Gaussian Mixture Model (GMM)

Assume: $X_1 \sim N(x \mid \mu_1, \Sigma_1)$ $\quad P(X_1) = \pi_1$

$X_2 \sim N(x \mid \mu_2, \Sigma_2)$ $\quad P(X_2) = \pi_2$

$$x = \begin{pmatrix} a \\ b \end{pmatrix} \quad P(x) = P(x \in X_1) + P(x \in X_2)$$

$$= \underbrace{\pi_1 \Phi(x)}_{} + \pi_2 \Phi(x)$$

Probability of observing x.

Probability of belonging to X_1

$\times$ Probability of observing the value under X_1

Let all observations be X.

$$\underset{\max}{P(X \mid \pi, \mu, \Sigma)} = \underbrace{\boxed{\prod_{n=1}^{N}}}_{\text{Assumes i.i.d}} \left[\sum_{k=1}^{K} \pi_k N(x_n \mid \mu_k, \Sigma_k) \right]$$

$$1_{x \in k}(x_n) = \begin{cases} 1 & \text{if true,} \\ 0 & \text{else} \end{cases}$$

$$\gamma_{nk}(1_{x \in k}(x_n)) = P(1_{x \in k}(x_n) = 1 \mid x_n)$$

$$= \frac{P(x_n \mid 1(x_n) = 1)\, P(1(x_n) = 1)}{P(x_n)}$$

$$= \frac{N(x_n \mid \mu_k, \Sigma_k)\, \pi_k}{\sum_{j=1}^{K} N(x_n \mid \mu_j, \Sigma_j)\, \pi_j}$$

$N_k = \sum_{n=1}^{N} \gamma_{nk}$: Expected Number of Observations assigned to class k.

Maximize $P(X \mid \pi, \mu, \Sigma)$ w.r.t:

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk}(\mathbb{1}(x_n))\, x_n$$

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk}(\mathbb{1}(x_n))(x_n - \mu_k)(x_n - \mu_k)^T$$

$$\pi_k = N_k / N$$

Intuitively:

μ_k = Avg the weighted avg of x_n belonging to class k.

Σ_k = Avg the weighted avg MSE w.r.t class k.

π_k = Proportion of samples in k to total.

Note that everything depends of γ_{nk}
but γ_{nk} depends on π, μ, Σ.
$\Big\}$ Circular dependency

To solve this: Expectation-Maximization

Mixture Autoregressive Model

$$\underbrace{F(y_t \mid \overbrace{\mathcal{F}_{t-1}}^{\text{Information up to time } t-1})}_{\text{Conditional CDF of } Y_t \text{ at } y_t} = \sum_{k=1}^{K} \underbrace{\alpha_k}_{\text{weight}} \Phi\underbrace{\left(\frac{y_t - \overbrace{\phi_{k0}}^{\text{Intercept}(c_t)} - \phi_{k1} y_{t-1} - \dots - \phi_{kp_k} y_{t-p_k}}{\sigma_k} \right)}_{\frac{y_t - AR(p_k)}{\sigma_k} = \frac{e_{t,k}}{\sigma_k} \to \frac{x-\mu}{\sigma}}$$

where $\alpha_k \geq 0$ and $\sum_{k=1}^{K} \alpha_k = 1$

Intuitively

Conditional CDF given information up to $t-1$ = Weighted sum of the probability of observing the z-scaled observation assuming a Gaussian distribution.

