1 Foundations

Time series data: $y_1, \ldots, y_T$ 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:

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

Moment identities.

- Independence implies factorization: if X, Y, Z independent then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ and $\mathbb{E}[X^i Y^j Z^k] = \mathbb{E}(X^i)\mathbb{E}(Y^j)\mathbb{E}(Z^k)$.
- $\mathrm{Var}(X) = \mathbb{E}\big[(X - \mu_X)^2\big] = \mathbb{E}(X^2) - \mu_X^2$ and $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mu_X \mu_Y$.
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$, $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.
- For linear combinations, $\mathrm{Cov}(aX + bY, cU + dV) = ac\,\mathrm{Cov}(X, U) + ad\,\mathrm{Cov}(X, V) + bc\,\mathrm{Cov}(Y, U) + bd\,\mathrm{Cov}(Y, V)$.

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 Known Time Series Processes

MA(1). $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$.

- $\gamma_0 = \sigma^2(1 + \theta^2)$, $\gamma_1 = \sigma^2\theta$, $\gamma_h = 0$ for $h > 1$.
- $\rho_1 = \frac{\theta}{1+\theta^2}$, $\rho_h = 0$ for $h > 1$.

MA(2). $Y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$.

- $\gamma_0 = \sigma^2(1 + \theta_1^2 + \theta_2^2)$.
- $\gamma_1 = \sigma^2(\theta_1 + \theta_1\theta_2)$, $\gamma_2 = \sigma^2\theta_2$, $\gamma_h = 0$ for $h > 2$.
- $\rho_1 = \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2}$, $\rho_2 = \frac{\theta_2}{1+\theta_1^2+\theta_2^2}$.

AR(1). $Y_t = \phi Y_{t-1} + \varepsilon_t$.

- $\gamma_0 = \frac{\sigma^2}{1-\phi^2}$.
- $\gamma_h = \phi^h \gamma_0 = \frac{\sigma^2\phi^h}{1-\phi^2}$.
- $\rho_h = \phi^h$.

AR(2). $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$.

- $\gamma_0 = \frac{\sigma^2(1-\phi_2)}{(1+\phi_2)(1-\phi_1-\phi_2)(1+\phi_1-\phi_2)}$.
- Yule-Walker: $\rho_1 = \phi_1 + \phi_2\rho_1 \implies \rho_1 = \frac{\phi_1}{1-\phi_2}$.
- $\rho_2 = \phi_1\rho_1 + \phi_2$. $\rho_h = \phi_1\rho_{h-1} + \phi_2\rho_{h-2}$ for $h \ge 2$.

ARMA(1,1). $Y_t = \phi Y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$.

- $\gamma_0 = \sigma^2 \frac{1+2\phi\theta+\theta^2}{1-\phi^2}$.
- $\gamma_1 = \sigma^2 \frac{(1+\phi\theta)(\phi+\theta)}{1-\phi^2}$.
- $\rho_1 = \frac{(1+\phi\theta)(\phi+\theta)}{1+2\phi\theta+\theta^2}$.
- $\gamma_h = \phi\gamma_{h-1}$, $\rho_h = \phi\rho_{h-1}$ for $h \ge 2$.

4 Stationarity

Definitions.

- Strict stationarity: joint distribution of $(Y_{t_1}, \ldots, Y_{t_k})$ equals that of $(Y_{t_1+h}, \ldots, Y_{t_k+h})$ for all integers h.
- Weak (second-order) stationarity: time series $\{Y_t; t = 0, \pm 1, \ldots\}$ satisfies $\mathbb{E}(Y_t) \equiv \mu$ for all t and $\mathrm{Cov}(Y_t, Y_{t+h}) = \mathrm{Cov}(Y_0, Y_h) \equiv \gamma_h$, a function of lag h only. 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.

Diagnosing stationarity.

- Check that observed mean and variance do not trend over time; plot rolling summaries to spot drifts.
- Empirical autocovariance $\hat{\gamma}_h$ should depend only on lag h; for strict stationarity test whether joint distributions appear time-shift invariant.
- Differencing or detrending (linear, seasonal, harmonic) often restores weak stationarity before fitting ARMA models.

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.

Canonical model equations.

- AR(p): $\phi(B)Y_t = \varepsilon_t$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$.
- MA(q): $Y_t = \mu + \theta(B)\varepsilon_t$, $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$.
- ARMA(p, q): $\phi(B)Y_t = \theta(B)\varepsilon_t$, combining both structures.

5 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.

Trend removal checklist.

- Seasonal pattern: include $S - 1$ dummies or sine/cosine pairs $\sin(2\pi t/S), \cos(2\pi t/S)$.
- Polynomial drift: add t, t^2 terms, or difference the series if a unit root is suspected.
- Compare before/after plots of y_t and residuals to confirm trend removal; interpret slopes in a practical context.

Residual diagnostics.

- Residual vs. time: look for outliers, curvature, or changing variance (nonstationarity).
- ACF of residuals: significant spikes indicate remaining structure and motivate refining the model.
- Q-Q plot: departures from diagonal reveal non-Gaussian innovations (heavy tails or skew).

6 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.

Autocovariance recipe.

1. Express Y_t in terms of innovations using model definition (e.g., backshift form).
2. Compute $\gamma_h = \mathbb{E}[(Y_t - \mu)(Y_{t+h} - \mu)]$ using independence of ε's.
3. For AR models use Yule–Walker recursion $\gamma_h = \sum_{j=1}^p \phi_j \gamma_{h-j}$ with γ_0 from variance equation.

Closed forms are tractable for small p or q; for ARMA combine both approaches or use software to obtain γ_h numerically.

ACF/PACF heuristics.
- Moving average MA(q): ACF cuts off after lag q; PACF tails off geometrically.
- Autoregressive AR(p): PACF cuts off after lag p; ACF decays (often exponential or damped sinusoidal).
- ARMA(p, q): both ACF and PACF tail off; rely on parameter parsimony and information criteria in addition to spike patterns.

7 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 $\gamma_h = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$.

Models to recognize.
- White noise: baseline benchmark with no autocorrelation.
- MA(q): dependence confined to last q shocks.
- AR(p): current value regresses on past p observations.
- ARMA(p, q): combines short-term shock memory with autoregression.

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.

Stationarity checks.
- AR(1): stationary when $|\phi_1| < 1$.
- AR(2): roots of $1 - \phi_1 z - \phi_2 z^2 = 0$ must lie outside unit circle; quick inequalities: $\phi_2 < 1$, $\phi_2 > -1$, and $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.
- Higher-order AR: factor $\phi(B)$; numerical eigenvalues of companion matrix reveal whether $|\lambda| < 1$.

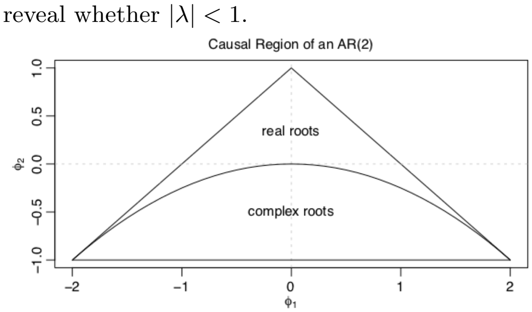


Fig. 3.3. Causal region for an AR(2) in terms of the parameters.

8 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.

MA invertibility checks.
- MA(1): invertible when $|\theta_1| < 1$, giving unique innovations.
- MA(2): solve $1 + \theta_1 z + \theta_2 z^2 = 0$; roots must satisfy $|z| > 1$.
- When $q > 2$, use numerical root finding or factorization to ensure all reciprocal roots lie inside unit circle.

