

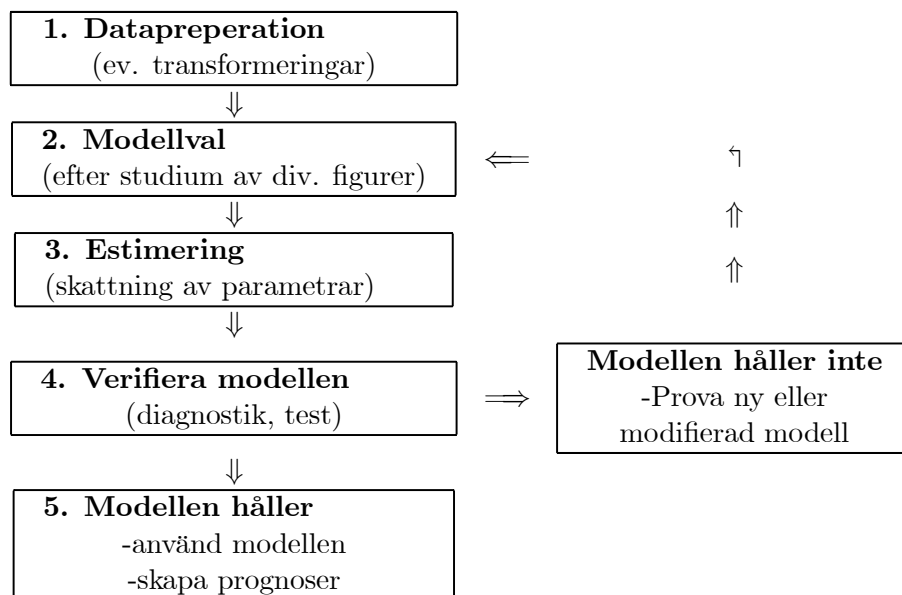
Kap 6. Regressionsmodeller för tidsserier

Modeller:

1. $y_t = \beta_0 + \epsilon_t$ (Ingen trend)
2. $y_t = \beta_0 + \beta_1 t + \epsilon_t$ (Linjär trend)
3. $y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$ (Kvadratisk trend)
4. $y_t = \beta_0 + \beta_1 t + \beta_{s1} x_{s1,t} + \beta_{s2} x_{s2,t} + \dots + \beta_{s(L-1)} x_{s(L-1),t} + \epsilon_t$ (Linjär trend+säsong)
5. $y_t = \beta_0 (\beta_1^t) \epsilon_t$ (Exponentiell trend)

I modellerna ovan antas feltermen ϵ_t vara $ON(0, \sigma^2)$

För att underlätta letandet efter den bästa möjliga modellen kommer vi genom hela kursen arbeta enligt följande schema:



Vilken modell är bäst lämpad? Se på avvikelsemått:

1. MAD - Mean Absolute Deviation

$$MAD = \frac{1}{n} \sum_{i=1}^n |e_t|$$

2. MSE - Mean Squared Error

$$MSD = \frac{1}{n} \sum_{i=1}^n e_t^2$$

där $e_t = y_t - \hat{y}_t$. Eftersom MSD är de kvadrerade avvikelserna är detta mått mer känslig för stora avvikelser i enskilda observationer.

Autokorrelation

Modellerna ovan kräver att feltermerna, ϵ_t , är oberoende och normalfördelade med väntevärde 0 och konstant varians σ^2 (dvs. $\epsilon_t \sim ON(0, \sigma^2)$). Normalfördelningsantagandet testas t.ex. med Anderson-Darlings normalitetstest, men hur testas oberoendeantagandet? Vad gör vi om feltermerna är beroende (autokorrelerade)? Om feltermerna visar sig vara beroende, innebär detta att ϵ_{t-1} påverkar ϵ_t . Ett sätt att beskriva denna påverkan är genom modellen:

$$\epsilon_t = \phi \epsilon_{t-1} + a_t$$

där $a_t \sim ON(0, \sigma^2)$, och ϕ är korrelationskoefficienten mellan feltermerna ϵ_{t-1} och ϵ_t . Om $\phi > 0$, indikerar detta på *positiv autokorrelation*, och om $\phi < 0$ finns misstankar om *negativ autokorrelation*.

En residualplot kan ge en fingervisning om feltermernas egenskaper, men det är ofta mycket svårt att med ögat avgöra beroende. Därför används ofta *Durbin-Watson's test*, där man testar hypoteserna:

H_0 : Feltermerna är inte autokorrelerade
 H_1 : Feltermerna är positivt/negativt autokorrelerade

Teststatistikan är

$$D = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

för stora värden på n kan man använda approximationen:

$$D \approx 2(1 - r_1)$$

Denna teststatistika har som synes utfallsrummet $\Omega = \{0 \leq D \leq 4\}$. Durbin och Watson skapade även en tabell över sin testvariabel (i boken tabell A5 och A6), där de kritiska värdena $d_{L,\alpha}$ och $d_{U,\alpha}$ finns tabellerade.

Följande gäller:

- nollhypotesen förkastas, och vi tror på positiv autokorrelation om $0 \leq d_{obs} \leq d_{L,\alpha}$
- nollhypotesen förkastas, och vi tror på negativ autokorrelation om $(4 - d_{L,\alpha}) \leq d_{obs} \leq 4$
- nollhypotesen accepteras om $d_{U,\alpha} \leq d_{obs} \leq (4 - d_{U,\alpha})$
- testet antas vara obestämt om:

$$d_{L,\alpha} \leq d_{obs} \leq d_{U,\alpha} \quad \text{eller om} \quad (4 - d_{U,\alpha}) \leq d_{obs} \leq (4 - d_{L,\alpha})$$