Naturally:

$$E[y_t \mid \mathcal{F}_{t-1}] = \sum_{k=1}^{K} \alpha_k \mu_{k,t}$$

$$E[V[y_t \mid k, \mathcal{F}_{t-1}]] = \sum_{k=1}^{K} \alpha_k \sigma_k^2$$

$$V[E[y_t \mid k, \mathcal{F}_{t-1}]] = \sum_{k=1}^{K} \alpha_k \mu_{k,t}^2 - \left(\sum_{k=1}^{K} \alpha_k \mu_{k,t}\right)^2$$

$$\therefore V[y_t \mid \mathcal{F}_{t-1}] = \sum_{k=1}^{K} \alpha_k \sigma_k^2 + \underbrace{\sum_{k=1}^{K} \alpha_k \mu_{k,t}^2 - \left(\sum_{k=1}^{K} \alpha_k \mu_{k,t}\right)^2}_{\geq 0 \text{ (variance cannot be negative)}}$$

* Law of total variance:

$$\underset{\text{random}}{V[Y]} = \underbrace{E[V[Y \mid X]]}_{\text{Average variance}} + \underbrace{V[E[Y \mid X]]}_{\text{Variance of } \mu}$$

within each group

$$\underbrace{F(y_{t+2} \mid \mathcal{F}_t)}_{\text{CDF of } t+2} = \underbrace{\int \underbrace{F(y_{t+2} \mid \mathcal{F}_t, y_{t+1})}_{\text{probability mass}} \, \underbrace{dF(y_{t+1} \mid \mathcal{F}_t)}_{\substack{\text{point probabilities} \\ \text{(weights)}}}}_{\text{weighted average of the CDF}}$$

weighted average of the CDF for y_{t+2} given information up to time t, integrating w.r.t all possible y_{t+1} values.

$$\Downarrow$$

* Note that the above is intractable

∴ Approximate using Monte Carlo

$$F(y_{t+2} \mid \mathcal{F}_t) = \frac{1}{N} \sum_{j=1}^{N} F(y_{t+2} \mid \mathcal{F}_t, \underbrace{y^{j}_{t+1}}_{\text{Sampled from } F(y_{t+1} / \mathcal{F}_t)})$$

??

Sampled from $\underbrace{F(y_{t+1} / \mathcal{F}_t)}$

Perhaps they use Expectation-Maximization to get the CDF.

$F(y_{t+1} \mid \mathcal{F}_t ; \theta)$

$$f(y_t \mid \mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi}\,\sigma_k} e^{-\frac{(y_t - \mu_{k,t})^2}{2\sigma_k^2}}$$

$$Z_{k,t} = \begin{cases} 1 & \text{if } y_t \in k \\ 0 & \text{else} \end{cases}$$

if $y_t \notin k$, we would get $f(x)^0 = 1$

$$f(y_t \mid Z_t, \mathcal{F}_{t-1}) = \prod_{k=1}^{K} \left[f_k(y_t \mid \mathcal{F}_{t-1}) \right]^{Z_{k,t}}$$

$$f(y_t, Z_t \mid \mathcal{F}_{t-1}) = \prod_{k=1}^{K} \left[\alpha_k f_k(y_t \mid \mathcal{F}_{t-1}) \right]^{Z_{k,t}}$$

??

Note the link

Remember that conditional PDF:

$$k(z \mid \theta, x) = \frac{h(x, z \mid \theta)}{g(x \mid \theta)}$$

$$f(Z_t \mid y_t, \mathcal{F}_{t-1}; \theta)$$

$$= \frac{\text{Complete Likelihood}}{\text{Observed Likelihood}} = \frac{f(y_t, z_t \mid \mathcal{F}_{t-1}; \theta)}{f(y_t \mid \mathcal{F}_{t-1}; \theta)}$$

$$= \frac{L^c(\theta \mid x, z)}{L(\theta \mid x)}$$

Does not depend on random variable Z

$$\log L(\theta \mid x) = \log L^c(\theta \mid x, z) - \log k(z \mid \theta, x)$$

$$\log L(\theta \mid x) = \underbrace{E_{\theta_0}[\log L^c(\theta \mid x, z)]}_{\text{Maximize this}} - E_{\theta_0}[\log k(z \mid \theta, x)]$$

$$E_{\theta_0}[\log L^c(\theta \mid x, z)] = E_{\theta_0}[\log f(y_t, z_t \mid \mathcal{F}_{t-1}; \theta)]$$