9 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 $\phi_1, \ldots, \phi_p$. 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.
- Identity operator I leaves the series unchanged; difference $\nabla = I - B$ removes a unit root (random walk trend).
- Seasonal difference $\nabla_S = I - B^S$ removes periodic mean shifts; combine with ∇ for seasonal ARIMA.
- Products of operators commute: $(1 - \phi_1 B)(1 - \phi_2 B) Y_t = (1 - (\phi_1 + \phi_2) B + \phi_1 \phi_2 B^2) Y_t$.

Example: Yule–Walker for AR(1). For $Y_t = 0.5 Y_{t-1} + \varepsilon_t$ with $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$,

$$\begin{aligned} \gamma_0 &= \mathrm{Cov}(Y_t, Y_t) = 0.5\,\gamma_1 + \sigma_\varepsilon^2, \\ \gamma_1 &= \mathrm{Cov}(Y_t, Y_{t-1}) = 0.5\,\gamma_0, \\ \Rightarrow\ \gamma_0 &= \frac{\sigma_\varepsilon^2}{1 - 0.5^2} = \frac{4}{3}\sigma_\varepsilon^2, \qquad \gamma_h = 0.5^{|h|}\gamma_0. \end{aligned}$$

10 Non-Stationary Models & ARIMA

Random Walk (RW). $Y_t = Y_{t-1} + \varepsilon_t \implies Y_t = \sum_{i=1}^{t} \varepsilon_i$ (assuming $Y_0 = 0$). Mean 0, Variance $\mathrm{Var}(Y_t) = t\sigma_\varepsilon^2$ (increases with time), limiting correlations $\rho_{t,s} \approx 1$. **RW with Drift:** $Y_t = c + Y_{t-1} + \varepsilon_t \implies Y_t = ct + \sum \varepsilon_i$ (deterministic linear trend).

ARIMA(p, d, q). $W_t = \nabla^d Y_t = (1 - B)^d Y_t$ is stationary ARMA(p, q). Model: $\phi(B)(1 - B)^d Y_t = \theta(B) \varepsilon_t$.
- $d = 1$: Linear trend behavior. $d = 2$: Quadratic trend behavior.
- **Drift in R**: 'include.drift=T' in 'Arima(0,1,0)' fits $Y_t = c + Y_{t-1} + \varepsilon_t$. Coefficient is slope c.

11 Model Identification

ACF and PACF Behavior.
- **AR(p):** ACF tails off; PACF cuts off after lag p.
- **MA(q):** ACF cuts off after lag q; PACF tails off.
- **ARMA(p, q):** Both ACF and PACF tail off.

Standard Error: $SE(r_k) \approx 1/\sqrt{T}$. Bounds $\pm 1.96/\sqrt{T}$.

EACF (Extended ACF). Identify p and q simultaneously. Look for the tip of a triangle of 'O' (non-significant values). Top-left vertex indicates (p, q).

Information Criteria. Minimize AIC ($-2\ln(L) + 2k$) or BIC ($-2\ln(L) + k\ln(T)$). $k = p + q + d + 1$. BIC penalizes complexity more.

Dickey-Fuller Test (ADF). H_0: Unit Root exists (Non-stationary). H_a: Stationary. Low p-value $\implies$ Stationary.

12 Parameter Estimation & Diagnostics

Estimation Methods.
- **Method of Moments (Yule-Walker)**: Equate ρ_k to r_k. Good initial guess.
- **MLE**: Maximize likelihood. Preferred.
- **CSS**: Minimize $\sum \varepsilon_t^2$ assuming initial zeros.

R Output: 'mean' in ARIMA(p,1,q) is drift. 'intercept' in ARMA(p,q) is $\delta = \mu(1 - \sum \phi_i)$.

Residual Diagnostics. Residuals $\hat{\varepsilon}_t$ should be White Noise.
- **Plot**: No trend, constant variance.
- **ACF**: All within $\pm 1.96/\sqrt{T}$.
- **Ljung-Box Test**: H_0: $\rho_1 = \cdots = \rho_K = 0$ (Adequate). High p-value (> 0.05) $\implies$ Good fit.
- **Q-Q Plot**: Normality check.

13 Forecasting

General Principle. Best MSE forecast: $\hat{Y}_{T+h} = \mathbb{E}[Y_{T+h} | Y_T, \ldots]$. Rules: $\mathbb{E}[\varepsilon_{T+h} | \text{Past}] = 0$ ($h > 0$), $\mathbb{E}[\varepsilon_{T-j} | \text{Past}] = \hat{\varepsilon}_{T-j}$, $\mathbb{E}[Y_{T-j} | \text{Past}] = Y_{T-j}$.

One-Step Ahead ($h=1$). AR(1): $\hat{Y}_{T+1} = \phi Y_T$. MA(1): $\hat{Y}_{T+1} = -\theta \hat{\varepsilon}_T$.

Multi-Step Ahead ($h>1$). AR(1): $\hat{Y}_{T+h} = \phi^h Y_T \to 0$. MA($q$): $\hat{Y}_{T+h} = 0$ for $h > q$. RW: $\hat{Y}_{T+h} = Y_T$. RW+Drift: $\hat{Y}_{T+h} = Y_T + hc$.

Forecast Error Variance. $e_h = Y_{T+h} - \hat{Y}_{T+h} = \sum_{j=0}^{h-1} \psi_j \varepsilon_{T+h-j}$. $\mathrm{Var}(e_h) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} \psi_j^2$. Weights ψ: from $\phi(x)\psi(x) = \theta(x)$.

- 1-step: σ_ε^2.
- AR(1) 2-step: $\sigma^2(1+\phi^2)$.
- MA(1) 2-step: $\sigma^2(1+\theta^2)$.
- RW: $h\sigma^2$.

Prediction Intervals (95%). $\hat{Y}_{T+h} \pm 1.96\sqrt{\mathrm{Var}(e_h)}$.

14 Transformations & Regression

Transformations. Variance grows with level $\implies$ $\ln(Y_t)$ or $\sqrt{Y_t}$. Forecast original: $\hat{Y}_{orig} = f^{-1}(\hat{Y}_{trans})$. Interval: $[f^{-1}(L), f^{-1}(U)]$.

Regression with Time Series Errors. $Y_t = \beta_0 + \beta_1 t + X_t$, X_t is ARMA. Fit OLS $\to$ Residuals $\to$ Identify ARMA $\to$ Re-fit GLS ('Arima' with 'xreg'). OLS SEs are wrong (usually too small) if errors autocorrelated.

15 Specific Exam Derivations

Non-Standard Process ($Y_t = Y_{t-1}\varepsilon_t$). Mean 0. Var $(\sigma^2)^t\,\mathrm{Var}(Y_0)$.

AR(2) Forecast ($Y_t = 2Y_{t-1} - Y_{t-2} + \varepsilon_t$). Equivalent to ARIMA(0,2,0). $\hat{Y}_{T+h}$ follows linear trend of last two points.

16 R Code to Math

`Arima(x, c(1,0,0))` $\to$ AR(1): $Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t$. `Arima(x, c(0,0,1))` $\to$ MA(1): $Y_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1}$. `Arima(x, c(0,1,0))` $\to$ RW: $Y_t - Y_{t-1} = \varepsilon_t$. `eacf` $\to$ upper left 'O'. `adf.test` $\to$ $p < 0.05$ stationary.

