ARMA(m, n)

Eq.: $y_t = \sum_{i=1}^{m} \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{n} \theta_j \varepsilon_{t-j}$, $\varepsilon_t \overset{\perp\!\!\!\perp}{\sim} \mathcal{N}(0, \sigma^2)$

Stationarity: $|\phi_i| < 1 \quad \forall i$

Invertibility: $|\theta_j| < 1 \quad \forall j$

Forecast: $\hat{y}_{t+h} := E(y_{t+h} | F_t)$
$= \sum_{i=1}^{m} \phi_i \hat{y}_{t+h-i|t} + \sum_{j=1}^{n} \theta_j \hat{\varepsilon}_{t+h-j|t}$, where $\hat{y}_{t+h-i|t} = \begin{cases} y_{t+h-i} & \text{for } h \le i \\ \hat{y}_{t+h-i} & \text{for } h > i \end{cases}$, $\hat{\varepsilon}_{t+h-j|t} = \begin{cases} \varepsilon_{t+h-j} & \text{for } h \le j \\ 0 & \text{for } h > j \end{cases}$

Forecasting Error: $\varepsilon_{t+h} = y_{t+h} - \hat{y}_{t+h}$

$Var(\varepsilon_{t+h}) = \sigma^2 \sum_{k=0}^{h-1} \psi_k^2$, where $\psi_k = \sum_{i=1}^{\min(k,m)} \phi_i \psi_{k-i} + \theta_k$ with $\psi_0 = 1$

Proof: MA(∞) representation: $y_t = \mu_t + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$ with $\psi_0 = 1$.

$$\begin{cases} y_{t+h} = \mu_{t+h} + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t+h-k} \\ \hat{y}_{t+h} = \mu_{t+h} + \sum_{k=h}^{\infty} \psi_k \varepsilon_{t+h-k} \end{cases} \Rightarrow \varepsilon_{t+h} = \sum_{k=0}^{h-1} \psi_k^2 \varepsilon_{t+h-k} = \sigma^2 \sum_{k=0}^{h-1} \psi_k^2 \quad \text{(Independent)}$$

Forecast Convergence: $\lim_{h\to\infty} \hat{y}_{t+h} = \mu_{t+h}$

Proof: $\hat{y}_{t+h} = \mu_{t+h} + \sum_{k=h}^{\infty} \psi_k \varepsilon_{t+h-k}$ and $\lim_{h\to\infty} \varepsilon_{t+h-k} = 0$

ADF test:

$$\Delta x_t = \underset{\substack{\uparrow \\ \text{drift}}}{\alpha} + \underset{\substack{\uparrow \\ \text{trend}}}{\beta t} + \gamma x_{t-1} + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + \varepsilon_t$$

$H_0: \gamma = 0$ (Unit root) against $H_1: \gamma < 0$

t-statistic $= \frac{\hat{\gamma}}{\sqrt{Var(\hat{\gamma})}}$: Reject H_0 if $\hat{t} < t_\alpha^*$ (i.e. -2.86)

p: Selected by ICs

Parameter Estimation ARIMA(p, d, q)

Moment Estimators	All models
Yule Walker Estimators	AR(p)
Least Squares Estimators	AR(p)
Conditional Least Squares Estimators	MA(q) or ARMA(p, q)
MLE	All models

Statistical Inference

Denote $M(\theta) = \left(\frac{\partial \mathcal{L}(\theta)}{\partial \theta_1}, \dots, \frac{\partial \mathcal{L}(\theta)}{\partial \theta_K}\right)^T$

$M(\hat{\theta}_{MLE}) = 0$ is unbiased.

$\sqrt{n}(\hat{\theta}_{MLE} - \theta^*) \approx -\sqrt{n}\,(\nabla_\theta M(\theta^*))^{-1} M(\theta^*) \sim \mathcal{N}(0, \Sigma)$

Order Selection:

Final Prediction Error $= \hat{\sigma}^2 \left(\frac{n+p}{n-p}\right)$ AR(p) only

AIC: $-2 \log \mathcal{L}(\theta) + 2(p+q+1)$

AICC: $-2 \log \mathcal{L}(\theta) + \frac{2(p+q+1)}{n-p-q-2}$

BIC: $(n-p-q) \log\left(\frac{n\hat{\sigma}^2}{n-p-q}\right) + n(1 + \log(\sqrt{2\pi})) + (p+q) \log\left(\frac{\sum_i x_i^2 - n\hat{\sigma}^2}{p+q}\right)$

Note: BIC is consistent, $\lim_{n\to\infty} P(\hat{\theta}_{BIC} = \theta^*) = 1$

Residual Analysis:

Residual $\sim$ White Noise

ACF plot

Portmanteau Statistics: $Q(h) = n(n+2) \sum_{i=1}^{h} \frac{r^2(i)}{n-j} < \chi_\alpha^2(h-p-q)$

GARCH(p,q)

Eq: $\varepsilon_t = \sigma_t z_t$, $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$, $z_t \overset{iid}{\sim} N(0,1)$