$$f(y_t, z_t \mid \mathcal{F}_{t-1}) = \prod_{k=1}^{K} \left[\alpha_k f_k(y_t \mid \mathcal{F}_{t-1})\right]^{z_{k,t}}$$

$$\log f(y_t, z_t \mid \mathcal{F}_{t-1}) = \sum_{k=1}^{K} z_{k,t} \left[\log \alpha_k + \log f_k(y_t \mid \mathcal{F}_{t-1})\right]$$

$$f(y_t \mid \mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi}\,\sigma_k} e^{-\frac{(y_t - \mu_{k,t})^2}{2\sigma_k^2}}$$

$$\log f(y_t \mid \mathcal{F}_{t-1}) = \log (\sqrt{2\pi}\,\sigma_k)^{-1} + \log e^{-\frac{(y_t - \mu_{k,t})^2}{2\sigma_k^2}}$$

$$= -\left[\log (2\pi)^{\frac{1}{2}} + \log \sigma_k\right] - \frac{(y_t - \mu_{k,t})^2}{2\sigma_k^2}$$

$$= -\frac{1}{2}\log(2\pi) - \log \sigma_k - \frac{(y_t - \mu_{k,t})^2}{2\sigma_k^2}$$

$$\log f(y_t, z_t \mid \mathcal{F}_{t-1}) = \sum_{k=1}^{K} z_{k,t} \left[\log \alpha_k \underbrace{- \frac{1}{2}\log(2\pi)}_{\text{Removed. Constant will not contribute any information for maximization.}} - \log \sigma_k - \frac{(y_t - \mu_{k,t})^2}{2\sigma_k^2}\right]$$

$$l_t = \sum_{k=1}^{K} z_{k,t} \left[\log \alpha_k - \log \sigma_k - \frac{\varepsilon_{k,t}^2}{2\sigma_k^2}\right]$$

$$l = \sum_{t=p+1}^{n} l_t = \sum_{t=p+1}^{n} \left\{ \sum_{k=1}^{K} z_{k,t} \left[\log \alpha_k - \log \sigma_k - \frac{\varepsilon_{k,t}^2}{2\sigma_k^2}\right]\right\}$$

where $\varepsilon_{k,t} = y_t - \phi_{k,0} - \phi_{k,1} y_{t-1} - \dots - \phi_{k,p_k} y_{t-p_k}$

θ refers to our unknown parameters: α, ϕ, σ

$$l = \sum_{t=p+1}^{n} \left\{ \sum_{k=1}^{K} z_{k,t} \left[\log \alpha_k - \log \sigma_k - \frac{\varepsilon_{k,t}^2}{2\sigma_k^2} \right] \right\}$$

$$\frac{\partial l}{\partial \alpha_k} = \frac{\partial}{\partial \alpha_k} \sum_{t=p+1}^{n} \sum_{k=1}^{K} z_{k,t} \log \alpha_k \,, \quad \text{s.t. } \sum_{k=1}^{K} \alpha_k = 1$$

$$= \frac{\partial}{\partial \alpha_k} \sum_{t=p+1}^{n} \left(\sum_{k=1}^{K-1} z_{k,t} \log \alpha_k + z_{K,t} \log \alpha_K \right)$$

$$= \frac{\partial}{\partial \alpha_k} \sum_{t=p+1}^{n} \left[\underbrace{\sum_{k=1}^{K-1} z_{k,t} \log \alpha_k}_{} + z_{K,t} \log \left(1 - \sum_{k=1}^{K-1} \alpha_k \right) \right] \,,$$

simplex reparameterization

note that only one term belongs to k.
If $k=1$, we only care about α_1.