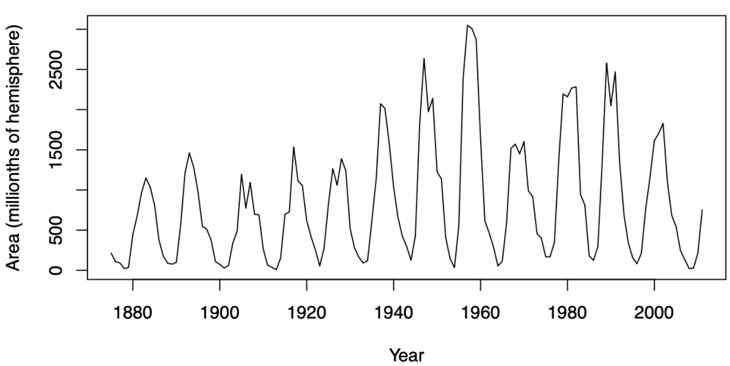
17 Communication & Interpretation

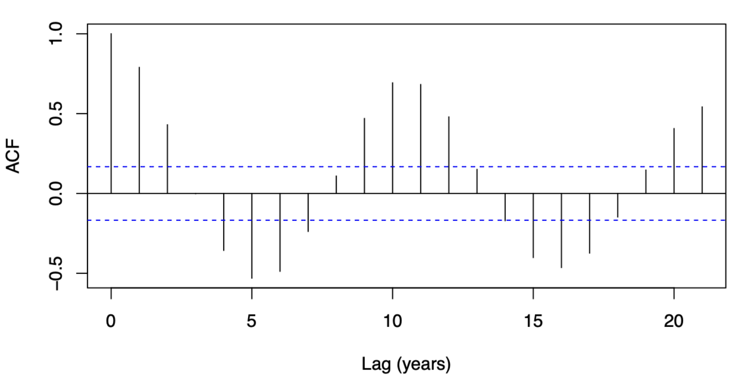
Describe a series. Reference level, trend, seasonality, and noise: e.g., "Monthly sales show an upward linear trend with seasonal peaks every December and weak autocorrelation after lag 2."

Describe a Q-Q plot. Compare quantiles of residuals to the theoretical line; systematic S-shape implies heavy tails, while slope changes suggest variance issues.

18 Sample Plots & Answers

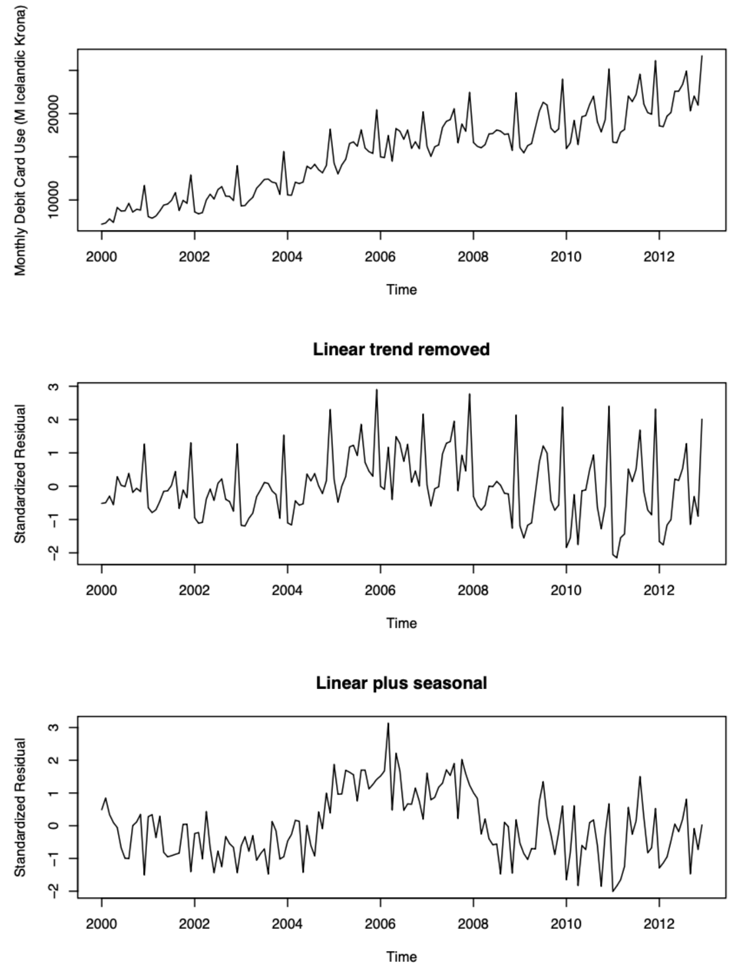
Sunspots (Sample 4).





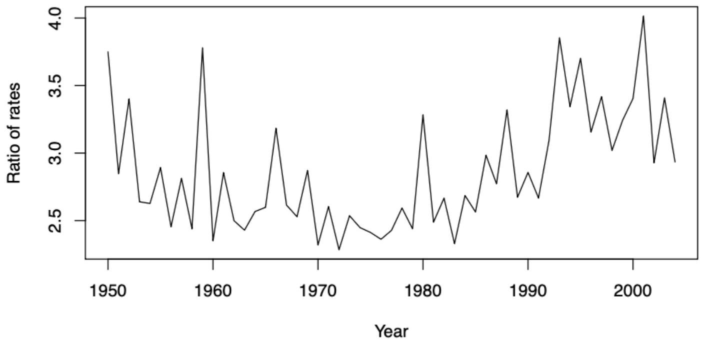
The sunspot series does **not show a clear long-term trend** in mean level, so it appears approximately **stationary in mean**, though the **variability** seems slightly higher in the middle decades. The ACF shows **strong periodic correlation**, consistent with the roughly 11-year solar cycle, but this does **not imply non-stationarity**—a perfectly stationary series with a cosine-like pattern can produce a slowly decaying ACF. Therefore, while there may be small changes in amplitude, the overall behaviour of the series is reasonably stationary.

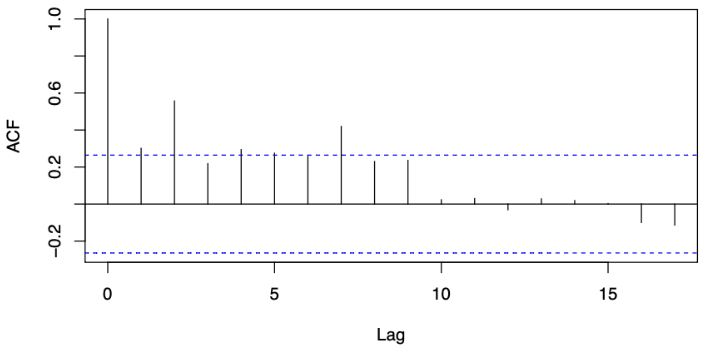
Debit Card Usage (Sample 5).



The debit card data show a **strong upward trend** over time, with usage increasing by about 1,000–1,100 million ISK per year, consistent with the regression output. The first model (linear trend only) explains about 83% of the variation ($R^2 = 0.83$), but the residuals clearly show **strong seasonality**. Adding monthly seasonal terms gives a much better fit ($R^2 = 0.94$), capturing the clear **annual cycle** with **very high December spending** and peaks during summer months (June–August). The residual plot for the seasonal model still shows **a large positive spike around 2007–2008**, likely tied to the economic boom ahead of Iceland's 2008 crash, but otherwise the residuals fluctuate randomly around zero. To further assess model quality, check residual diagnostics: a month plot to confirm seasonality removal, hist/Q–Q plots for normality, and the residual ACF for remaining autocorrelation. Overall, the linear plus seasonal model fits well, whereas the simple linear trend alone is inadequate.

