ARMA Processes

Definition of ARMA(1,1) process

A time series $\{X_t\}$ is an ARMA$(1,1)$ process if it is stationary and satisfies:

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad Z_t \sim WN(0, \sigma^2)$$

where ϕ (autoregressive coefficient) and θ (moving average coefficient) are constants while keeping $\phi \neq \theta$ to avoid redundancy in the model.

Using the **backshift operator** $B(BX_t = X_{t-1})$, this becomes:

$$(1 - \phi B)X_t = (1 + \theta B)Z_t \quad \textbf{or} \quad \phi(B)X_t = \theta(B)Z_t$$

where $\phi(B) = 1 - \phi B$ and $\theta(B) = 1 + \theta B$.

Case $|\phi| < 1$:

The operator $\phi(B)$ is invertible. Its inverse is:

$$\chi(B) = \frac{1}{1 - \phi B} = \sum_{j=0}^{\infty} \phi^j B^j$$

which converges absolutely for $|\phi| < 1$. On applying $\chi(B)$ to both sides of $\phi(B)X_t = \theta(B)Z_t$, we get :

$$X_t = \chi(B)\theta(B)Z_t = \left(\sum_{j=0}^{\infty} \phi^j B^j\right)(1 + \theta B)Z_t$$

Expanding this will give us:

$$X_t = Z_0 + (\phi + \theta)\sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j}$$

Using the same approach that we used in AR(1) process, we can see that this is unique stationary solution. The solution is **casual** as X_t depends only on current and past noise terms $\{Z_t - j, j \geq 0\}$.

Case $|\phi| > 1$:

Now, we will rewrite $\phi(B)^{-1}$ using forward shifts (negative powers of B):

$$\frac{1}{\phi(B)} = -\sum_{j=1}^{\infty} \phi^{-j} B^{-j}$$

On applying this operator to $\theta(B)Z_t$, we get:

$$X_t = -\theta\phi^{-1}Z_t - (\theta+\phi)\sum_{j=1}^{\infty}\phi^{-j-1}Z_{t+j}$$

We can use the same logic like the previous case as here $|\phi^{-1}| < 1$ and so stationary solution exists but it is **non-casual** as X_t depends on future noise terms $\{Z_{t+j}, j \geq 1\}$, making it impractical for real-world applications.

Case $\phi = \pm 1$: Following the same approach, our model will lead to unbounded variance over time and hence it won't be stationary.

ARMA$(1,1)$ **has a unique stationary solution if** $|\phi| \neq 1$:

$$\textbf{For } |\phi| < 1: \quad X_t = Z_0 + (\phi+\theta)\sum_{j=1}^{\infty}\phi^{j-1}Z_{t-j} \textbf{ (casual)}$$

$$\textbf{For } |\phi| > 1: \quad X_t = -\theta\phi^{-1}Z_t - (\theta+\phi)\sum_{j=1}^{\infty}\phi^{-j-1}Z_{t+j} \textbf{ (non-casual)}$$

$$\textbf{For } |\phi| = 1: \textbf{ No stationary solution exists}$$

Invertibility:

A process is invertible if the white noise term Z_t can be expressed as a function of current and past observations $X_s, s \leq t$. This allows recovering the noise sequence from observed data. Invertibility depends on the moving average coefficient θ, analogous to how causality depends on the autoregressive coefficient ϕ. Just like we do calculation of casuality in the case of ϕ, we can do the same for θ and it is straight forward to show that:

ARMA$(1,1)$ **is invertible if and only if** $|\theta| < 1$ **(noise depends on past** X_t**). If** $|\ \theta\ | > 1$**, it is non-invertible (noise depends on future** X_t**). Invertibility ensures recoverability of noise terms for practical modeling.**