Positivity: $\omega > 0$, $\alpha_i \geq 0 \ \forall i$, $\beta_j \geq 0 \ \forall j$

Stationarity: $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$

LT-variance: $\sigma_{LT}^2 = \dfrac{\omega}{1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j}$

Forecast: $\hat{\sigma}_{t+h}^2 = \omega + \sum_{i=1}^{p} \alpha_i \hat{\varepsilon}_{t+h-i}^2 + \sum_{j=1}^{q} \beta_j \hat{\sigma}_{t+h-j}^2$, where $\hat{\varepsilon}_{t+h-i}^2 = \begin{cases} \varepsilon_{t+h-i}^2 & \text{for } h \leq i \\ \hat{\sigma}_{t+h-i}^2 & \text{for } h > i \end{cases}$

GARCH(1,1): $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$

$\hat{\sigma}_{t+1}^2 = \omega + \alpha \varepsilon_t^2 + \beta \sigma_t^2$

Unconditional Kurtosis: $\dfrac{\mathbb{E}(\varepsilon_t^4)}{\mathbb{E}(\varepsilon_t^2)^2} = \dfrac{3(1-\omega^2)}{1-2\alpha^2-\omega^2} \geq 3$

Parameter Estimation: $\varepsilon_t = \sigma_t z_t \sim N(0, \sigma_t^2) \Rightarrow f(\varepsilon_t) = \dfrac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\dfrac{\varepsilon_t^2}{2\sigma_t^2}\right)$

$\Rightarrow \mathcal{L}(\omega, \alpha, \beta) = -\sum_{i=1}^{n} \left(\ln(\sigma_t^2) + \dfrac{\varepsilon_t^2}{\sigma_t^2} \right)$

EGARCH(p,q)

Eq.: $\varepsilon_t = \sigma_t z_t$, $z_t \overset{iid}{\sim} N(0,1)$, $\ln(\sigma_t^2) = \omega + \sum_{i=1}^{p} \left(\alpha_i [|z_{t-i}| - E(|z_{t-i}|)] + \gamma_i z_{t-i} \right) + \sum_{j=1}^{q} \beta_j \ln(\sigma_{t-j}^2)$

$\gamma < 0$ (Leverage effect)

Stationarity: $\sum_{j=1}^{q} |\beta_j| < 1$

LT-variance: $\ln(\sigma_{LT}^2) = \dfrac{\omega}{1 - \sum_{j=1}^{q}\beta_j}$

EGARCH(1,1): $\ln(\sigma_t^2) = \omega + \alpha(\theta z_{t-1} + \gamma |z_{t-1}|) + \beta \ln(\sigma_{t-1}^2)$

$\ln(\hat{\sigma}_{t+1}^2) = \omega + \alpha(\theta z_t + \gamma |z_t|) + \beta \ln(\sigma_t^2)$

Heteroskedasticity Test

1) ACF plot on $\hat{\varepsilon}_t^2$

2) Ljung-Box test: $Q = n(n+2) \sum_{k=1}^{h} \dfrac{\hat{\rho}_k^2}{n-k}$, where $\hat{\rho}_k$: Sample auto-correlation at lag k

H_0: No auto-correlation up to lag h. Reject H_0 if $Q > \chi^2_{\alpha,h}$

ARMA(m,n) - EGARCH(1,1)

Eq.: $r_t = \sum_{i=1}^{m} \phi_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{n} \theta_j \varepsilon_{t-j}$

$\varepsilon_t = \sigma_t z_t$, $z_t \overset{iid}{\sim} N(0,1)$

$\ln(\sigma_t^2) = \omega + \alpha[|z_{t-1}| - E(|z_{t-1}|)] + \gamma z_{t-1} + \beta \ln(\sigma_{t-1}^2)$

Mean stationary: $|\phi_i| < 1 \quad \forall i$

Variance stationary: $|\beta| < 1$

Note: MA(n) target auto-corr. in mean of ε_t vs EGARCH target auto-corr in variance of ε_t

Scaling: If $\varepsilon_t \approx 0$, $\ln(\sigma_t^2)$ maybe unstable (Overflow / Gradient vanish & explode)

$r_{scaled,t} = \frac{r_t}{s_t}$, where s_t is rolling sample s.d.

Forecasting

Conditional Mean: $\hat{y}_{t+h} = \sum_{i=1}^{m} \phi_i \hat{y}_{t+h-i|t} + \sum_{j=1}^{n} \theta_j \hat{\varepsilon}_{t+h-j|t}$

Conditional Variance: $\ln(\hat{\sigma}_{t+1}^2) = \omega + \alpha[|z_t| - E(|z_t|)] + \gamma z_t + \beta \ln(\sigma_t^2)$

$\hat{\sigma}_{t+1}^2 \approx \exp[\ln(\hat{\sigma}_{t+1}^2)]$ (Jensen's ineq.)

