Heteroscedastic Censored Regression for Weather Forecasts

Jakob W. Messner, Georg J. Mayr, Achim Zeileis
Weather forecasts

**Numerical Weather Prediction (NWP)**

- Observations $\rightarrow$ estimate current atmospheric state.
- Simulate atmospheric processes with numerical models.

$\Rightarrow$ Compute future weather
Weather forecasts

Numerical Weather Prediction (NWP)

- Observations → estimate current atmospheric state.
- Simulate atmospheric processes with numerical models.

⇒ Compute future weather

Problems:

- Few observations
- Observation errors
- Not perfectly known atmospheric processes
- Unresolved processes

⇒ NWP errors
NWP errors

3 days accumulated precipitation

precipitation [mm/3 days]

Date


NWP forecast
NWP errors

3 days accumulated precipitation

Date
precipitation [mm/3 days]
3 days accumulated precipitation
NWP forecast
observation

precipitation [mm/3 days]

NWP errors

![NWP errors graph](image-url)
NWP errors

The graph shows a scatter plot comparing NWP forecast values against observations. The orange line represents the linear regression line, indicating the trend between the forecast and observed data. The points deviate from the line, suggesting some errors in the NWP forecast.

The x-axis represents the NWP forecast, while the y-axis represents the observations.
Ensemble prediction

NWP error sources:

- Initial conditions
- Model formulations
Ensemble prediction

**NWP error sources:**
- Initial conditions
- Model formulations

**Idea:**
- Perturbed initial conditions
- Different model formulations

⇒ Compute different weather scenarios
NWP errors

3 days accumulated precipitation

Date
precipitation [mm/3 days]

observation
NWP forecast

0 20 40 60 80 100

precipitation [mm/3 days]
NWP errors

3 days accumulated precipitation

observation
ensemble forecasts
ensemble mean

Date
precipitation [mm/3 days]
Statistical models

Challenges:
- utilize uncertainty information from ensemble forecasts
- limited (non-negative) response
Heteroscedastic censored regression

\[ \text{rain}^* \sim \mathcal{N}(\mu, \sigma^2) \]

\[ \mu = \beta_0 + \beta_1 \times \text{ensmean} \]

\[ \log(\sigma) = \gamma_0 + \gamma_1 \times \text{enssd} \]

- \text{rain}^*: (latent) precipitation
- \text{ensmean}: ensemble mean forecast
- \text{enssd}: ensemble standard deviation
- \beta_0, \beta_1, \gamma_0, \gamma_1: regression coefficients
Heteroscedastic censored regression

\[ \text{rain}^* \sim \mathcal{N}(\mu, \sigma^2) \]

\[ \mu = \beta_0 + \beta_1 \times \text{ensmean} \]

\[ \log(\sigma) = \gamma_0 + \gamma_1 \times \text{enssd} \]

Consider non-negativity:

\[ \text{rain} = \begin{cases} 
0 & \text{rain}^* \leq 0 \\
\text{rain}^* & \text{rain}^* > 0 
\end{cases} \]
Heteroscedastic censored regression
Heteroscedastic censored regression

![Graph showing heteroscedastic censored regression](image-url)
Heteroscedastic censored regression
Heteroscedastic censored regression

\[ \text{rain}^* \sim \mathcal{N}(\mu, \sigma^2) \]

\[ \mu = \beta_0 + \beta_1 \ast \text{ensmean} \]

\[ \log(\sigma) = \gamma_0 + \gamma_1 \ast \text{enssd} \]

Consider non-negativity:

\[ \text{rain} = \begin{cases} 
0 & \text{rain}^* \leq 0 \\
\text{rain}^* & \text{rain}^* > 0
\end{cases} \]
Implementation in R

Model fitting

- `crch()` from package `crch`
- `glm()`-like interface
Implementation in R

Model fitting

- `crch()` from package `crch`
- `glm()`-like interface
- model specification via formula plus data
- two part formula, e.g., \( y \sim x_1 + x_2 + x_3 \mid z_1 + z_2 \)
Implementation in R

Model fitting

- `crch()` from package `crch`
- `glm()`-like interface
- model specification via formula plus data
- two part formula, e.g., `y ~ x1 + x2 + x3 | z1 + z2`
- log-likelihood maximized numerically via `optim()`
- distributions: Gaussian, logistic, student-t
- censored and truncated
Implementation in R

Model fitting

- `crch()` from package `crch`
- `glm()`-like interface
- model specification via formula plus data
- two part formula, e.g., \( y \sim x1 + x2 + x3 \mid z1 + z2 \)
- log-likelihood maximized numerically via `optim()`
- distributions: Gaussian, logistic, student-t
- censored and truncated
- methods: `summary()`, `coef()`, `residuals()`, `logLik()`, `predict()`, ...
R> CRCH <- crch(rain ~ ensmean | ensssd, data = Rain, left = 0)
R>
R> CRCH <- crch(rain ~ ensmean | enssd, data = Rain, left = 0)
R> summary(CRCH)
R> CRCH <- crch(rain ~ ensmean | enssd, data = Rain, left = 0)
R> summary(CRCH)

Call:
crch(formula = rain ~ ensmean | enssd, data = Rain, left = 0)

Standardized residuals:

   Min  1Q Median  3Q   Max
-3.7622 -0.3298  0.2448  0.7536 3.8235

Coefficients (location model):

        Estimate Std. Error  z value Pr(>|z|)  
(Intercept)  -1.36061 0.04609    -29.52  <2e-16 ***
enssmean     0.78533 0.00962     81.63  <2e-16 ***

Coefficients (scale model with log link):

        Estimate Std. Error  z value Pr(>|z|)  
(Intercept)  0.33189 0.02078     15.975 < 2e-16 ***
ensssd      0.25445 0.03827      6.649 2.96e-11 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Distribution: gaussian
Log-likelihood: -1.028e+04 on 4 Df
Number of iterations in BFGS optimization: 12
Censored regression

![Graph showing censored regression data with linear regression line]
Censored regression

- Linear regression
- Censored regression

Graph showing the relationship between ensemble mean (x-axis) and rain (y-axis), with linear and censored regression lines.
Predictions

![Graph showing predictions for rain* with predictive density on the y-axis and rain* on the x-axis. The data point for 2013-10-27 is marked with an orange line.](image)
R> location <- predict(CRCH, newdata = Rain[j,], type = "location")
R> scale <- predict(CRCH, newdata = Rain[j,], type = "scale")
Predictions

```
R> location <- predict(CRCH, newdata = Rain[j,], type = "location")
R> scale    <- predict(CRCH, newdata = Rain[j,], type = "scale")
```
Will it rain in Aalborg the next 3 days?
UseR forecast

- $P(\text{rain} > 0) = 38\%$
- $E(\text{rain}) = 0.5\text{ mm}$
Summary

Censored regression with conditional heteroscedasticity:

- effective usage of ensemble information
- non-negativity of precipitation considered
- **crch** package:
  - model fitting with `crch()`
  - various methods
Summary

Censored regression with conditional heteroscedasticity:
- effective usage of ensemble information
- non-negativity of precipitation considered
- **crch** package:
  - model fitting with `crch()`
  - various methods

sunny weather for UseR!
Thank you!


jakob.messner@uibk.ac.at