Noncausal/noninvertible ARMA$(1,1)$ **processes can be reparameterized using a new white noise sequence** $\{W_t\}$ **to become causal and invertible without losing statistical properties (mean, variance, autocovariance). Thus, focusing on causal/invertible models suffices for analysis, even for higher-order ARMA models.**

Properties of the sample mean and autocorrelation function

Estimation of Sample mean μ

Given a stationary process $\{X_t\}$, we can compute the sample mean as:

$$\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

It will be an unbiased estimator of μ since

$$E[\bar{X_n}] = \frac{1}{n}(E(X_1) + \cdots + E(X_n)) = \mu$$

The variance (or mean squared error, as the estimator is unbiased) is:

$$Var(\bar{X}_n\ ) = E[(\bar{X}_n\ - \mu)^2] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)$$

For a stationary process, $Cov(X_i, X_j) = \gamma(i-j)$, where $\gamma(h)$ is the autocovariance at lag h. Let $h = i - j$. The double sum becomes:

$$Var(\bar{X}_n) = \frac{1}{n^2}\sum_{h=-n+1}^{n-1}(n - |h|)\gamma(|h|)$$

For each lag h, there are $n - |h|$ pairs (i, j) such that $i - j = h$. We now adjust the summation limits by extending the sum to include $h = \pm n$ (terms vanish because $n - |h| = 0$):

$$Var(\bar{X}_n) = \frac{1}{n^2}\sum_{h=-n}^{n}(n - |h|)\gamma(|h|)$$

Therefore:

$$Var(\bar{X}_n) = \frac{1}{n}\sum_{h=-n}^{n}(1 - \frac{|h|}{n})\gamma(|h|)$$

The term $(1 - \frac{|h|}{n})$ acts as a weight for each lag h, decaying linearly with $|h|$. Lags closer to 0 contribute more due to more observation pairs, while distant lags contribute less.

Proposition 3

If $\{X_t\}$ is a stationary time series with mean μ and autocovariance function $\gamma(\cdot)$, then as $n \to \infty$,

$$Var(\bar{X}_n) = \mathbb{E}(\bar{X}_n - \mu)^2 \longrightarrow 0 \quad \text{if } \gamma(n) \to 0$$

$$n\mathbb{E}(\bar{X}_n - \mu)^2 \longrightarrow \sum_{h=-\infty}^{\infty} \gamma(h) \quad \text{if } \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$$

For large n, under Gaussianity or linearity (e.g., ARMA models), the Central Limit Theorem for dependent data applies:

$$\sqrt{n}\,(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, v)$$

where

$$v = \sum_{h=-\infty}^{\infty} \gamma(h)$$

is the long-run variance, capturing the cumulative effect of autocorrelations. This leads to:

$$\bar{X}_n \approx \mathcal{N}(\mu, \frac{v}{n})$$

An approximate 95% confidence interval for μ is:

$$\left(\bar{X}_n - 1.96\sqrt{\frac{\hat{v}}{n}}, \bar{X}_n + 1.96\sqrt{\frac{\hat{v}}{n}}\right)$$

where $\hat{v}$ estimates v. The estimator $\hat{v}$ accounts for autocovariances up to lag $\sqrt{n}$ with **Bartlett weights**:

$$\hat{v} = \sum_{|h|<\sqrt{n}} \left(1 - \frac{|h|}{n}\right) \hat{\gamma}(h)$$

Truncation at $\sqrt{n}$ balances bias (including sufficient lags) and variance (excluding noisy estimates at large lags). These weights $(1 - \frac{|h|}{n})$ reduces the influence of higher lags, akin to the Bartlett kernel in HAC estimators.