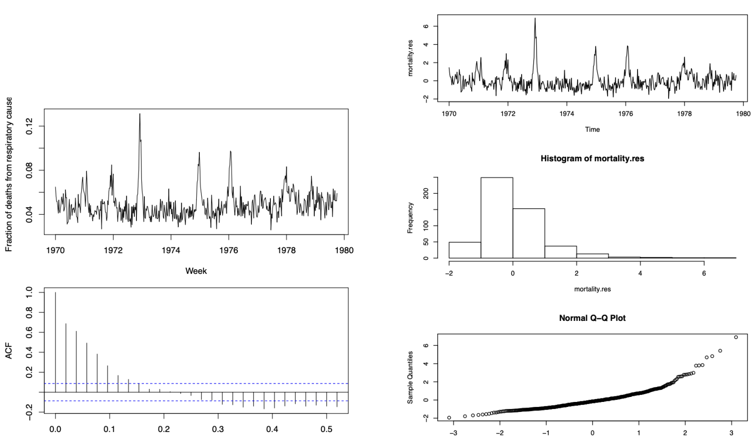
Female Homicide Ratio (Sample 6).





The U.S.–to–Canada female homicide rate ratio appears roughly stationary, fluctuating between about 2.5 and 4 without a strong long-term trend. There may be mild cyclical variation, but no clear systematic drift. To check this formally, inspect the ACF plot, which already suggests modest short-lag autocorrelation (notably at lags 2 and 7). In R, plot the series and its ACF (‘acf()‘), then run an Augmented Dickey–Fuller test (‘adf.test()‘ from `tseries`). If a slow trend is suspected, fit `lm(ratio ~ time + I(time^2))`, examine the residuals, and re-plot their ACF. Finish with residual diagnostics (standardized residuals, histograms, Shapiro–Wilk) and, if needed, analyze the original U.S. and Canada series jointly to interpret why certain lags show correlation.

Mortality Ratio Trend (Sample 7).



The fitted regression of the mortality ratio on time gives a **small but statistically significant positive slope** (estimate ≈ 0.00049, $p = 0.0138$). Over 10 years the ratio increases by about 0.005, modest relative to the mean (≈ 0.05) and standard deviation (≈ 0.01). The very low $R^2 \approx 0.01$ shows that time explains little variation, so the linear trend adds little forecasting power even though the t-test is significant. The ACF plot was labeled incorrectly ("Week" misread as "Year"), creating spurious negative correlations at 0.5, yet the data still show **strong short-term autocorrelation** and likely **seasonal patterns**. Residual diagnostics reveal **non-normal, right-skewed errors** and a clear **outlier around 1973** with a standardized residual near 6. These issues violate model assumptions; model seasonality explicitly (e.g., harmonic terms or differencing) and investigate the 1973 spike before trusting the fit.

- Differencing (∇): Removes polynomial trends.
- – ∇Y_t: Removes linear trend.
- – $\nabla^2 Y_t$: Removes quadratic trend.
- – $\nabla_s Y_t = Y_t - Y_{t-s}$: Removes seasonality of period s.

1. EXPECTATIONS & OPERATORS

Rules for Expectations & Variances

- $E[\sum c_iU_i] = \sum c_iE[U_i]$
- $Var(U) = E[(U-\mu)^2] = E[U^2] - (E[U])^2$
- $Cov(U,V) = E[(U-\mu_u)(V-\mu_v)] = E[UV] - E[U]E[V]$
- **Bilinearity:** $Cov(\sum c_iU_i, \sum d_jV_j) = \sum_i \sum_j c_id_jCov(U_i,V_j)$
- **Variance Sum:** $Var(\sum c_iU_i) = \sum c_i^2Var(U_i) + \sum_{i\neq j} c_ic_jCov(U_i,U_j)$
- **Independence:** U,V indep $\implies Cov(U,V) = 0, E[UV] = E[U]E[V]$.

Backshift Operator (B) & Difference (∇)

- $BY_t = Y_{t-1}$, $B^kY_t = Y_{t-k}$
- $\nabla Y_t = (1-B)Y_t = Y_t - Y_{t-1}$
- $\nabla^2Y_t = (1-B)^2Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$

2. STATIONARITY & FUNCTIONS

Strict Stationarity: Joint dist. of $Y_{t_1},\dots,Y_{t_k}$ same as $Y_{t_1+h},\dots,Y_{t_k+h}$. **Weak Stationarity:**

1. Mean constant: $E[Y_t] = \mu$ for all t.
2. Variance constant: $Var(Y_t) = \gamma_0 < \infty$.
3. Covariance depends only on lag: $Cov(Y_t, Y_{t+h}) = \gamma_h$.

Autocovariance (ACVF) & Autocorrelation (ACF)

$$\gamma_h = E[(Y_t-\mu)(Y_{t+h}-\mu)] \quad \text{and} \quad \rho_h = \frac{\gamma_h}{\gamma_0}$$

Properties: $\gamma_0 = Var(Y_t)$, $|\rho_h| \leq 1$, $\gamma_{-h} = \gamma_h$, $\rho_0 = 1$.

Sample Mean Statistics Estimator: $\hat{\mu} = \bar{Y} = \frac{1}{T}\sum Y_t$.

$$Var(\bar{Y}) = \frac{1}{T^2}\sum_{i=1}^{T}\sum_{j=1}^{T}\gamma_{i-j} = \frac{1}{T}\sum_{h=-(T-1)}^{T-1}\left(1 - \frac{|h|}{T}\right)\gamma_h$$

Standard Error (Large T): $SE \approx \sqrt{\frac{\gamma_0}{T}\sum_{h=-\infty}^{\infty}\rho_h}$.

3. BASIC MODELS

White Noise (WN) ϵ_t: $E[\epsilon_t] = 0$, $Var(\epsilon_t) = \sigma^2$, $\gamma_h = 0$ if $h \neq 0$.

Random Walk (RW): $Y_t = Y_{t-1} + \epsilon_t = \sum_{j=1}^{t}\epsilon_j$ (assume $Y_0 = 0$).

- Mean: $E[Y_t] = 0$.
- Variance: $Var(Y_t) = t\sigma^2$ (Increases with time, non-stationary).
- ACVF: $\gamma_{t,s} = \min(t,s)\sigma^2$. $\rho_{t,s} = \sqrt{\frac{\min(t,s)}{\max(t,s)}}$.

General Linear Process (GLP) $Y_t = \mu + \sum_{j=0}^{\infty}\psi_j\epsilon_{t-j}$ with $\psi_0 = 1, \sum\psi_j^2 < \infty$.

$$\gamma_h = \sigma^2\sum_{j=0}^{\infty}\psi_j\psi_{j+h}$$

4. ARMA MODELS

Notation: $\phi(B)Y_t = \theta(B)\epsilon_t$

$$(1 - \phi_1B - \dots - \phi_pB^p)Y_t = (1 - \theta_1B - \dots - \theta_qB^q)\epsilon_t$$

**Note: Text uses minus signs for θ. Lecture sometimes varies. Check polynomial roots.*

Stationarity: Roots of $\phi(x) = 0$ must lie **outside** the unit circle ($|x| > 1$). **Invertibility:** Roots of $\theta(x) = 0$ must lie **outside** the unit circle.