Simulation Algo: Simulate $\{z_{t+1,i}\}_{i=1}^{n}$, then $\{\ln(\hat{\sigma}_{t+1,i}^2)\}_{i=1}^{n}$ hence $E(\hat{\sigma}_{t+1}) = \frac{1}{n}\sum_{i=1}^{n} \ln(\hat{\sigma}_{t+1,i}^2)$

Parameter Estimation

Joint MLE: $\Theta = (\phi_{1:m}, \theta_{1:n}, \omega, \alpha, \beta, \gamma)$

$f(r_t|\mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right)$

$l_t(\Theta) \propto -\frac{1}{2}\ln(\hat{\sigma}_t^2) - \frac{\hat{\varepsilon}_t^2}{2\hat{\sigma}_t^2}$

$\mathcal{L}(\Theta) = \sum_{t=\max(p,q)+1}^{T} l_t(\Theta) \triangleq -\frac{1}{2}\sum_{t=t}^{T}\left(\ln(\hat{\sigma}_t^2(\Theta)) + \frac{\hat{\varepsilon}_t^2(\Theta)}{\hat{\sigma}_t^2(\Theta)}\right)$

Recursion:

Initiation: $(\hat{\phi}_{1:m,0}, \hat{\theta}_{1:n,0})$ by ARMA(m,n) estimation

$(\omega_0, \alpha_0, \beta_0, \gamma_0)$ by EGARCH(1,1) estimation

$\hat{\sigma}_0^2 = \frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})^2$, $\hat{\varepsilon}_{t\,|\,unknown} = 0$

* Different initialization leads to different $\hat{\Theta}_{MLE}$

for $t = \max(p,q)+1$ to T: $O(T)$ per iteration

1) $\hat{\varepsilon}_t = r_t - \sum_{i=1}^{m}\phi_i r_{t-i} - \sum_{j=1}^{n}\theta_j \varepsilon_{t-j}$
2) $\hat{z}_{t-1} = \frac{\hat{\varepsilon}_{t-1}}{\hat{\sigma}_{t-1}}$
3) $\ln(\hat{\sigma}_t^2) = \omega + \alpha[|\hat{z}_{t-1}| - E(|z_{t-1}|)] + \gamma \hat{z}_{t-1} + \beta\ln(\hat{\sigma}_{t-1}^2)$
4) $l_t(\Theta) \propto -\frac{1}{2}\ln(\hat{\sigma}_t^2) - \frac{\hat{\varepsilon}_t^2}{2\hat{\sigma}_t^2}$

$\hat{\Theta}_{MLE} = \arg\max_{\Theta} \mathcal{L}(\Theta)$ by BFGS

Consistency:

- ARMA process is causal and invertible, and no common roots
- EGARCH(1,1) has $|\beta| < 1$
- Strictly stationary & Ergodic
- $\hat{\Theta}_{MLE} \in \Theta$ and Θ is compact ($\alpha, \gamma > 0$ and $E[\ln(\sigma_t^2)] < \infty$)
- $E(r_t^4) < \infty$
- $\sigma_t^2 > 0$ a.s. and $\inf_{\Theta} \sigma_t^2 > 0$

By Ergodic thm., $\sup_{\theta\in\Theta}\left|\frac{1}{n}\mathcal{L}(\Theta) - \tilde{l}(\Theta)\right| \xrightarrow{a.s.} 0$, where $\tilde{l}(\Theta) = \mathbb{E}\left[-\frac{1}{2}\left(\ln(\hat{\sigma}_t^2(\Theta)) + \frac{\hat{\varepsilon}_t^2(\Theta)}{\hat{\sigma}_t^2(\Theta)}\right)\right]$

Asymptotic properties:

$\sqrt{n}(\hat{\Theta}_{MLE} - \Theta_0) \sim N(0, \Sigma)$ where $\Sigma = J^{-1} I J^{-1}$ and $\begin{cases} I = \mathbb{E}\left(\frac{\partial\mathcal{L}(\Theta_0)}{\partial\Theta}\frac{\partial\mathcal{L}(\Theta_0)}{\partial\Theta^T}\right) \\ J = \mathbb{E}\left(\frac{\partial^2\mathcal{L}(\Theta_0)}{\partial\Theta\,\partial\Theta^T}\right)\end{cases}$ (4^{th} moment assumption)

* Asymptotic accuracy of EGARCH estimators isn't affected by the presence of ARMA

Adv: 1) Asymptotic efficient
2) Align with DGP

Disadv: 1) Time complexity
2) Initial guess affects convergence

Two-step Estimation: Fit ARMA(m,n) to r_t $\longrightarrow$ Fit EGARCH(1,1) to ε_t

Problem: EGARCH estimates inherit bias from ARMA specification

Experiment setup:

Data input

$\downarrow$

ADF Test

$\downarrow$ Stationary

ARMA(m,n) $\Rightarrow \{ \varepsilon_{ARMA}, \tilde{\phi}_{1:m}, \tilde{\theta}_{1:n} \}$

$\downarrow$

Heteroskedasticity Test

$\downarrow$ Yes

Estimate $\tilde{\Theta}_{MLE}$

$\downarrow$

Goodness of fit $\begin{cases} \text{res. and res.}^2 \text{ have no auto-corr} \\ \text{res.} \sim \text{Normal} \end{cases}$