$$= \sum_{t=p+1}^{n} \frac{\partial}{\partial \alpha_k} \left[z_{k,t} \log \alpha_k + z_{K,t} \log \left(1 - \sum_{k=1}^{K-1} \alpha_k \right) \right]$$

$$= \sum_{t=p+1}^{n} \left(\frac{z_{k,t}}{\alpha_k} + \frac{z_{K,t}}{1 - \sum_{k=1}^{K-1} \alpha_k} (-1) \right)$$

$$= \sum_{t=p+1}^{n} \left(\frac{z_{k,t}}{\alpha_k} - \frac{z_{K,t}}{\alpha_K} \right)$$ where $k = 1, \dots, K-1$

$$l = \sum_{t=p+1}^{n} \left\{ \sum_{k=1}^{K} z_{k,t} \left[\log \alpha_k - \log \sigma_k - \frac{\varepsilon_{k,t}^2}{2\sigma_k^2} \right] \right\}$$

$$\frac{\partial l}{\partial \phi_{k,i}} = \frac{\partial}{\partial \phi_{k,i}} \sum_{t=p+1}^{n} \sum_{k=1}^{K} z_{k,t} \left(- \frac{\varepsilon_{k,t}^2}{2\sigma_k^2} \right)$$

$$= \sum_{t=p+1}^{n} \frac{\partial l}{\partial \varepsilon_{k,t}} \frac{\partial \varepsilon_{k,t}}{\partial \phi_{k,i}}$$

$$\frac{\partial l}{\partial \varepsilon_{k,t}} = z_{k,t} \left(- \frac{\varepsilon_{k,t}}{\sigma_k^2} \right)$$ we only care about a particular k.

$$\frac{\partial \varepsilon_{k,t}}{\partial \phi_{k,i}} = \frac{\partial}{\partial \phi_{k,i}} y_t - \sum_{i=0}^{p_k} \phi_{k,i} y_{t-i}$$ we only care about a particular i.

$$= -y_{t-i}$$

Let $y_{t-i} = u(y_t, i)$

$$u(y_t, i) = \begin{cases} 1 & \text{if } i = 0 \\ y_{t-i} & \text{if } i > 0 \end{cases}$$

$$\frac{\partial l}{\partial \phi_{k,i}} = \sum_{t=p+1}^{n} z_{k,t} \left(- \frac{\varepsilon_{k,t}}{\sigma_k^2} \right) \left(- u(y_t, i) \right)$$

$$= \sum_{t=p+1}^{n} \frac{z_{k,t}\, u(y_t, i)\, \varepsilon_{k,t}}{\sigma_k^2}$$ where $k = 1, \dots, K$, $i = 0, \dots, p_k$

$$l = \sum_{t=p+1}^{n} \left\{ \sum_{k=1}^{K} z_{k,t} \left[\log \alpha_k - \log \sigma_k - \frac{\varepsilon_{k,t}^2}{2\sigma_k^2} \right] \right\}$$

$$\frac{\partial l}{\partial \sigma_k} = \sum_{t=p+1}^{n} \frac{z_{k,t}}{\sigma_k} - \left[- \frac{z_{k,t}\, \varepsilon_{k,t}^2}{2\sigma_k^3} (-2) \right]$$

$$= \sum_{t=p+1}^{n} \frac{z_{k,t}}{\sigma_k} - \left(\frac{z_{k,t}\, \varepsilon_{k,t}^2}{\sigma_k^3} \right)$$

$$= \sum_{t=p+1}^{n} \frac{z_{k,t}}{\sigma_k} \left(\frac{\varepsilon_{k,t}^2}{\sigma_k^2} - 1 \right) \quad k = 1, \ldots, K$$

Suppose $\theta(\alpha, \phi, \sigma)$ is known.

$$Z \leftarrow E[Z \mid Y; \theta]$$

$$z_{k,t} \leftarrow E[z_k \mid Y; \theta] = P(y_i \in k \mid Y, \theta)$$

$$= P(z_{k,t} = 1 \mid y_t, \theta)$$

$$= \tau_{k,t}$$

$$Q(\theta \mid \theta_0, x) = E_{\theta_0}[\log L^c(\theta \mid x, z)]$$

$$= E_{\theta_0}[\log f(y_t, z_t \mid \mathcal{F}_{t-1}; \theta)]$$

$$Q = E_{Z_t|Y,\theta}\left[\sum_{k=1}^{K} Z_{k,t} \left[\log \alpha_k + \log f_k(y_t \mid \mathcal{F}_{t-1})\right] \right]$$

$$Q = E_{Z_t|Y,\theta}\left[\sum_{k=1}^{K} Z_{k,t} \left[\log \alpha_k - \frac{1}{2}\log(2\pi) - \log \sigma_k - \frac{(y_t - \mu_{k,t})^2}{2\sigma_k^2}\right] \right]$$