Estimation of Autocovariance $\gamma(\cdot)$ and Autocorrelation $\rho(\cdot)$ Functions

We will compute ACVF $\hat{\gamma}(h)$ and ACF $\hat{\rho}(h)$ as:

$$\hat{\gamma}(h) = n^{-1} \sum_{i=1}^{n-|h|} (X_{i+|h|} - \bar{X}_n)(X_i - \bar{X}_n)$$

and

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

these estimators quantify the covariance and correlation between observations at lag h. While $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ are biased (even if normalized by $n-|h|$ instead of n), they become nearly unbiased for large n. A critical property of $\hat{\gamma}(h)$ is that the k-dimensional sample covariance matrix:

$$\hat{\Gamma}_k = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(k-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(k-1) & \hat{\gamma}(k-2) & \cdots & \hat{\gamma}(0) \end{bmatrix}$$

is **nonnegative definite**. This ensures valid covariance structures, as negative eigenvalues would imply implausible negative variances.

The matrix $\hat{\Gamma}_k$ is constructed using a transformation matrix T of size $k \times 2k$, where each row contains lagged deviations $Y_i = X_i - \bar{X}_n$ (with $Y_i = 0$ for $i > n$). For any real vector a:

$$a'\hat{\Gamma}_k a = n^{-1}(a'T)(T'a) \geq 0$$

which guarantees nonnegative definiteness.

Using n^{-1} (instead of $(n-|h|)^{-1}$) preserves nonnegative definiteness but introduces bias. For example, the divisor n ensures structural validity of $\hat{\Gamma}_k$, while $(n-|h|)^{-1}$ adjusts for reduced pairs at larger lags but risks invalid covariance matrices.

The normalized matrix (Sample Autocorrelation Matrix) :

$$\hat{R}_k = \frac{\hat{\Gamma}_k}{\hat{\gamma}(0)}$$

The estimators $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ balance bias and structural integrity. While slightly biased, their design ensures covariance/correlation matrices remain valid, enabling reliable statistical inference in time series analysis.

Sample ACVF and ACF estimators prioritize nonnegative definiteness over bias reduction.Using n^{-1} ensures valid covariance matrices, critical for model identification and inference.

The matrices $\hat{\Gamma}_k$ and $\hat{R}_k$ are in fact nonsingular if there is at least one nonzero Y_i, or equivalently if $\hat{\gamma}(0) > 0$.

The standard methods suggest n should be atleast 50 and $h \leq n/4$.

For systematic inference concerning $\rho(h)$, understanding the sampling distribution of the sample ACF estimator $\hat{\rho}(h)$ is essential. While exact distributions are complex, asymptotic theory provides a practical approximation: for large sample sizes n, the vector $\hat{\rho}_k = (r\hat{h}o(1), \cdots, \hat{\rho}(k))'$ follows a multivariate normal distribution. Specifically, under linear model assumptions (e.g., ARMA models),

$$\hat{\rho}_k \approx N(\rho_k, n^{-1}W)$$

where $\rho_k = (\rho(1), \cdots, \rho(k))'$ contains the true autocorrelations, and W is the covariance matrix defined by Bartlett's formula. The (i, j)-th element of W is given by:

$$w_{ij} = \sum_{k=-\infty}^{\infty} \{\rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i)\}$$

This expression simplifies computationally to a finite sum:

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\}$$

The covariance structure W captures dependencies between autocorrelation estimates at different lags, which is vital for constructing confidence intervals and testing hypotheses (e.g., whether multiple autocorrelations jointly differ from zero).

Example 1: IID Noise

Consider a time series $\{X_t\} \sim IID(0, \sigma^2)$, where each observation has mean 0, variance σ^2, and no serial dependence. For such a process, the true autocorrelation function (ACF) satisfies:

$$\rho(h) = \begin{cases} 1 & \textbf{if} \quad h = 0 \\ 0 & \textbf{otherwise} \end{cases}$$

