Model estimation and comparison

Linear Difference Equation with white noise

- $t = 1, 2, \ldots, N$ observations

Linear Difference Equation with white noise

- $t = 1, 2, ..., N$ observations
- $y(t)$ single uni variate output variable

Linear Difference Equation with white noise

- $t = 1, 2, ..., N$ observations
- $y(t)$ single uni variate output variable
- $x_j(t)$ j-th univariate input variable $j = 1, 2, .., r$

Linear Difference Equation with white noise

- $t = 1, 2, ..., N$ observations
- $y(t)$ single uni variate output variable
- $x_j(t)$ j-th univariate input variable $j = 1, 2, .., r$
- $$y(t) = \sum_{i=1}^{p} \alpha_i \cdot y(t-i) + \sum_{j=1}^{r} \sum_{k=0}^{q} \beta_{jk} \cdot x_j(t-k) + w(t)$$
  $$w(t) \text{ ~ white noise}$$

Linear Difference Equation with white noise

- $t = 1, 2, ..., N$ observations
- $y(t)$ single uni variate output variable
- $x_j(t)$ j-th univariate input variable $j = 1, 2, .., r$
- $$y(t) = \sum_{i=1}^{p} \alpha_i \cdot y(t-i) + \sum_{j=1}^{r} \sum_{k=0}^{q} \beta_{jk} \cdot x_j(t-k) + w(t)$$
  $$w(t) \text{ ~ white noise}$$
- Example with $p = 1$ and $r = 2$ and $q = 0$

Linear Difference Equation with white noise

- $t = 1, 2, ..., N$ observations
- $y(t)$ single uni variate output variable
- $x_j(t)$ j-th univariate input variable $j = 1, 2, .., r$
- $$y(t) = \sum_{i=1}^{p} \alpha_i \cdot y(t-i) + \sum_{j=1}^{r} \sum_{k=0}^{q} \beta_{jk} \cdot x_j(t-k) + w(t)$$
  $$w(t) \text{ ~ white noise}$$
- Example with $p = 1$ and $r = 2$ and $q = 0$
- $$y(t) = \alpha \cdot y(t-1) + \beta_1 \cdot x_1(t) + \beta_2 \cdot x_2(t) + w(t)$$

Least Squares estimation

$$
\begin{aligned}
Y :=& \quad N \times 1 \text{ output vector} \\
X =& \quad N \times (r \cdot q) \text{ matrix of inputs} \\
Z =& \quad N \times p \text{ matrix of lagged outputs} \\
U =& \quad [Z|X] = N \times (p + r \cdot q) \text{ matrix of predictors} \\
W =& \quad N \times 1 \text{ vector of errors} \\
B =& \quad [B_y|B_x] = (p + r \cdot q) \times 1 \text{ vector of parameters} \\
Y =& \quad U \cdot B + W \\
\hat{B} =& \quad (U' \cdot U)^{-1} \cdot (U' \cdot Y) \\
\hat{Y} =& \quad U \cdot \hat{B}, \; E = Y - \hat{Y}, \\
SSE =& \quad E' \cdot E
\end{aligned}
$$

Model comparison for Lin stoch Diff + white noise

$$
\begin{aligned}
SSE_C \quad : &= \text{ sum of squared errors from model } c \\
\ln like_C &= -(N/2)\cdot \ln\left(SSE_C\,/\,N\right) - N/2\cdot\left[\ln\left(2\pi\right)+1\right] \\
BIC_C &= N\cdot\ln(SSE_C\,/\,N) + p_C\cdot\ln\left(N\right) + N\cdot\left[\ln\left(2\pi\right)+1\right] \\
p_C &= \text{no. parms in model C} \\
SSE_R \quad : &= \text{ sum of squared errors for restricted version of C} \\
p_R &= \text{no. parms in model R (after linearly constraing parms in C)}
\end{aligned}
$$

$$
\begin{aligned}
MSR &= \left(SSE_R - SSE_C\right)/\left(p_C - p_R\right) \\
MSE &= SSE_C\,/\,\left(N - p_C\right) \\
F &= MSR\,/\,MSE, \quad \text{dist under H0 as central } F \text{ with} \\
df_N &= \left(p_C - p_R\right),\ df_D = N - p_C
\end{aligned}
$$

