Maximum Likelihood Estimation of State-Space Models

The Kalman filter forecasts $\hat{\boldsymbol{\xi}}_{t|t-1}$ and $\hat{\boldsymbol{y}}_{t|t-1}$ are linear projections of $\boldsymbol{\xi}_t$ and $\boldsymbol{y}_t$ on $(\boldsymbol{x}_t, \mathscr{Y}_{t-1})$

They are optimal among all forecasts that are linear functions of $(\boldsymbol{x}_t, \mathscr{Y}_{t-1})$.

If $\boldsymbol{\xi}_1$ and $\{\boldsymbol{w}_t, \boldsymbol{v}_t\}_{t=1}^{T}$ are multivariate Gaussian

$\hat{\boldsymbol{\xi}}_{t|t-1}$ and $\hat{\boldsymbol{y}}_{t|t-1}$ are optimal among all forecasts that are functions of $(\boldsymbol{x}_t, \mathscr{Y}_{t-1})$ (linear and non-linear)

The distribution of $\boldsymbol{y}_t$ given $(\boldsymbol{x}_t, \mathscr{Y}_{t-1})$ is also multivariate Gaussian, of the form:

$\boldsymbol{y}_t|\boldsymbol{x}_t, \mathscr{Y}_{t-1} \sim MVN(A'\boldsymbol{x}_t + H'\hat{\boldsymbol{\xi}}_{t|t-1}, H'P_{t|t-1}H + R)$

Thus, the density function is

$f_{\boldsymbol{Y}_t|\boldsymbol{X}_t, \mathscr{Y}_{t-1}}(\boldsymbol{y}_t|\boldsymbol{x}_t, \mathscr{Y}_{t-1}, \boldsymbol{\theta})$

$= (2\pi)^{-n/2}|H'P_{t|t-1}H + R|^{-1/2}$ x $exp\{-1/2(\boldsymbol{y}_t - A'\boldsymbol{x}_t - H'\hat{\boldsymbol{\xi}}_{t|t-1}$ x $(H'P_{t|t-1}H + R)^{-1}(\boldsymbol{y}_t - A'\boldsymbol{x}_t - H'\hat{\boldsymbol{\xi}}_{t|t-1})'\}$

where $\boldsymbol{\theta}$ aggregates all known parameters in F, A, H, Q, and R

The log-likelihood is the joint density

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{T} log(f_{\boldsymbol{Y}_t|\boldsymbol{X}_t, \mathscr{Y}_{t-1}}(\boldsymbol{y}_t|\boldsymbol{x}_t, \mathscr{Y}_{t-1}, \boldsymbol{\theta}))$$

The log-likelihood can be maximized numerically with respect to $F(\boldsymbol{\theta}), A(\boldsymbol{\theta}), H(\boldsymbol{\theta}), Q(\boldsymbol{\theta})$, and $R(\boldsymbol{\theta})$

This is an exact log likelihood and yields exact MLEs.

Max likelihood estimation for MA and ARMA can be computed in this manner.

Basic Prescription

1. Guess $\boldsymbol{\theta}^{(0)}$
2. Guess $\boldsymbol{\theta}^{(s)}$, compute $F(\boldsymbol{\theta}^{(s)}), A(\boldsymbol{\theta}^{(s)}), H(\boldsymbol{\theta}^{(s)}), Q(\boldsymbol{\theta}^{(s)})$, and $R(\boldsymbol{\theta}^{(s)})$
3. Using the Kalman Filter to iteratively compute $\hat{\boldsymbol{\xi}}_{t|t-1}$ and $P_{t|t-1}$, $t = 1, \dots, T$
4. Compute the log-likelihood using $H(\boldsymbol{\theta}^{(s)}), A(\boldsymbol{\theta}^{(s)}), R(\boldsymbol{\theta}^{(s)})$, and $\{\hat{\boldsymbol{\xi}}_{t|t-1}, P_{t|t-1}\}_{t=1}^{T}$
5. Use a numerical method to update $\boldsymbol{\theta}^{(s+1)}$
6. If $||\boldsymbol{\theta}^{(s+1)} - \boldsymbol{\theta}^{(s)}|| < \tau$, stop. Otherwise, set $i = i + 1$ and return to step 7.

Updating $\boldsymbol{\theta}^{(i)} \to \boldsymbol{\theta}^{(i+1)}$ may involve numerical or analytical derivatives.

Analytical derivatives of the log likelihood with respect to each θ_i will involve $\dfrac{\partial \hat{\boldsymbol{\xi}}_{t|t-1}(\boldsymbol{\theta})}{\partial \theta_i}$ $\dfrac{\partial P_{t|t-1}}{\partial \theta_i}$

These derivatives can be updated recursively similar to $\hat{\boldsymbol{\xi}}_{t|t-1}$ and $P_{t|t-1}$