AR(1): $Y_t = \phi Y_{t-1} + \epsilon_t$ ($|\phi| < 1$)

- GLP form: $Y_t = \sum_{j=0}^{\infty}\phi^j\epsilon_{t-j}$ ($\psi_j = \phi^j$)
- Variance: $\gamma_0 = \frac{\sigma^2}{1-\phi^2}$
- ACF: $\rho_h = \phi^h$ (Decays exponentially)

MA(1): $Y_t = \epsilon_t - \theta\epsilon_{t-1}$

- Mean: μ. Variance: $\gamma_0 = \sigma^2(1+\theta^2)$.
- ACF: $\rho_1 = \frac{-\theta}{1+\theta^2}$, $\rho_h = 0$ for $h > 1$.

AR(2): $Y_t = \phi_1Y_{t-1} + \phi_2Y_{t-2} + \epsilon_t$

- Stationarity: $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, $|\phi_2| < 1$.
- Yule-Walker ($k > 0$): $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$.
- $\rho_1 = \frac{\phi_1}{1-\phi_2}$, $\rho_2 = \phi_1\rho_1 + \phi_2$.

ARMA(1,1): $Y_t - \phi Y_{t-1} = \epsilon_t - \theta\epsilon_{t-1}$

- $\gamma_0 = \sigma^2\frac{1-2\phi\theta+\theta^2}{1-\phi^2}$
- $\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1-2\phi\theta+\theta^2}$
- $\rho_k = \phi\rho_{k-1}$ for $k \geq 2$.

ARIMA(p,d,q): $W_t = \nabla^dY_t$ follows $ARMA(p,q)$.

- If $d > 0$, original Y_t is non-stationary (variance grows).
- Constant term in ARIMA $\implies$ deterministic polynomial trend in Y_t.

5. IDENTIFICATION & DIAGNOSTICS

ACF/PACF Behavior

Model	ACF	PACF
$AR(p)$	Tails off	Cuts off after lag p
$MA(q)$	Cuts off after lag q	Tails off
$ARMA$	Tails off	Tails off

EACF (Extended ACF): Look for triangle of '0's. Top-left vertex implies order (p,q). **Information Criteria (Minimize these):**

$$AIC = -2\ln(L) + 2k \qquad BIC = -2\ln(L) + k\ln(T)$$

where $k = p + q + 1$ (parameters). BIC penalizes parameters more heavily (selects simpler models).

Unit Root Tests (Dickey-Fuller) H_0: Unit root present (Non-stationary, needs differencing). H_a: Stationary. Low p-value (< 0.05): Reject H_0, assume stationary.

Diagnostics

- **Residuals:** Should be White Noise.
- **Ljung-Box Test:** H_0: Residuals are independent (WN). Significant p-value implies lack of fit (residuals correlated).
- **QQ Plot:** Check for Normality (straight line).

6. FORECASTING

General Principle: Minimize Mean Squared Prediction Error (MSPE). Optimal Forecast: $\hat{Y}_{T+h} = E[Y_{T+h}|Y_1,\dots,Y_T]$.

Conditional Expectation Rules for Forecasting

1. **Past Data:** $E[Y_{T-j}|\mathcal{F}_T] = Y_{T-j}$ for $j \geq 0$.
2. **Past Noise:** $E[\epsilon_{T-j}|\mathcal{F}_T] = \epsilon_{T-j}$ (Calculated as residuals e_t).
3. **Future Noise:** $E[\epsilon_{T+h}|\mathcal{F}_T] = 0$ for $h > 0$.

Calculation Procedure (Box-Jenkins) 1. Write model for $T+h$: $Y_{T+h} = \mu + \phi Y_{T+h-1} + \dots + \epsilon_{T+h} - \theta\epsilon_{T+h-1}$. 2. Apply conditional expectation $E[\cdot|Y_1 \dots Y_T]$. 3. Replace future ϵ with 0, past ϵ with residuals, future Y with forecasts $\hat{Y}$, past Y with data.

Forecast Error & Variance Error: $e_T(h) = Y_{T+h} - \hat{Y}_{T+h}$. Structure: $e_T(h) = \sum_{j=0}^{h-1}\psi_j\epsilon_{T+h-j}$ (Truncated GLP). **Variance:**

$$Var(e_T(h)) = \sigma^2\sum_{j=0}^{h-1}\psi_j^2$$

- $h = 1$: $Var = \sigma^2$.
- $h \to \infty$ (Stationary): $Var \to \gamma_0$ (Process Variance).
- $h \to \infty$ (Non-stationary): $Var \to \infty$.

Prediction Intervals (95%)

$$\hat{Y}_{T+h} \pm 1.96\sqrt{Var(e_T(h))}$$

*If parameters are estimated, this ignores estimation error (okay for large T).

Example: AR(1) Forecasts Model: $Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t$. $\hat{Y}_{T+1} = \mu + \phi(Y_T - \mu)$. $\hat{Y}_{T+h} = \mu + \phi^h(Y_T - \mu)$. Variance ($h = 1$): σ^2. Variance ($h = 2$): $\sigma^2(1+\phi^2)$.

Example: MA(1) Forecasts Model: $Y_t = \mu + \epsilon_t - \theta\epsilon_{t-1}$. $\hat{Y}_{T+1} = \mu - \theta\epsilon_T = \mu - \theta(Y_T - \hat{Y}_T)$. $\hat{Y}_{T+h} = \mu$ for all $h \geq 2$ (Memory is short). Variance ($h = 1$): σ^2. Variance ($h \geq 2$): $\sigma^2(1+\theta^2)$.

Forecasting Transformed Series If $W_t = \ln(Y_t)$ and you forecast $\hat{W}_{T+h} \pm C$: Interval for Y: $[e^{\hat{W}_{T+h}-C}, e^{\hat{W}_{T+h}+C}]$. *Note: Point forecast $e^{\hat{W}}$ is biased for mean of Y.

7. TRENDS & ESTIMATION

Estimation Methods

- **Yule-Walker:** Method of Moments. Solves $\rho_k = \sum\phi_j\rho_{k-j}$. Good for AR initial estimates.
- **MLE:** Maximizes likelihood $L(\beta,\phi,\theta,\sigma^2)$. Assumes Normality.
- **Least Squares (CLS):** Minimizes $\sum\epsilon_t^2$.

Trends

- **Linear:** $\mu_t = \beta_0 + \beta_1t$. Fit OLS. Residuals may be ARMA.
- **Seasonal:** Dummy variables (indicators).

Final Exam Preparation

- December 10, 2025 from 7:00 PM to 10:00 PM.
- Room: SSC B9200 (big lecture theatre on 3000 level of AQ).
- Crowdmark Exam out of 40.
- Cheat-sheets: 6 pieces of paper written or printed on both sides.
- No font size restrictions.
- REQUIRED: Bring SFU Student card – the physical card is needed.
- Calculator permitted – no internet connection.
- No internet capable device is permitted.

TENTATIVE Office Hours

- Monday 8 Dec, 9 AM to 11 AM.
- Tuesday 9 Dec, 2 PM to 4 PM by zoom.
- Wednesday 10 Dec, 9AM to 11 AM.
- Dec 5 by appointment.

Coverage

- Chapters 1 through 9.
- Some technical bits not covered in class will not be on final.
- Final slightly more focused on material after the midterm.
- But the material is very much cumulative.
- There will be theoretical calculations involving means, variances, covariances, correlations, standard deviations, confidence intervals, hypothesis tests, forecasts, forecast errors, forecast standard errors, and prediction / forecast intervals.
- I will not ask you to write R code.
- But I will expect you to read and interpret output.
- And do arithmetic with output sometimes to compute intervals etc.
- There will be questions like the first 3 on the midterm.
- Answers to such questions must be based on time series ideas from course.