PY Suppose $\theta(\alpha, \phi, \sigma)$ is known. E-STEP

$$Q = \sum_{k=1}^{K} E[Z_k \mid Y;\theta] \left[\log \alpha_k - \frac{1}{2}\log(2\pi) - \log \sigma_k - \frac{(y_t - \mu_{k,t})^2}{2\sigma_k^2}\right]$$

$$= \sum_{k=1}^{K} \tau_{k,t} \log \alpha_k + \sum_{k=1}^{K} \tau_{k,t} \log f_k(y_t \mid \mathcal{F}_{t-1})$$

Note that Q depends on $\tau_{k,t}$.

$\therefore$ our E-step should be $\tau_{k,t}$.

$\tau_{k,t} = P(Z_{k,t} = 1 \mid y_t;\theta)$, observing exact value $\rightarrow$ PDF

$$= \frac{f(y_t, Z_{k,t}=1 \mid \theta)}{f(y_t \mid \theta)}, \quad \frac{\text{Joint distribution}}{\text{Marginal distribution}}$$

$$= \frac{f(y_t \mid Z_{k,t}=1,\theta)\, P(Z_{k,t}=1 \mid \theta)}{\sum_{k=1}^{K} f(y_t \mid Z_{k,t}=1,\theta)\, P(Z_{k,t}=1 \mid \theta)}, \quad P(Z_{k,t}=1 \mid \theta) = \alpha_k \text{ by design of mixture models.}$$

$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\quad \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ $\quad f(x) = \frac{1}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)$

$$= \frac{\alpha_k (1/\sigma_k)\, \phi(\varepsilon_{k,t}/\sigma_k)}{\sum_{k=1}^{K} \alpha_k (1/\sigma_k)\, \phi(\varepsilon_{k,t}/\sigma_k)} \quad \text{where } k = 1, \ldots, K$$

M-Step

Assume Z is known

$$\frac{\partial \mathcal{L}}{\partial \alpha_k} = \sum_{t=p+1}^{n} \left(\frac{z_{t,k}}{\alpha_k} - \frac{z_{t,K}}{\alpha_K} \right) = 0 \quad \text{where } k = 1, \dots, K-1$$

$$\frac{\sum_{t=p+1}^{n} \tau_{t,k}}{\alpha_k} - \frac{\sum_{t=p+1}^{n} \tau_{t,K}}{\alpha_K} = 0,$$

$$\boxed{\text{Let } \sum_{t=p+1}^{n} \tau_{t,k} = N_k \qquad \sum_{t=p+1}^{n} \tau_{t,K} = N_K}$$

$$\alpha_k = \frac{N_k}{N_K} \alpha_K$$

$$\sum_{k=1}^{K-1} \alpha_k + \alpha_K = 1, \qquad \boxed{\alpha_1 + \alpha_2 + \alpha_3 + \dots + \alpha_K = 1}$$

$$\sum_{k=1}^{K-1} \frac{N_k}{N_K} \alpha_K + \alpha_K = 1$$

$$\alpha_K \left(\sum_{k=1}^{K-1} \frac{N_k}{N_K} + 1 \right) = 1, \qquad \boxed{1 = \frac{N_K}{N_K}}$$

$$\alpha_K \left(\frac{1}{N_K} \sum_{k=1}^{K} N_k \right) = 1$$

$$\alpha_K = \frac{N_K}{\sum_{k=1}^{K} N_k}$$

$$\alpha_k = \frac{N_k}{\sum_{p=1}^{K} N_p}, \qquad \underbrace{\sum_{p=1}^{K} \sum_{t=p+1}^{n}}_{\text{finite sums are commutative}} \tau_{t,K} = \sum_{t=p+1}^{n} \underbrace{1}_{\sum_{p=1}^{K} P(Z_{t,k}=1 \mid y_t; \theta)}$$

$$\alpha_k = \frac{N_k}{\sum_{t=p+1}^{n} 1}$$

$$\therefore \hat{\alpha}_k = \frac{\sum_{t=p+1}^{n} \tau_{t,k}}{n-p} \qquad k = 1, \dots, K$$

