1 Foundations

Time series data: $y_1, \dots, y_T$ observed from random variables $\{Y_t\}$. Index t typically records evenly spaced time. Goals: describe dynamics, diagnose structure, forecast Y_{T+h}, quantify forecast uncertainty. Random variables need not be independent; dependence structure guides how much information additional observations provide.

Notation. $\mu_t = \mathbb{E}(Y_t)$ (mean function), $\gamma_{t,s} = \mathrm{Cov}(Y_t, Y_s)$, $\rho_{t,s} = \gamma_{t,s}/\sqrt{\gamma_{t,t}\gamma_{s,s}}$, $\sigma_t^2 = \gamma_{t,t}$. White noise (WN) $\{\varepsilon_t\}$: mean 0, variance σ_ε^2, uncorrelated across lags.

2 Second-Order Structure

Core formulas.

$$\mathbb{E}(aU + bV) = a\,\mathbb{E}(U) + b\,\mathbb{E}(V),$$
$$|\rho_{t,s}| \le 1, \quad \rho_{t,t} = 1,$$
$$\mathrm{Cov}\left(\sum_{i=1}^m c_i Y_i, \sum_{j=1}^n d_j Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n c_i d_j\, \gamma_{i,j}.$$

Variance of a linear combination:

$$\mathrm{Var}\left(\sum_{i=1}^m c_i Y_i\right) = \sum_{i=1}^m c_i^2 \gamma_{i,i} + 2\sum_{i=2}^m \sum_{j=1}^{i-1} c_i c_j\, \gamma_{i,j}.$$

Representative processes.

- WN: $Y_t = \varepsilon_t$, so $\mu_t = 0$, $\gamma_{t,s} = \sigma_\varepsilon^2 \mathbf{1}\{t = s\}$, $\rho_{t,s} = \mathbf{1}\{t = s\}$.
- Moving average MA(1): $Y_t = \varepsilon_t + a\,\varepsilon_{t-1}$ gives $\gamma_0 = (1 + a^2)\sigma_\varepsilon^2$, $\gamma_1 = a\,\sigma_\varepsilon^2$, $\gamma_k = 0$ for $|k| > 1$, so $\rho_1 = a/(1 + a^2)$.
- Random walk: $Y_t = \sum_{j=1}^t \varepsilon_j$ has $\mu_t = 0$, $\gamma_{t,s} = \sigma_\varepsilon^2 \min\{t, s\}$, $\rho_{t,s} = \min\{t, s\}/\sqrt{ts}$ (nonstationary).

3 Stationarity

Definitions.

- Strict stationarity: joint distribution of $(Y_{t_1}, \dots, Y_{t_k})$ equals that of $(Y_{t_1+h}, \dots, Y_{t_k+h})$ for all integers h.
- Weak (second-order) stationarity: $\mu_t = \mu$ for all t and $\gamma_{t,t+h} = \gamma_h$ depends only on lag h. Toeplitz covariance matrix: Γ_T has γ_0 on the diagonal, γ_1 on first off-diagonals, etc.

Strict $\Rightarrow$ weak when $\mathrm{Var}(Y_t) < \infty$; for Gaussian series, weak $\Rightarrow$ strict.

Stationary covariance. If $\gamma_h = \mathrm{Cov}(Y_t, Y_{t+h})$, then $\rho_h = \gamma_h/\gamma_0$. For stationary Gaussian processes, γ_h fully determines the joint distribution.

4 Estimating the Mean and Trends

Sample mean. $Y = T^{-1}\sum_{t=1}^T Y_t$ estimates a constant mean μ. Variance:

$$\mathrm{Var}(\bar{Y}) = \frac{1}{T^2}\sum_{i=1}^T \sum_{j=1}^T \gamma_{i,j} = \frac{\gamma_0}{T}\left(1 + 2\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right)\rho_h\right).$$

White noise: $\mathrm{SE}(\bar{Y}) = \sigma_\varepsilon/\sqrt{T}$. MA(1): replace $\rho_1 = a/(1 + a^2)$ in expression above. Random walk: $\mathrm{Var}(\bar{Y}) = \sigma_\varepsilon^2(2T + 1)T + 1\,/\,6T$ grows with T.

Trend structures.