Rules for expected values

- Random variables: $U_1, \ldots, U_t$ and $V_1, \ldots, V_t$:
- Constants: $c_1, \ldots, c_s$, $d_1, \ldots, d_t$.
- Expected Value:

$$\mathrm{E}\left(\sum_{i=1}^{s} c_i U_i\right) = \sum_{i=1}^{s} c_i \mathrm{E}(U_i).$$

- Constant random variables:

$$\mathrm{E}(a) = a.$$

- Common notation $\mu_U = \mathrm{E}(U)$.

Rules for variances, covariances

- Variance: $\mathrm{Var}(U) = \mathrm{E}\left\{(U-\mu_U)^2\right\} = \mathrm{E}(U^2) - \mu_U^2$.
- Covariance:

$$\mathrm{Cov}\left(\sum_{i=1}^{s} c_i U_i, \sum_{j=1}^{t} d_j V_j\right) = \sum_{i=1}^{s}\sum_{j=1}^{t} c_i d_j \mathrm{Cov}(U_i, V_j)$$

- Variance:

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{s} c_i U_i\right) &= \mathrm{Cov}\left(\sum_{i=1}^{s} c_i U_i, \sum_{j=1}^{t} c_j U_j\right) \\ &= \sum_{i=1}^{s}\sum_{j=1}^{s} c_i c_j \mathrm{Cov}(U_i, U_j)\end{aligned}$$

Effect of independence

- If U and V are independent then

$$\begin{aligned}\mathrm{E}\left\{g(U)h(V)\right\} &= \mathrm{E}\left\{g(U)\right\}\mathrm{E}\left\{h(V)\right\} \\ \mathrm{Cov}\left\{g(U), h(V)\right\} &= 0.\end{aligned}$$

- Variance: if $U_1, \ldots, U_s$ are independent then

$$\mathrm{Var}\left(\sum_{i=1}^{s} c_i U_i\right) = \sum_{i=1}^{s} c_i^2 \mathrm{Var}(U_i).$$

- You need to be able to apply all these rules to time series.

Time series models

- Recognize AR(p), MA(q), ARMA(p, q), ARIMA(p, d, q).
- Understand the properties of mean 0 white noise and mean μ white noise.
- Understand what *stationary* means.
- Know the basic models for non-stationary trend.
- Know how to check if a covariance function is stationary.

Mean, autocovariance, and autocorrelation functions

- For a general series Y:
- The mean function is $\mu_Y(t) = \mu(t) = \mathrm{E})Y_t$.
- The autocovariance function is

$$\gamma_{s,t} = \mathrm{Cov}(Y_s, Y_t).$$

- The autocorrelation function is

$$\rho_{s,t} = \frac{\gamma_{s,t}}{\sqrt{\gamma_{s,s}\gamma_{t,t}}}.$$

- If Y is stationary then μ does not depend on t and

$$\begin{aligned}\gamma_{s,s+h} &= \gamma_{0,h} = \gamma_{h,0} \equiv \gamma_h \\ \rho_{s,s+h} &= \rho_{0,h} = \rho_{h,0} \equiv \rho_h\end{aligned}$$

Mean structures covered

- Zero.
- Constant.
- Linear (straight line function of t).
- Seasonal – different mean each month or each quarter.
- Understand the difference between seasonal models with and without an intercept.
- Linear plus seasonal.
- Quadratic trend.
- Cosine waves – not on final.

Estimating the mean

- If the mean is not constant must detrend (or difference) to get stationary series.
- *Preliminary* estimate using `lm` in `R`.
- Do not rely on standard errors.
- In `R` use `time(y)` to create a variable which contains observation times.
- In `R` use `season(y)` to create a variable which contains dummy variables for months or quarter.
- For a constant mean use $\hat{\mu} = \overline{Y}$.
- Know how to use variance formula:

$$\mathrm{Var}(\bar{Y}) = \frac{1}{T^2}\sum_{ij}\mathrm{Cov}(Y_i, Y_j). = \frac{1}{T^2}\sum_{ij}\gamma_{i,j}.$$

Special Cases

- Stationary series

$$SE = \frac{1}{\sqrt{T}}\sqrt{\gamma_0 + 2\sum_{j=1}^{T-1}\left(1 - \frac{j}{T}\right))\gamma_j}$$

- I gave formulas for AR(1), MA(1), Random Walk.
- These are examples of calculations I might ask you to do.

ARMA(p, q) Processes

- Know how to get equations to solve to find autocovariance and / or autocorrelation.
- Understand how to write AR(1) as MA(∞).
- Understand how to write MA(1) as *causal* AR(∞).
- Understand how to find the AR polynomial and the MA polynomial.
- Understand (for $p \le 2$) how to check if AR(p) is stationary.
- Understand (for $q \le 2$) how to check if MA(q) is invertible.
- Understand notation for operators I, B and polynomials like

$$\phi(B) = I + \phi_1 B + \phi_2 B^2$$

and understand what time series is defined by

$$\phi(B)Y_t = \epsilon_t$$

- Understand how to multiply out a product of two small polynomials.

Compute covariances

- Be able to compute autocovariance of ARMA(p, q) for small p, q.
- Be ready to use Yule Walker equations – remember $\gamma_{-k} = \gamma_k$.
- Remember $\mathrm{Cov}(U + a, V = b) = \mathrm{Cov}(U, V)$.
- Remember that common factors in the two polynomials must be cancelled.
- Understand that if you fit ARMA(2,2) when ARMA(1,1) is right you have a problem.

Understand differencing

- Know the notation

$$\nabla Y_t = (I - B)Y_t = Y_t - Y_{t-1}.$$

- Know how to use things like

$$\nabla^2 = (I - B)^2.$$

- Know how to find covariance for Random Walk.
- Know what differencing does to constant, linear, and quadratic trend.
- Know what differencing does if the original series is stationary.

Transformations

- We touched on logs, square roots, fractional differencing, Box-Cox.
- Only logs and square roots could be on final.
- Understand goal: simplify trend, reduce or eliminate variation in the variance over time.

Properties of the sample acf and partial acf

- Know how we compute the sample autocorrelation.
- Know how to diagnose q for an MA(q) using sample acf.
- Know how to diagnose p for and AR(p) using sample partial acf.
- Understand use of dotted line cut-offs in acf and pacf plots.

Conditional Expectation

- Know and use the basic properties of conditional expectation, conditional covariance.
- See Chapter 6.
- I gave you formula for conditional covariance of U and V given third variable W.
- Know how to find covariance for Random Walk.
- Know what differencing does to constant, linear, and quadratic trend.
- Know what differencing does if the original series is stationary.

Transformations

- We touched on logs, square roots, fractional differencing, Box-Cox.
- Only logs and square roots could be on final.
- Understand goal: simplify trend, reduce or eliminate variation in the variance over time.

Properties of the sample acf and partial acf

- Know how we compute the sample autocorrelation.
- Know how to diagnose q for an MA(q) using sample acf.
- Know how to diagnose p for and AR(p) using sample partial acf.
- Understand use of dotted line cut-offs in acf and pacf plots.

Conditional Expectation

- Know and use the basic properties of conditional expectation, conditional covariance.
- See Chapter 6.
- I gave you formula for conditional covariance of U and V given third variable W.
- Likely to be too hard for the final.
- But know formulas needed for forecasting:
- Data $Y_1, \ldots, Y_T$ – stationary causal series. For all $k \geq 1$:

$$\mathrm{E}(\epsilon_{T+k}|Y_1, \ldots, Y_T) = 0$$

and

$$\mathrm{Cov}(\epsilon_{t+k}, \epsilon_t) = 0.$$

Extended AutoCorrelation Function, eacf

- Know how table works by hypothesis testing – no need to know how to do the tests.
- More important: know how to look at eacf and choose the simplest reasonable model.
- Be sure to explain your choice.
- Understand the dangers in doing multiple hypothesis tests.