M-Step   Assume Z is known

$$\frac{\partial \ell}{\partial \sigma_k} = \sum_{t=p+1}^{n} \frac{Z_{t,k}}{\sigma_k}\left(\frac{\varepsilon^2_{t,k}}{\sigma^2_k} - 1\right) \qquad k=1,\dots,K$$

where $\varepsilon_{t,k} = y_t - \phi_{k,0} - \phi_{k,1}y_{t-1} - \dots - \phi_{k,p_k} y_{t-p_k}$

$$\sum_{t=p+1}^{n} \frac{\tau_{t,k}}{\sigma_k}\left(\frac{\varepsilon^2_{t,k}}{\sigma^2_k} - 1\right) = 0$$

$$\frac{1}{\sigma_k^2}\sum_{t=p+1}^{n} \tau_{t,k}\,\varepsilon^2_{t,k} = \sum_{t=p+1}^{n} \tau_{t,k}$$

$$\hat{\sigma}_k = \left(\frac{\sum_{t=p+1}^{n} \tau_{t,k}\,\varepsilon^2_{t,k}}{\sum_{t=1}^{n} \tau_{t,k}}\right)^{\frac{1}{2}}$$

$$\frac{\partial \ell}{\partial \phi_{k,i}} = \sum_{t=p+1}^{n} \frac{Z_{t,k}\,u(y_t,i)\,\varepsilon_{t,k}}{\sigma^2_k} \qquad \begin{array}{l} k=1,\dots,K \\ \text{where } i=0,\dots,p_k \end{array}$$

$$\sum_{t=p+1}^{n} \tau_{t,k}\,u(y_t,i)\,\varepsilon_{t,k} = 0 \qquad \boxed{u(y_t,i) = \begin{cases} 1 & \text{if } i=0 \\ y_{t-i} & \text{if } i>0 \end{cases}}$$

$$\sum_{t=p+1}^{n} \tau_{t,k}\,u(y_t,i)\,y_t = \sum_{t=p+1}^{n} \tau_{t,k}\,u(y_t,i) \sum_{j=0}^{p_k} \hat{\phi}_{k,j}\,u(y_t,j)$$

$$\sum_{t=p+1}^{n} \tau_{t,k}\,u(y_t,i)\,y_t = \sum_{j=0}^{p_k} \hat{\phi}_{k,j} \sum_{t=p+1}^{n} \tau_{t,k}\,u(y_t,j)\,u(y_t,i)$$

where $i=0,\dots,p_k$

$$\sum_{t=p+1} \tau_{t,k}\, u(y_t, i)\, y_t = \sum_{j=0}^{p_k} \hat{\phi}_{k,j} \sum_{t=p+1} \tau_{t,k}\, u(y_t, j)\, u(y_t, i)$$

where $i = 0, \dots, p_k$

$$\underbrace{\begin{bmatrix} \sum \tau y_t u_0 \\ \sum \tau y_t u_1 \\ \vdots \\ \sum \tau y_t u_{p_k} \end{bmatrix}}_{b_k} = \underbrace{\begin{bmatrix} \sum \tau u_0 u_0 & \sum \tau u_1 u_0 & \dots & \sum \tau u_{p_k} u_0 \\ \sum \tau u_0 u_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \sum \tau u_0 u_{p_k} & \dots & \dots & \sum \tau u_{p_k} u_{p_k} \end{bmatrix}}_{A_k} \underbrace{\begin{bmatrix} \hat{\phi}_{k_0} \\ \vdots \\ \hat{\phi}_{k p_k} \end{bmatrix}}_{\hat{\phi}_k}$$

<u>Note</u>

τ is the probability of observing y_t in cluster k.

