
  - dynamic models
  - microeconometric models (GLMs with panel data, ...)

- GMM and instrumental variables

- non- and semiparametric regression

AER: Book and package

Christian Kleiber and Achim Zeileis: *Applied Econometrics with R*, Springer-Verlag, New York, 2006 (?)

Contents:

- R Basics
- Linear Regression and Extensions
- Validating Linear Models
- Models of Microeconometrics
- Time Series Models
- Programming Your Own Analysis

`AER`: **Book and package**

Package `AER` contains more than 60 data sets (with examples) from

- textbooks

  – B. Baltagi: Econometrics, 3e
  – W.H. Greene: Econometric Analysis, 5e
  – P.H. Franses: Time Series Models for Business and Economic Forecasting

- data archives of the *Journal of Applied Econometrics*, *Journal of Business and Economic Statistics*

- selected further sources (*Empirical Economics*, PARADE magazine ...)

Be sure to attend

Econometrics and Social Science

(Spotlights: HS 0.3, Forum: Aula 3)

Friday 16   15:00-18:30