Dickey-Fuller test, information criteria

- Be prepared to use computer output for `adf.test` to test null hypothesis that the AR polynomial has a unit root.
- Be prepared to use table of values of AIC or BIC to select model order.
- Understand which tasks AIC is best for and which BIC is best for.

Parameter Estimation

- Usually use maximum likelihood to estimate continuous parameters.
- Parameters to estimate are ϕs (AR), θs (MA), β (mean structure), σ (noise SD).
- There won't be questions on transformation parameters like Box-Cox.
- Yule-Walker plus linear regression can give initial estimates.
- Understand that maximum likelihood chooses parameters to maximize the joint density of the data or the log of that.
- Understand that the maximization is essentially always an iteration needing initialization.
- Maximum likelihood permits estimation of standard errors.
- In `R`, the algorithms `arima` (in `stats` and `TSA`) or `Arima` (in `forecast`) produce MLEs and estimated SEs.
- Know how to use SEs and estimates to get confidence intervals and test hypotheses.
- Know how to interpret the results of such.

Forecasts, Forecast SEs, Prediction intervals

- Know basic properties of conditional expectation.
- Know what means squared prediction error is.
- Know why it is minimized by a conditional expectation.
- Compute conditional expectations, given data, of model equations to find forecasts (in terms of unknown parameters).
- Know how to convert those formulas into actual point forecasts using estimates.
- Know how to write out ideal forecast error in terms of ϵs.
- Know how to compute the Variance of this forecast error.
- Know how to turn this into a prediction interval.

How to compute Conditional Expectations

- **Rule 0**:

$$\mathrm{E}(aU + bV + c|W) = a\mathrm{E}(U|W) + b\mathrm{E}(V|W) + c.$$

- **Rule 1**: If W_N is independent of W_O then

$$\mathrm{E}(W_\mathrm{N}|W_\mathrm{O}) = \mathrm{E}(W_\mathrm{N}).$$

- **Rule 2**: treat conditioning variables as constants.
- **Rule 3**:

$$\mathrm{E}\left\{\mathrm{E}(U|W)\right\} = \mathrm{E}(U).$$

Compute 1, 2, and more steps ahead forecasts

- No model harder than ARIMA(p, d, q) with p, d, q all ≤ 3.
- But all will be substantially simpler than ARIMA(3,3,3).
- Derive theoretical formulas for forecasts and variances.
- Use `arima` output to do estimation and compute estimated forecast SE.
- Convert those to a forecast interval.
- Be able to compute forecast interval on original scale if a log, square root, or similarly easy transformation has been applied.
- Know how to accommodate estimate mean structure and differencing.

Residual analysis

- Understand role of residual analysis in selecting a model for the mean before fitting an ARIMA model.
- Interpret a time series plot of fitted residuals
- Interpret acf and pacf of fitted residuals.
- Interpret Q-Q plot of fitted residuals.
- Spot heavy tails, skewness.
- Understand influence of heavy tails on forecast interval performance.
- Understand how residuals are computed for simple ARIMA processes.

1. Suppose ϵ_1, ϵ_2, and so on are independent (white) noise with mean 0 and standard deviation σ. Let $Y_1 = \epsilon_1$ and

$$Y_t = Y_{t-1}\epsilon_t$$

for $t = 2, 3, \cdots$.

(a) Compute the mean and and standard deviation of Y_t. [4 marks]

$\Rightarrow \mathbb{E}[Y_t] = \mathbb{E}[Y_{t-1}\varepsilon_t] = \mathbb{E}[Y_{t-1}]\mathbb{E}[\varepsilon_t]$
$= \text{''} \cdot 0 = 0$

$\Rightarrow Var[Y_t] \overset{?}{=} \overset{\text{mean}=0}{\mathbb{E}[Y_t^2]} = \mathbb{E}[Y_{t-1}^2 \varepsilon_t^2] = \mathbb{E}[Y_{t-1}^2]\mathbb{E}[\varepsilon_t^2]$
$= \mathbb{E}[Y_{t-1}^2]\cdot\sigma^2$
$= \mathbb{E}[Y_{t-2}^2]\cdot(\sigma^2)^2$
$= \dots$
$= (\sigma^2)^t = \sigma^{2t}$

$\Rightarrow SD[Y_t] = \sqrt{Var(Y_t)} = \sqrt{\sigma^{2t}} = \sigma^t$

(b) Is Y_t weakly stationary? (Explain how you know.) [2 marks]

No $\quad Cov(Y_t, Y_t) = \sigma^{2t} \neq Cov(Y_s, Y_s) = \sigma^{2s}$

→ variance depends on the time t → not only the lag

(c) If you have data $Y_1, \ldots, Y_T$ what is the forecast for Y_{T+1}? Explain. [2 marks]

$\mathbb{E}[Y_{T+1} | Y_T \ldots Y_1] = \mathbb{E}[\varepsilon_{T+1} \cdot Y_T | Y_T]$
$= 0 \cdot Y_T = \underline{\underline{0}}$

from conditional expected value

(a) On the basis of this output which ARIMA model might you fit to the data? Why? [3 marks]

→ MA(1) → only first value in ACF seems significant w.r.t. 95% confidence intervalle under white noise assumption

→ EACF would also suggest MA(1)

(b) Not being sure what your answer might be to the previous part I provide on the next three pages the output from a variety of fits. Which would you select (and why)? [3 marks]

→ MA(1) has best AIC and BIC

→ adding in MA(2) → second MA param is not significant

→ similar for ARMA(1,1); AR param is not significant

3. Suppose that Y_t is a time series following the model

$$Y_t = 2Y_{t-1} - Y_{t-2} + \epsilon_t$$

and you have data $Y_1, \ldots, Y_T$.

(a) Find formulas for the forecasts $\hat{Y}_{T+1}$, $\hat{Y}_{T+2}$ and $\hat{Y}_{T+3}$. [3 marks]

$\hat{y}_{T+1} = \mathbb{E}[y_{T+1} \mid y_T \ldots y_1] = \mathbb{E}[2y_T - y_{T-1} + \varepsilon_{T+1} \mid y_T \ldots y_1]$

$= 2y_T - y_{T-1} + \underbrace{\mathbb{E}[\varepsilon_{T+1} \mid y_T \ldots]}_{=0}$

$\Rightarrow \hat{y}_{T+1} = 2y_T - y_{T-1}$

$\Rightarrow \hat{y}_{T+2} = \mathbb{E}[2y_{T+1} - y_T + \varepsilon_{T+2} \mid y_T \ldots]$

$= 2\mathbb{E}[y_{T+1} \mid y_T \ldots] - y_T$

$\Rightarrow \hat{y}_{T+2} = 2\hat{y}_{T+1} - y_T = 3y_T - 2y_{T-1}$

similarly

$\Rightarrow \hat{y}_{T+3} = 2\hat{y}_{T+2} - \hat{y}_{T+1}$

$= 6y_T - 4y_{T-1} - 2y_T + y_{T-1}$

$= 4y_T - 3y_{T-1}$

(b) Find a formula for $\mathrm{Var}(Y_{T+2} - \hat{Y}_{T+2})$. [3 marks]