State Space model

$$
\begin{aligned}
Z &:= p \times 1 \text{ state vector} \\
X &:= q \times 1 \text{ input vector} \\
W &:= p \times 1 \text{ white noise vector} \\
Y &:= r \times 1 \text{ output vector} \\
\epsilon &:= r \times 1 \text{ measurement error vector} \\
Z(t) &= F \cdot Z(t-1) + G \cdot X(t) + W(t) \\
Y &= H \cdot Z(t) + \epsilon(t)
\end{aligned}
$$

Likelihood function for State Space model

Assume $W\tilde{}N(0,\Omega)$ and $\epsilon\tilde{}N(0,\Psi)$,
$\theta = F, G, H, \Omega, \Psi$ vector of parameters that need to be estimated

$$
\begin{aligned}
\hat{Z}(t) &= E\left[Z(t)|Y(1)...Y(t-1)\right] \\
\varepsilon(t) &= Z(t) - \hat{Z}(t) \\
P_t &= E\left[\varepsilon(t)\cdot\varepsilon(t)'\right] \\
\hat{\epsilon}(t) &= Y(t) - H\cdot\hat{Z}(t) \\
&= H\cdot Z + \epsilon - H\cdot\hat{Z} = H\cdot\left(Z-\hat{Z}\right) + \epsilon \\
V\left[\hat{\epsilon}(t)\right] &= V\left(H\cdot\varepsilon + \epsilon\right) = E\left[\hat{\epsilon}(t)\cdot\hat{\epsilon}(t)'\right] \\
&= H\cdot P_t\cdot H' + \Psi = \Sigma_t \\
\ln L(\theta) &= \ln f\left(Y_1,...,Y_N|\theta\right) = \ln f\left(\hat{\epsilon}_1,...,\hat{\epsilon}_N|\theta\right) \\
-2\cdot\ln L(\theta) &= \sum_{t=1}^{N}\ln|\Sigma_t| + \sum_{t=1}^{N}\hat{\epsilon}'(t)\Sigma_t^{-1}\hat{\epsilon}(t)
\end{aligned}
$$

Chi square test

Assume Simple model is nested within complex model

$$
\begin{aligned}
-2 \cdot \ln L &= G^2 \\
G_C^2 &\rightarrow \text{Complex model with } p \text{ parameters} \\
G_S^2 &\rightarrow \text{Simple model with } q < p \text{ parameters} \\
& \left(G_S^2 - G_C^2\right) \tilde{\ } \chi^2 (p - q)
\end{aligned}
$$

BIC Bayesian information (Schwartz) Criterion Comparisonn

Models do not need to be nested
Model A has p parameters
Model B has q parameters
Suppose $p > q$
Asympototically equivalent to a Bayesian model comparison
Choose model with lowest BIC

$$
\begin{aligned}
-2 \cdot \ln L &= G^2 \\
BIC_A &= G_A^2 + p \cdot \ln(N) \\
BIC_B &= G_B^2 + q \cdot \ln(N) \\
BIC_{Diff} &= \left(G_B^2 - G_A^2\right) + (q - p) \cdot \ln(N)
\end{aligned}
$$

AIC (Akaike) information Criterion Comparisonn

Models do not need to be nested
Model A has p parameters
Model B has q parameters
Suppose $p > q$
Measures (Kullback-Leibler) Distance between model and true distribution
Choose model with lowest AIC

$$\begin{aligned} -2 \cdot \ln L &= G^2 \\ AIC_A &= G_A^2 + p \cdot 2 \\ AIC_B &= G_B^2 + q \cdot 2 \\ AIC_{Diff} &= \left(G_B^2 - G_A^2\right) + (q - p) \cdot 2 \end{aligned}$$

Newton - Raphson Parameter Search

Quadratic Approximation of the objective function at some point in the parameter space

$$
\begin{aligned}
F(\theta) &= -2 \ln L(\theta) \\
\nabla &= \frac{\partial}{\partial \theta} F(\theta) \;:=\text{gradient} \\
H &= \frac{\partial^2}{\partial \theta \cdot \partial \theta'} F(\theta) \;:= \text{Hessian} \\
F(\theta + \delta) &\approx F(\theta) + \delta' \cdot \nabla + \frac{1}{2}\left[\delta' \cdot H \cdot \delta\right] \\
\frac{\partial}{\partial \delta} F(\theta + \delta) &= \nabla + H \cdot \delta = 0 \\
\delta &= -H^{-1} \cdot \nabla \\
\theta_k &= \theta_{k-1} + s \cdot \delta \\
\text{as } N &\rightarrow \infty \,,\; H^{-1} \approx Var\left[\hat{\theta}\right]
\end{aligned}
$$