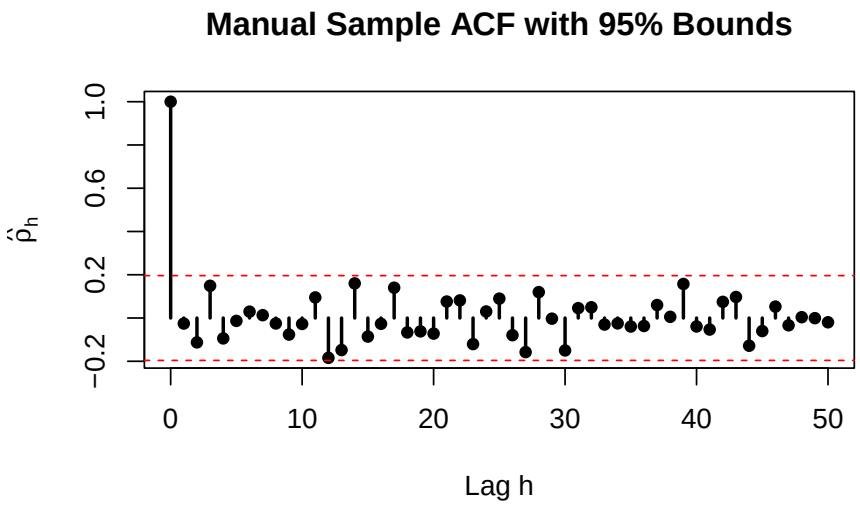
Since $\rho(k) = 0$ for all $k \neq 0$, every term in the sum vanishes except when $i = j$, leading to:

$$w_{ij} = \begin{cases} 1 & \textbf{if} \quad i = j \\ 0 & \textbf{otherwise} \end{cases}$$

Thus, W reduces to the identity matrix.

We take lags $h = 1, 2, \cdots, k$ on X-axis and values of $\hat{\rho}(h)$ on Y-axis. We draw two dashed lines at $y = \pm 1.96/n$. We now plot $\hat{\rho}(h)$ as vertical bars at each lag h.

```r
set.seed(123)
n     <- 100      # sample size
sigma <- 1        # noise sd
nlags <- 50       # number of lags to compute
X <- rnorm(n, mean = 0, sd = sigma)
Xbar  <- mean(X)
denom <- sum((X - Xbar)^2)
rho <- numeric(nlags + 1)
for (h in 0:nlags) {
  # numerator: covariance at lag h
  num <- sum((X[(1+h):n] - Xbar) * (X[1:(n-h)] - Xbar))
  rho[h+1] <- num / denom
}
bound <- qnorm(0.975) / sqrt(n)   # same as 1.96/sqrt(n)
plot(0:nlags, rho,
     type = "h",          # vertical lines
     lwd  = 2,
     xlab = "Lag h",
     ylab = expression(hat(rho)[h]),
     main = "Manual Sample ACF with 95% Bounds")
points(0:nlags, rho, pch = 16)    # add dots
abline(h = c(+bound, -bound),
       col = "red", lty = 2)
```

