Regression för tidsserier

Mattias Villani

Statistiska institutionen
Stockholms universitet

Institutionen för datavetenskap
Linköpings universitet

mattiasvillani.com @matvil mattiasvillani

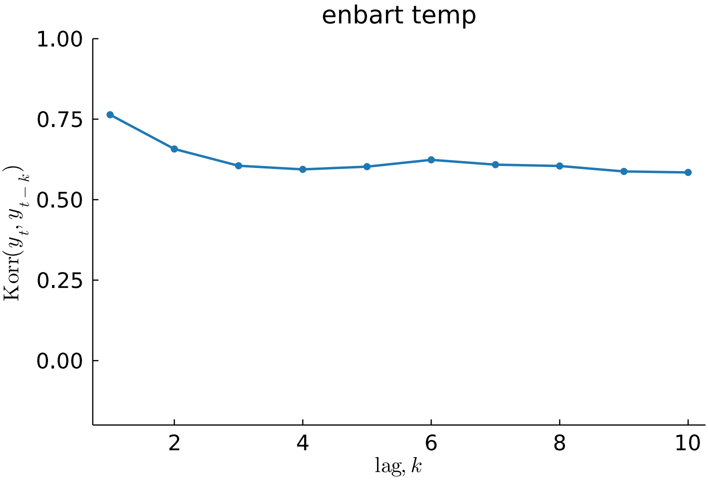
Översikt

- **Regression för tidsserier**
- **Durbin-Watson test**
- **Modellval** för **tidsserieregression** genom **prognosförmåga**.

Regression för tidsserier

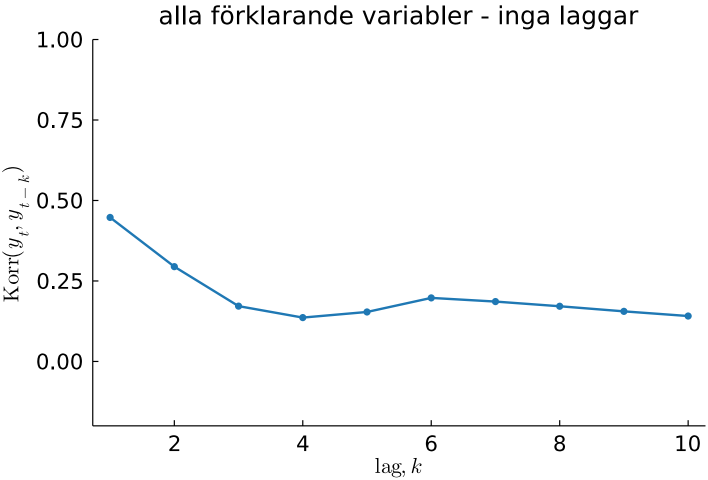
- Regression
$$y = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$$
där **feltermerna** ε antas bara **oberoende** från $N(0, \sigma_\varepsilon^2)$.
- **Oberoende = okorrelerade** för **normalfördelade** variabler.
- Regressionen skattas med
$$y = a + b_1 x_1 + \ldots + b_k x_k$$
och vi får residualer
$$e_t = y_t - \hat{y}_t.$$
- Vi kan undersöka om **residualerna är okorrelerade**.
- Två metoder:
  - Visuellt genom att **plotta autokorrelationsfunktionen för** e_t
  - **Durbin-Watson test**

ACF residualer - temp



ACF residualer - alla variabler

- Regression med alla förklarande variabler:
  `temp,hum,windspeed,holiday,workingday,säsong,yr`.



Regression för tidsserier

- **Regressionsmodeller för tidsserier**

$$y_t = \alpha + \beta_1 x_t + \varepsilon_t$$

  får ofta korrelerade residualer. 🥺

- Kombinera **enkel regression och** AR(1) 😁

$$y_t = \alpha + \beta_1 x_t + \beta_2 y_{t-1} + \varepsilon_t$$

- Kombinera **multipel regression och** AR(p) 😍

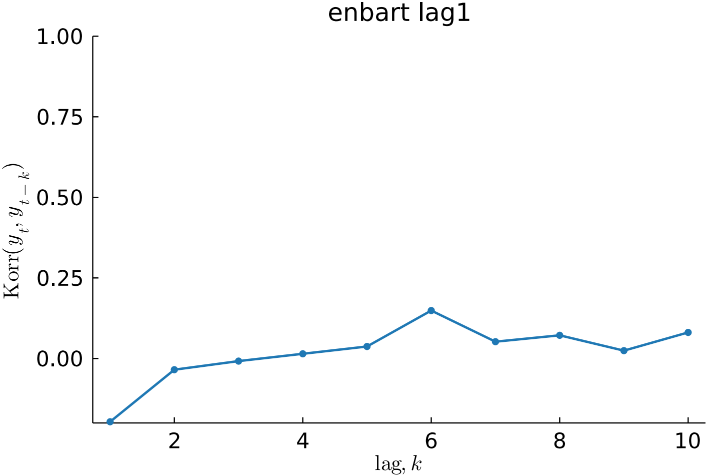
$$y_t = \alpha + \beta_1 x_t + \ldots + \beta_k x_{kt} + \beta_{k+1} y_{t-1} + \ldots + \beta_{k+p} y_{t-p} + \varepsilon_t$$

- Cykeluthyrning:

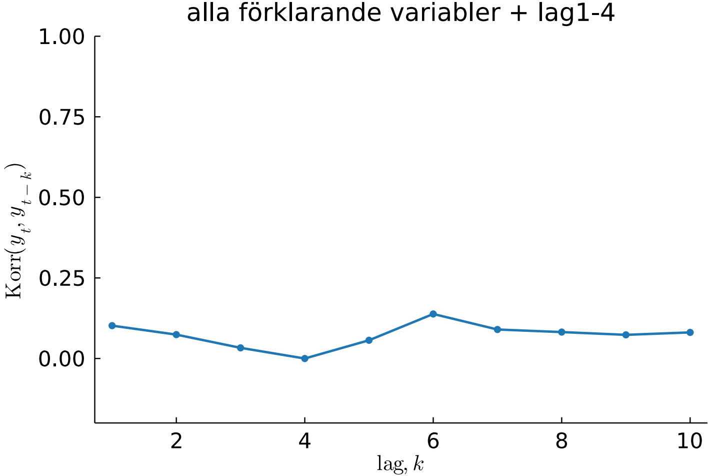
$$\texttt{AntalUthyr}_{\text{idag}} = a + b_1 \cdot \texttt{temp}_{\text{idag}} + b_2 \cdot \texttt{AntalUthyr}_{\text{igar}}$$

- Standardfel och hypotestest måste korrigeras om laggar av y_t används som förklarande variabel.

ACF residualer - enbart lag 1



ACF residualer - alla variabler + lag 1-4



Durbin-Watson test

- Test för autokorrelation (i feltermer).
- Teststatistika
$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$
- Durbin-Watson testar första autokorrelationen (AJÅ)
$$d \approx 2(1 - r_1)$$
- Teststatistikan uppfyller
$$0 \leq d \leq 4$$
- Grova kritiska gränser:
$$\begin{aligned} d \text{ nära } 2 &\implies \text{ej signifikant} \\ d < 1 &\implies \text{signifikant positiv autokorrelation} \\ d > 1 &\implies \text{signifikant negativ autokorrelation} \end{aligned}$$
- Durbin-Watson test kan inte användas när man har laggar av målvariabeln (y_{t-1} etc) som förklarande variabler.

Durbin-Watson test - cykeluthyrning

```
> library(car)
> lmfit = lm(nRides ~ temp , data = bike)
> durbinWatsonTest(lmfit)
 lag Autocorrelation D-W Statistic p-value
   1       0.7641582     0.4678707       0
 Alternative hypothesis: rho != 0
```

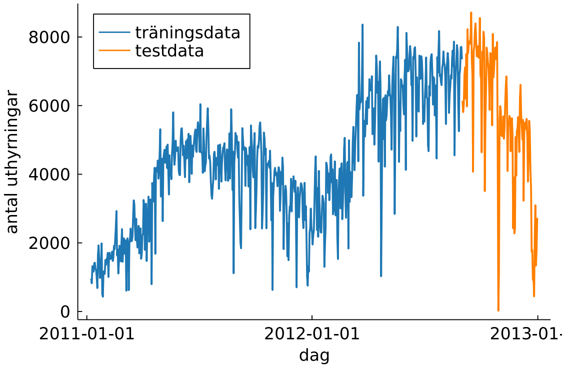
```
> lmfit = lm(nRides ~ temp + hum + windspeed + holiday + workingday + as.factor(season) + yr, data = bike)
> durbinWatsonTest(lmfit)
 lag Autocorrelation D-W Statistic p-value
   1       0.4472755      1.104221       0
 Alternative hypothesis: rho != 0
```

Förklarande variabler	R^2	$r_1^{(\mathrm{res})}$	d	p-värde
`temp`	0.385	0.764	$0.471^{\star\star\star}$	$<$ 1e-93
`temp,hum,windspeed,holiday,workingday,säsong,yr`	0.795	0.447	$1.104^{\star\star\star}$	$<$ 1e-33

Cykeluthyrningar - utvärdera prognosförmåga

- **Träningsdata**: Jan 1, 2011 - Aug 31, 2012.
- **Testdata**: Sept 1, 2012 - Dec 31, 2012.
- **Prediktionsmått RMSE**

$$\mathrm{RMSE}_{\mathrm{test}} = \sqrt{\frac{1}{n_{\mathrm{test}}} \sum_{t \in \mathrm{Testdata}} (y_t - \hat{y}_t)^2}$$



Cykeluthyrningar

- Träningsdata: Jan 1, 2011 - Aug 31, 2012.
- Testdata: Sept 1, 2012 - Dec 31, 2012.

Förklarande variabler	R^2	$\mathrm{RMSE}_{\mathrm{test}}$
`temp`	0.385	2346.60
`temp,hum,windspeed,holiday,workingday,säsong,yr`	0.795	1292.07
`lag1`	0.714	1274.32
`lag1,lag2`	0.730	1279.30
`lag1-lag4`	0.746	1267.84
`lag1-lag6`	0.764	1262.10
`temp,hum,windspeed,holiday,workingday,säsong,yr,lag1`	0.825	1127.63
`temp,hum,windspeed,holiday,workingday,säsong,yr,lag1-lag4`	0.827	1118.83
`temp,hum,windspeed,holiday,workingday,säsong,yr,lag1-lag6`	0.830	1117.63
`temp,hum,windspeed,holiday,workingday,säsong,yr,lag1-lag6,Lasso`	NA	1118.34