- Constant: $\mu_t = \mu$.
- Periodic (seasonality S): $\mu_{t+S} = \mu_t$.
- Linear: $\mu_t = \beta_0 + \beta_1 t$.
- Polynomial/cosine terms extend to quadratic or harmonic regressors.

OLS estimation. Fit $Y_t = \beta_0 + \beta_1 t + X_t$ by minimizing $\sum_{t=1}^T (y_t - \beta_0 - \beta_1 t)^2$. Closed forms:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^T (t - \bar{t})y_t}{\sum_{t=1}^T (t - \bar{t})^2}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{t}, \quad \bar{t} = \frac{T + 1}{2}.$$

Linear model representation: $Y = D\beta + e$, where D collects regressors (e.g., seasonal indicators). Residuals $\hat{X}_t = y_t - \hat{\mu}_t$ should behave approximately stationary; assess via plots, autocorrelation, and normality diagnostics.

5 Sample Autocorrelation

For weakly stationary data with sample mean Y,

$$\hat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T-k}(y_t - \bar{Y})(y_{t+k} - \bar{Y}), \qquad \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}.$$

In practice the $1/T$ factor is often replaced by $1/(T - k)$, but emphasis is on the shape of $\hat{\rho}_k$. The `acf` function in R implements these estimates.

6 Linear Time Series Models

General linear process. $Y_t = \mu + \sum_{j=0}^\infty \psi_j \varepsilon_{t-j}$ with $\sum_{j=0}^\infty \psi_j^2 < \infty$ ensures $\gamma_h = \sigma_\varepsilon^2 \sum_{j=0}^\infty \psi_j \psi_{j+h}$.

MA(q).

$$Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q},$$

with $\gamma_h = 0$ for $|h| > q$ and $\gamma_h = \sigma_\varepsilon^2 \sum_{j=0}^{q-h} \theta_{j+h}\theta_j$ for $0 \le h \le q$ (set $\theta_0 = 1$). Identified by a finite autocorrelation tail.

AR(1).

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t, \qquad |\phi| < 1.$$

Equivalent infinite MA: $Y_t = \mu + \sum_{j=0}^\infty \phi^j \varepsilon_{t-j}$. Variance $\gamma_0 = \sigma_\varepsilon^2/(1 - \phi^2)$, autocorrelation $\rho_h = \phi^{|h|}$.

AR(p). Backshift operator B gives $\phi(B)Y_t = \varepsilon_t$ with $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$. Stationarity requires all roots of $\phi(z) = 0$ satisfy $|z| > 1$ (equivalently $|\phi| < 1$ in AR(1)). Variance and covariance solve Yule–Walker equations.

7 ARMA Models and Invertibility

ARMA(p, q).

$$\phi(B)Y_t = \theta(B)\varepsilon_t, \quad \theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

Require stationarity (roots of ϕ outside unit circle) and invertibility (roots of θ outside unit circle) for identification and to express Y_t as both an infinite MA and an infinite AR. Noninvertible MA parameters can yield identical autocovariances (e.g., a vs. $1/a$); choose invertible form to ensure uniqueness.

Invertibility via backshift. If $\theta(B)$ is invertible, then $(1 + \theta_1 B + \cdots + \theta_q B^q)^{-1}$ expands as a convergent power series, leading to $Y_t = \sum_{j=1}^\infty \pi_j Y_{t-j} + \varepsilon_t$ with coefficients decaying geometrically.

8 Yule–Walker Relations

For stationary AR(p) with innovations variance σ_ε^2 and autocorrelations ρ_h:

$$\rho_k = \sum_{j=1}^p \phi_j \rho_{k-j}, \qquad k \ge 1.$$

The first p equations involve both positive and negative lags ($\rho_{-h} = \rho_h$) and form a linear system in $\phi_1, \dots, \phi_p$. For $k > p$, recursion supplies ρ_k. Once ρ_h are known, $\gamma_h = \gamma_0\rho_h$ with $\gamma_0 = \sigma_\varepsilon^2/\left(1 - \sum_{j=1}^p \phi_j\rho_j\right)$.

Backshift calculus. $BY_t = Y_{t-1}$, $B^k Y_t = Y_{t-k}$. Operators satisfy $(I - \phi B)^{-1} = \sum_{j=0}^\infty \phi^j B^j$ when $|\phi| < 1$; similarly for higher-order polynomials after factoring into linear terms.