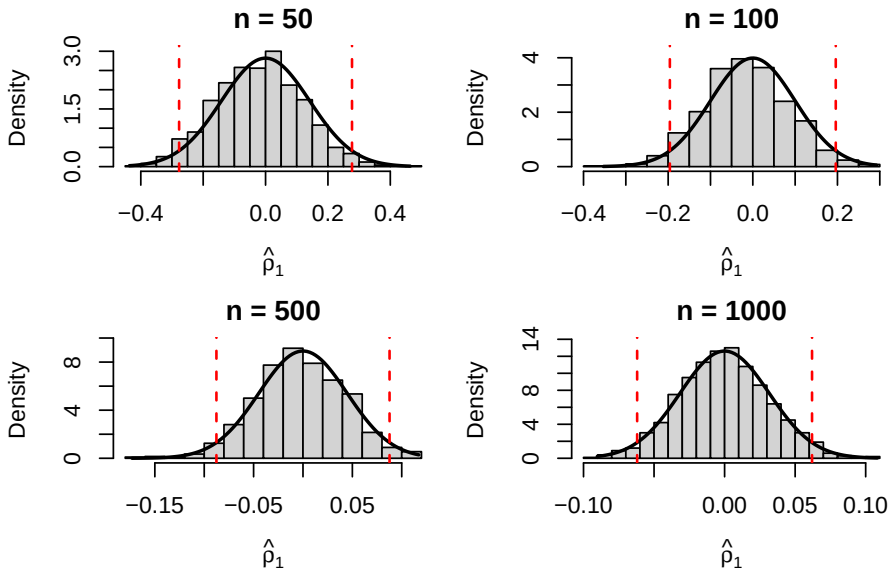


Approximately 95% of $\hat{\rho}(h)$ values lie within the confidence bounds. No systematic pattern (e.g., decaying trends or spikes) appeared.

We now see the verification that the sample ACF converges to the theoretical normal distribution as the number of points n in the time series increases.

```r
set.seed(123)
n_vec <- c(50, 100, 500, 1000)  # different series lengths
M     <- 1000                   # number of simulations per n
h     <- 1                      # lag to examine
old_par <- par(mfrow = c(2, 2), mar = c(4, 4, 2, 1))

for (n in n_vec) {
  # 3a. Simulate M series and compute rho_hat(h) manually
  rho_vals <- replicate(M, {
    x    <- rnorm(n)           # iid N(0,1)
    xbar <- mean(x)
    num  <- sum((x[(1+h):n] - xbar) * (x[1:(n-h)] - xbar))
    den  <- sum((x - xbar)^2)
    num / den
  })

  # 3b. Plot histogram of empirical rho
  hist(rho_vals,
       breaks = 15,
       freq   = FALSE,
       main   = paste0("n = ", n),
```

```
       xlab   = expression(hat(rho)[1]),
       ylim   = c(0, max(density(rho_vals)$y)*1.1))

  # 3c. Overlay theoretical N(0, 1/n) density
  curve(dnorm(x, mean = 0, sd = 1/sqrt(n)),
        from = min(rho_vals),
        to   = max(rho_vals),
        add  = TRUE,
        lwd  = 2)

  # 3d. Add ±1.96/√n bounds
  abline(v = c(-1.96/sqrt(n), 1.96/sqrt(n)),
         col = "red", lty = 2, lwd = 1.5)
}
```




```
# 4. Reset plotting parameters
par(old_par)
```

As $n \uparrow, \hat{\rho}(h)$ clusters closer to 0. The distribution becomes Gaussian. The spread of $\hat{\rho}(h)$ diminishes proportionally to $1/n$.

Example 2: An MA(1) Process

For an MA(1) process, the true autocorrelation function (ACF) satisfies:

$$\rho(h) = \begin{cases} 1 & \textbf{if } h = 0 \\ \frac{\theta}{1+\theta^2} & \textbf{if } h = \pm 1 \\ 0 & \textbf{if } |h| > 1 \end{cases}$$

GARCH Process

Let $\{r_t\}_{t\in\mathbb{Z}}$ be a (zero-mean) return series, or demeaned log-returns of an asset. We write

$$r_t = \epsilon_t$$

where the innovations ϵ_t decompose as:

$$\epsilon_t = \sigma_t z_t$$

with $z_t \sim IID(0,1)$ (often gaussian or student-t), $\sigma_t^2 = Var(\epsilon_t|\mathcal{F}_{t-1})$ is the conditional variance, and $\mathcal{F}_{t-1}$ is the $\sigma-$ algebra generated by $\{r_{t-1}, r_{t-2}, \cdots\}$.

Definition: GARCH(p,q)

A **GARCH** (p,q)model posits that

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

subject to

$$\alpha_0 > 0, \;\; \alpha_i \geq 0, \;\; \beta_j \geq 0, \;\; \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$$

α_i measure the immediate "ARCH" effects (past shocks), β_j capture the "GARCH" effects (persistence of past variances) and the stationary condition $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ ensures a finite unconditional variance.