SYSTEM ESTIMATION METHODS II: STRUCTURED ESTIMATION

ADVANCED TOPICS

- **Optimal one-step ahead predictor for LTI systems**
- **ML estimation of LTI models**
- **Concentrated ML for MIMO models**
- **Exact likelihood function for an AR(1) model**

OPTIMAL ONE-STEP AHEAD PREDICTOR FOR LTI SYSTEMS

$$y_t = G(q)u_t + H(q)w_t = G(q)u_t + [H(q) - I]w_t + w_t$$

$G(q)u_t$ depends on u_{t-1}, ..., so it can be computed at time $t-1$.
Notice that the elements of $H(q) - I$ have relative degree > 0, so $[H(q) - I]w_t$ depends on past values of w_t. This term can be computed by noting that

$$w_t = H^{-1}(q)[y_t - G(q)u_t]$$

Note that w_t depends on $y_t, y_{t-1}, \ldots$ and $u_{t-1}, u_{t-2}, \ldots$, so it *cannot* be computed at time $t-1$. But $w_{t-1}, \ldots$ *can* be computed at time $t-1$.

Therefore:

$$\begin{aligned} y_t &= G(q;\theta)u_t + [H(q;\theta) - I]\{H^{-1}(q;\theta)[y_t - G(q;\theta)u_t]\} + w_t \\ &= \underbrace{[I - H^{-1}(q;\theta)]y_t + H^{-1}(q;\theta)G(q;\theta)u_t}_{\hat{y}_{t|t-1}(\theta) := E\{y_t | Y_{t-1}, U_{t-1}\}} + \underbrace{w_t}_{\varepsilon_t} \end{aligned}$$

ML ESTIMATION OF LTI MODELS

To compute the MLE of θ given $U_N := \{u_N, \ldots, u_1\}$ and $Y_N := \{y_N, \ldots, y_1\}$, we need $P\{Y_N \mid U_N, \theta\}$. By Bayes' Theorem,

$$P\{Y_N \mid U_N, \theta\} = \prod_{t=1}^{N} P\{y_t \mid Y_{t-1}, U_t; \theta\}$$

Now, by the rule for transforming random variables:

$$P\{y_t \mid Y_{t-1}, U_t; \theta\} = P_\varepsilon\{\varepsilon_t \mid Y_{t-1}, U_t; \theta\}\big|_{\varepsilon_t = y_t - f(Y_{t-1}, U_t, t; \theta)} \cdot \left| \det\left(\frac{\partial \varepsilon_t}{\partial y_t} \right) \right| = P_\varepsilon\{\varepsilon_t(\theta); \theta\}$$

where $\varepsilon_t(\theta) := y_t - f(Y_{t-1}, U_t, t; \theta)$, and $P_\varepsilon\{\,\cdot\,; \theta\}$ is the conditional density for ε_t given Y_{t-1} and U_t, and may depend on θ. Thus,

$$P\{Y_N \mid U_N, \theta\} = \prod_{t=1}^{N} P_\varepsilon\{\varepsilon_t(\theta); \theta\}$$

ML Estimation of LTI models (cont.)

When $\{\varepsilon_t\}$ are IID Gaussian with covariance Σ,

$$P\{Y_N \mid U_N, \theta\} = \prod_{t=1}^{N} \left([(2\pi)^m \det \Sigma]^{-1/2} \exp\left\{ -\frac{1}{2} \varepsilon_t^T(\theta) \Sigma^{-1} \varepsilon_t(\theta) \right\} \right)$$

$$= [(2\pi)^m \det \Sigma]^{-N/2} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{N} \varepsilon_t^T(\theta) \Sigma^{-1} \varepsilon_t(\theta) \right\}$$

Hence the log likelihood function is

$$l(\theta) = \ln P\{Y_N \mid U_N, \theta\} = -\frac{Nm}{2} \ln 2\pi - \frac{N}{2} \ln \det \Sigma - \frac{1}{2} \sum_{t=1}^{N} \varepsilon_t^T(\theta) \Sigma^{-1} \varepsilon_t(\theta)$$

Therefore, the MLE is

$$\boxed{\hat{\theta} = \arg\max_{\theta \in \Theta} l(\theta) = \arg\min_{\theta \in \Theta} \sum_{t=1}^{N} \varepsilon_t^T(\theta) \Sigma^{-1} \varepsilon_t(\theta)}$$

CONCENTRATED ML FOR MIMO MODELS

If Σ is unknown, it has to be estimated together with θ. Here we can use the fact that

$$\max_{\theta,\Sigma} l(\theta,\Sigma) = \max_{\theta}\left[\max_{\Sigma} l(\theta,\Sigma)\right]$$

Now,

$$\frac{\partial l}{\partial \Sigma} = -\frac{N}{2}\Sigma^{-1} - \frac{1}{2}\Sigma^{-1}\left[\sum_{t=1}^{N}\varepsilon_t(\theta)\varepsilon_t^T(\theta)\right]\Sigma^{-1}$$

Forcing $\partial l / \partial \Sigma = 0$ gives

$$\boxed{\hat{\Sigma}_{ML} = \frac{1}{N}\sum_{t=1}^{N}\varepsilon_t(\theta)\varepsilon_t^T(\theta)}$$

CONCENTRATED ML FOR MIMO MODELS (CONT.)