$u_0 = u(y_t, 0) = 1$ & $u_1 = u(y_t, 1) = y_{t-1}$

$\hookrightarrow \therefore$ weighted least squares. $A_k \hat{\phi}_k = b_k$

$$A_k = X_k^T W_k X_k \qquad b_k = X_k^T W_k Y$$

$$X_k = \begin{bmatrix} 1 & y_{t-1} & y_{t-2} & \dots & y_{t-p_k} \\ 1 & y_t & y_{t-1} & \dots & y_{t+1-p_k} \\ 1 & y_{n-1} & y_{n-2} & \dots & y_{n-p_k} \end{bmatrix} \rightarrow n-p \text{ rows}, \; p_k+1 \text{ cols}$$

$$Y = [y_t \;\; y_{t+1} \;\; \dots \;\; y_n] \qquad W_k = \mathrm{diag}(\tau_{t,k})$$

where $t = p+1, p+2, \dots, n$

$$\boxed{\hat{\phi}_k = A_k^{-1} b_k}$$

Log-Likelihood

$$f(y_t \mid \mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi}\,\sigma_k} e^{-\frac{(y_t-\mu_{t,k})^2}{2\sigma_k^2}}$$

$$L(\theta) = \prod_{t=p+1}^{n} \left[\sum_{k=1}^{K} \alpha_k \frac{1}{\sqrt{2\pi}\,\sigma_k} e^{-\frac{(y_t-\mu_{t,k})^2}{2\sigma_k^2}} \right]$$

$$\ell(\theta) = \sum_{t=p+1}^{n} \log \left[\sum_{k=1}^{K} \alpha_k \frac{1}{\sqrt{2\pi}\,\sigma_k} e^{-\frac{(y_t-\mu_{t,k})^2}{2\sigma_k^2}} \right]$$

$$= \sum_{t=p+1}^{n} \log \left[\sum_{k=1}^{K} \alpha_k \frac{1}{\sigma_k} e^{-\frac{(y_t-\mu_{t,k})^2}{2\sigma_k^2}} \right] + \boxed{\log(2\pi)^{-\frac{n-p}{2}}}$$

Constant can be removed.

$$\therefore \ell(\theta) = \sum_{t=p+1}^{n} \log \left[\sum_{k=1}^{K} \alpha_k \frac{1}{\sigma_k} e^{-\frac{\varepsilon_{t,k}^2}{2\sigma_k^2}} \right]$$

$$\varepsilon_t = \Delta y_t - \phi_0 - \phi_1 \Delta y_{t-1} - \phi_2 \Delta y_{t-2}$$

$$= y_t - y_{t-1} - \phi_0 - \phi_1(y_{t-1} - y_{t-2}) - \phi_2 (y_{t-2} - y_{t-3})$$

$$= y_t - \phi_0 - y_{t-1} - \phi_1 y_{t-1} + \phi_1 y_{t-2} - \phi_2 y_{t-2} + \phi_2 y_{t-3}$$

$$= y_t + (-1 - \phi_1) y_{t-1} + (\phi_1 - \phi_2) y_{t-2} + \phi_2 y_{t-3} - \phi_0$$

Let $\phi_{11} = -0.3208$, $\phi_{21} = 0.6711$ $\quad \phi_{0k} = 0$

$$\varepsilon_{t,1} = y_t - \phi_{01} - (1 + \phi_{11}) y_{t-1} + \phi_{11} y_{t-2}$$

$$= y_t - 0.6792\, y_{t-1} - 0.3208\, y_{t-2}$$

$$\varepsilon_{t,2} = y_t - 1.6711\, y_{t-1} + 0.6711\, y_{t-2}$$

$$\varepsilon_{t,3} = \Delta y_t - \phi_0$$

$$= y_t - y_{t-1}$$