$= \mathrm{Var}(2y_{T+1} - y_T + \varepsilon_{T+2} - 3y_T + 2y_{T-1})$

$= \mathrm{Var}(\cancel{3y_T} - \cancel{2y_{T-1}} + 2\varepsilon_{T+1} + \varepsilon_{T+2} - \cancel{3y_T} + \cancel{2y_{T-1}})$

$= \mathrm{Var}(2\varepsilon_{T+1} + \varepsilon_{T+2})$, ε_{T+1} and ε_{T+2} are iid

$= 4\sigma^2 + \sigma^2$

$= 5\sigma^2$

(c) The last 4 observations in the series are -8.775, -8.136, -7.989, and -8.533; that is, these are $Y_{T-3}, \ldots, Y_T$. I fitted the model above to the data and got the following as part of the output:

T-3 T-2 T-1 T

```
sigma^2 estimated as 0.7426:  log likelihood=-25.4
AIC=52.81   AICc=53.03   BIC=53.8
> sqrt(0.7426)
[1] 0.8617424
```

$\to \sigma$

Find a 95% prediction interval for Y_{T+2}. [3 marks]

$\hat{y}_{T+2} = 3y_T - 2y_{T-1} = -9.621$

$CI_{95} = \hat{y}_{T+2} \pm 1.96 \cdot \sqrt{5 \cdot \hat{\sigma}^2}$

$= -9.621 \pm 1.96 \cdot .8617 \cdot \sqrt{5}$

$= [-13.398, -5.844]$

(a) Do I need the drift term in my model? Explain. [3 marks]

No -> The drift is not significant w.r.t its S.E.

-> Better BIC and AIC with no drift

(b) Using the coefficient estimates printed provide an estimate of the lag 1 and lag 2 autocorrelations. [2 marks] **Warning**: this may be hard so I have cut back the marks to just 2.

no drift model

$$Y_t = \underbrace{\hat{\mu}}_{\text{intercept}} + \varepsilon_t + \underbrace{0.2145}_{\hat{\theta}_1}\varepsilon_{t-1} - \underbrace{.3182}_{\hat{\theta}_2}\varepsilon_{t-2}$$

discard $\hat{\mu}$ since cst.

$$\hat{\gamma}_0 = cov(Y_t, Y_t) = cov(\varepsilon_t + \hat{\theta}_1\varepsilon_{t-1} + \hat{\theta}_2\varepsilon_{t-2}, \varepsilon_t + \hat{\theta}_1\varepsilon_{t-1} + \hat{\theta}_2\varepsilon_{t-2})$$

$$= \hat{\sigma}^2(1 + \hat{\theta}_1^2 + \hat{\theta}_2^2)$$

$$\hat{\gamma}_1 = cov(Y_t, Y_{t-1}) = cov(\cancel{\varepsilon_t} + \hat{\theta}_1\varepsilon_{t-1} + \hat{\theta}_2\varepsilon_{t-2}, \varepsilon_{t-1} + \hat{\theta}_1\varepsilon_{t-2} + \cancel{\hat{\theta}_2\varepsilon_{t-3}})$$

$$= \hat{\theta}_1 \cdot \hat{\sigma}^2 + \hat{\theta}_2\hat{\theta}_1\hat{\sigma}^2 = \hat{\sigma}^2\hat{\theta}_1(1 + \hat{\theta}_2)$$

$$\hat{\gamma}_2 = cov(Y_t, Y_{t-1} \overset{\text{similarly}}{=} \hat{\theta}_2\hat{\sigma}^2$$

$$\Rightarrow \hat{\rho}_1 = \frac{\hat{\theta}_1(1+\hat{\theta}_2)}{(1+\hat{\theta}_1^2+\hat{\theta}_2^2)} \Rightarrow \hat{\rho}_2 = \frac{\hat{\theta}_2}{(1+\hat{\theta}_1^2+\hat{\theta}_2^2)}$$

plug in

On the basis of this I chose to fit an ARMA(1,2) model.

```
     #
     #  Now fit ARMA(1,2) with a time trend.
     #
     Fit.mort = arima(ratiomort,order=c(1,0,2),xreg=time(mortality.res))
     Fit.mort
Call:
 arima(x = ratiomort, order = c(1, 0, 2), xreg = time(mortality.res))

 Coefficients:
          ar1      ma1     ma2  intercept  time(mortality.res)
       0.7990  -0.3085  0.1433    -0.8197                4e-04
 s.e.  0.0454   0.0583  0.0486     1.1150                6e-04

 sigma^2 estimated as 7.815e-05:  log likelihood = 1680.84,  aic =
 -3349.68
```

Finally some diagnostic plots on the residuals of this fit and a Box-Ljung test.

```
     #
     #  Graphs concerning the residuals from that fit.
     #
     acf(residuals(Fit.mort))
     pacf(residuals(Fit.mort))
     qqnorm(residuals(Fit.mort))
     qqline(residuals(Fit.mort))
     #
     #  Box-Ljung test for the residuals
     #
     > LB.test(Fit.mort)

     Box-Ljung test

     data:  residuals from  Fit.mort
     X-squared = 10.0765, df = 9, p-value = 0.3443
```

can't reject null hypothesis that residuals terms are uncorrelated

(a) Do I need the linear drift term in my model? Explain. [3 marks]

No, when fitting the trend using MLE the estimate is smaller then its S.E. Indicating that the term is non-significant

(b) On the basis of the outputs and graphs discuss the ARMA(1,2) fit. Identify strengths and weaknesses if any. [4 marks]

- Both the intercept (base mean) and slope are smaller then their S.e. -> non significant, removing them would possibly simplify the model and improve AIC.

- The ARMA(1,2) seems like a good model choice
  - -> all 3 parameters estimates are significant
  - -> residual don't seem corr. by looking at acf & pacf where no lag is significant And from Box-Ljung test, null Hypothesis can't be rejected
  - -> residual seems reasonably normally dist. from qq-plot

6. The series **rtem** is the square root ($R_t = \sqrt{Y_t}$) of total annual electricity use (Y_t) in the US for a 33 year period starting in 1973. I decided to fit an ARIMA(0,1,0) model to this series and get forecasts using the **forecast** function.

```
> fit.root = Arima(rtem,order=c(0,1,0))
> fit.root
Series: rtem
ARIMA(0,1,0)

sigma^2 estimated as 706.6:  log likelihood=-150.37
AIC=302.75   AICc=302.88   BIC=304.21
> forecast(fit.root)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2006       2009.475 1975.409 2043.541 1957.376 2061.574
2007       2009.475 1961.298 2057.651 1935.795 2083.154
2008       2009.475 1950.471 2068.479 1919.236 2099.713
2009       2009.475 1941.343 2077.606 1905.276 2113.673
2010       2009.475 1933.301 2085.648 1892.978 2125.972
2011       2009.475 1926.031 2092.919 1881.858 2137.091
2012       2009.475 1919.345 2099.604 1871.633 2147.316
2013       2009.475 1913.122 2105.827 1862.116 2156.833
2014       2009.475 1907.277 2111.672 1853.177 2165.772
2015       2009.475 1901.749 2117.200 1844.723 2174.227
```

Question: Give a forecast and a 95% forecast (prediction) interval for the original series Y_t for the year 2006. [2 marks]

$R_{06} = \sqrt{Y_{06}} \Rightarrow \hat{R}_{06}^2 = \hat{Y}_{06} = (2009.475)^2$ -> compute

$IC_{95}^{Y} = [(1957.37)^2, (2061.57)^2]$

= compute