Substituting $\hat{\Sigma}_{ML}$ in $l(\theta,\Sigma)$ yields

$$
\begin{aligned}
l(\theta,\hat{\Sigma}_{ML})
&= -\frac{Nm}{2}\ln 2\pi - \frac{N}{2}\ln\det\left[\frac{1}{N}\sum_{t=1}^{N}\varepsilon_t(\theta)\varepsilon_t^T(\theta)\right] - \frac{1}{2}\sum_{t=1}^{N}\varepsilon_t^T(\theta)\left[\frac{1}{N}\sum_{s=1}^{N}\varepsilon_s(\theta)\varepsilon_s^T(\theta)\right]^{-1}\varepsilon_t(\theta) \\
&= -\frac{Nm}{2}\ln 2\pi - \frac{N}{2}\ln\det\left[\frac{1}{N}\sum_{t=1}^{N}\varepsilon_t(\theta)\varepsilon_t^T(\theta)\right] - \frac{1}{2}\operatorname{tr}\left\{\left[\frac{1}{N}\sum_{s=1}^{N}\varepsilon_s(\theta)\varepsilon_s^T(\theta)\right]^{-1}\sum_{t=1}^{N}\varepsilon_t(\theta)\varepsilon_t^T(\theta)\right\} \\
&= -\frac{Nm}{2}(\ln 2\pi + 1) - \frac{N}{2}\ln\det\left[\frac{1}{N}\sum_{t=1}^{N}\varepsilon_t(\theta)\varepsilon_t^T(\theta)\right]
\end{aligned}
$$

Therefore, $\hat{\theta}_{ML}$ satisfies

$$
\boxed{\hat{\theta}_{ML} = \arg\min_{\theta\in\Theta}\det\left[\frac{1}{N}\sum_{t=1}^{N}\varepsilon_t(\theta)\varepsilon_t^T(\theta)\right]}
$$

EXACT LIKELIHOOD FUNCTION FOR AN AR(1) MODEL

$$y_t + ay_{t-1} = w_t, \qquad |a| < 1$$

with $\{w_t\}$ Gaussian white noise of zero mean and variance λ^2. We have measurements $\{y_1, \ldots, y_N\}$, i.e.

$$\begin{aligned} y_1 &= -ay_0 + w_1 \\ y_2 &= -ay_1 + w_2 \\ &\vdots \\ y_N &= -ay_{N-1} + w_N \end{aligned}$$

Possible assumptions on initial condition y_0:

- $y_0 = 0$ $\Rightarrow$ *Conditional likelihood*
- y_0 is a parameter to be estimated
- y_0 has a prior (Bayesian) distribution
- $\{y_1, \ldots, y_N\}$ are taken from a stationary process $\Rightarrow$ *Exact likelihood*

EXACT LIKELIHOOD FUNCTION FOR AN AR(1) MODEL (CONT.)

$$
\begin{aligned}
P\{y_1,\ldots,y_N;a\} &= P\{y_N,\ldots,y_2 \mid y_1;a\}P\{y_1;a\} \\
&= P\{y_N,\ldots,y_3 \mid y_1,y_2;a\}P\{y_2 \mid y_1;a\}P\{y_1;a\} \\
&= \left[\prod_{t=2}^{N} P\{y_t \mid Y_{t-1};a\}\right] P\{y_1;a\}
\end{aligned}
$$

For $t \geq 2$:

$$
P\{y_t \mid Y_{t-1};a\} = P_w\{\varepsilon_t(a)\}\Big|_{\varepsilon_t(a)=y_t+ay_{t-1}} = \frac{1}{\sqrt{2\pi}\lambda}\exp\left[-\frac{(y_t+ay_{t-1})^2}{2\lambda^2}\right]
$$

and

$$
P\{y_1;a\} = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{y_1^2}{2\sigma^2}\right]; \qquad \sigma^2 = E\{y_1^2\} = \frac{\lambda^2}{1-a^2}
$$

EXACT LIKELIHOOD FUNCTION FOR AN AR(1) MODEL (CONT.)

The *exact* log likelihood function is, then,

$$l_{exact}(a) = \sum_{t=2}^{N} \ln\left[\frac{1}{\sqrt{2\pi}\lambda}\exp\left[-\frac{(y_t + a y_{t-1})^2}{2\lambda^2}\right]\right] + \ln\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{y_1^2}{2\sigma^2}\right]\right]$$

$$= -\frac{N}{2}\ln 2\pi - (N-1)\ln\lambda - \frac{1}{2\lambda^2}\sum_{t=2}^{N}(y_t + a y_{t-1})^2 - \ln\frac{\lambda^2}{1-a^2} - \frac{1-a^2}{2\lambda^2}y_1^2$$

In contrast, for the conditional log likelihood function (obtained by assuming $y_0 = 0$) we have

$$P\{y_1; a\} = \frac{1}{\sqrt{2\pi}\lambda}\exp\left[-\frac{y_1^2}{2\lambda^2}\right]$$

I.e.,

$$l_{cond}(a) = -\frac{N}{2}\ln 2\pi - N\ln\lambda - \frac{1}{2\lambda^2}\sum_{t=2}^{N}(y_t + a y_{t-1})^2 - \frac{y_1^2}{2\lambda^2}$$

EXACT LIKELIHOOD FUNCTION FOR AN AR(1) MODEL (CONT.)

Notice that $l_{exact}(a) = O(N)$ and $l_{cond}(a) = O(N)$, but $l_{exact}(a) - l_{cond}(a) = O(1)$, hence

$$\hat{a}_{exact} = \hat{a}_{cond} + O(N^{-1})$$

(A rigorous derivation of this follows e.g. from P. M. Robinson, "The stochastic difference between econometric statistics", *Econometrica*, vol. 56(3), 1988)