Stat 551 Final Formulas

ACGF

Definition. Suppose $\{X_t\}$ is WS with ACVF $\gamma(\cdot)$. Then its *autocovariance generating function* is $G(z) = \sum_{h=-\infty}^{\infty} \gamma(h) z^h$, provided this converges for all $z \in \mathbb{C}$ with $r^-1 < |z| < 1$ for some $|z| > 1$.

Result (Linear Processes). For $\{Z_t\} \sim WN(0, \sigma^2)$ and real-coefficients ψ_j with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, consider the linear process (WS) $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} = \psi(B) Z_t$, where $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$. Then, the ACGF of $\{X_t\}$ is given by $G(z) = \sigma^2 \psi(z) \psi(z^{-1})$.

Result (Filtered Processes). For WS $\{X_t\}$ with ACGF $G_X(z)$, and real-coefficients ψ_j with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, consider the filtered process (WS) $Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} = \psi(B) X_t$, where $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$. Then, the ACGF of $\{Y_t\}$ is given by $G_Y(z) = G_X(z) \psi(z) \psi(z^{-1})$.

Example. For $\{Z_t\} \sim WN(0, \sigma^2)$, $Y_t - \mu = Z_t + \theta_1 Z_{t-1} + \theta_{12} Z_{t-12}$, $G_Y(z) = \sigma^2 \left[(1 + \theta_1^2 + \theta_{12}^2) z^0 + \theta_1(z^1 + z^{-1}) + \theta_1 \theta_{12}(z^{11} + z^{-11}) + \theta_{12}(z^{12} + z^{-12})\right]$

Example. For $\{X_t\}$ WS ARMA (not necessarily causal or invertible), $G_X(z) = \sigma^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}$

Example. For $\{X_{1,t}\}$, $\{X_{2,t}\}$ uncorrelated WS ARMA, and $X_t = X_{1,t} + X_{2,t}$, $G_X(z) = G_{X_1}(z) + G_{X_2}(z)$, thus $X_t \sim \text{ARMA}(p, q)$ with $p \le p_1 + p_2, q \le \max(q_1 + p_2, q_2 + p_1)$

Frequency Domain

Definition (Fourier Frequencies). Defined as $\omega_j = 2\pi j/n, j \in \mathcal{F}_n$ where $\mathcal{F}_n = \{\lfloor (n-1)/2 \rfloor, \dots, -1, 0, 1, \dots, \lfloor n/2 \rfloor\}$. For $j \in \mathcal{F}_n, -\pi < w_j \le \pi$.

Definition. For $j \in \mathcal{F}_n$, define $\underset{\sim}{e}_j = \frac{1}{\sqrt{n}} \left(e^{i\omega_j}, \dots, e^{in\omega_j}\right) \in \mathcal{C}^n$.

Result. $\left\{\underset{\sim}{e}_j : j \in \mathcal{F}_n\right\}$ is an orthonormal basis for $\mathcal{C}^2$, meaning for any $\underset{\sim}{y} \in \mathcal{C}^n$, there exist complex numbers a_j's depending on $\underset{\sim}{y}$ such that $\underset{\sim}{y} = \sum_{j \in \mathcal{F}_n} a_j \underset{\sim}{e}_j$ and with respect to the inner product ($\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle = \sum_{i=1}^n x_i \overline{y_i}$), it holds

$$\langle \underset{\sim}{e}_j, \underset{\sim}{e}_k \rangle = \begin{cases} 0 & \text{if } j \ne k \in \mathcal{F}_n \\ 1 & \text{if } j = k \in \mathcal{F}_n \end{cases}$$

Result. For time series data $\underset{\sim}{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$, it holds that $\underset{\sim}{X} = \sum_{j \in \mathcal{F}_n} d_j \underset{\sim}{e}_j$ where $d_j = \{\underset{\sim}{X}, \underset{\sim}{e}_j\} \in \mathbb{C}, j \in \mathcal{F}_n$.

DFT & Periodogram

Definition. The collection $\left\{d_j = \{\underset{\sim}{X}, \underset{\sim}{e}_j\} : j \in \mathcal{F}_n\right\}$ is called the *discrete Fourier transform* of time series data $\underset{\sim}{X} = (X_1, \dots, X_n)$.

Definition. The *periodogram* of $\underset{\sim}{X} = (X_1, \dots, X_n)$ at frequency ω_j is $I_n(\omega_j) = d_j \overline{d_j} = |d_j|^2 = \frac{1}{n} |\sum_{t=1}^n X_t e^{-it\omega_j}|^2$.

Properties of $I_n(\omega_j), i \in \mathcal{F}_n$

1. At $j = 0$ or $\omega_j = \omega_0 = 0$, $I_n(0) = n(\overline{X}_n)^2$.
2. If the sample mean is subtracted and the DFT is computed for $\underset{\sim}{X} - \overline{X}_n \underset{\sim}{1} = \underset{\sim}{X} - \overline{X}_n \sqrt{n} \underset{\sim}{e}_0$, then $d_j^* = d_j - n\overline{X}_n \mathbb{I}(j = 0)$, i.e. $d_j, I_n(\omega_j)$ are unaffected by mean correction with frequencies $j \ne 0$.
3. The periodogram is symmetric: $I_n(\omega_j) = I_n(\omega_{-j})$.
4. Sum of squares total property: $\sum_{t=1}^n X_t^2 = \sum_{j \in \mathcal{F}_n} I_n(\omega_j)$. This implies that we can explain sources of variability in time series data by the size of periodogram values at different frequencies.
5. $I_n(\omega_j) = \sum_{k=-(n-1)}^{n-1} \hat{\gamma(k)} e^{-ik\omega_j}$

Spectral Densities

Definition. For a WS process $\{X_t\}$ with ACVF $\gamma(\cdot)$, the *spectral density* of $\{X_t\}$ is defined as $f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\omega}, \omega \in [-\pi, \pi] = \frac{1}{2\pi} G(e^{-i\omega})$.

Properties of f

1. $f(\omega) \ge 0, \omega \in [-\pi, \pi]$
2. f is even: $f(\omega) = f(-\omega)$.
3. for any integer k, $\int_{-\pi}^{\pi} e^{ik\omega} f(\omega) d\omega = \gamma(k)$.

Example. $\{Z_t\} \sim WN(0, \sigma^2)$, $f(\omega) = \sigma^2/[2\pi]$.

Result (Filtered Processes). For WS $\{X_t\}$ with ACGF $G_X(z)$ and spectral density $f_X(\cdot)$, and real-coefficients ψ_j with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, consider the filtered process (WS) $Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} = \psi(B) X_t$, where $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$. Then, the ACGF of $\{Y_t\}$ is given by $G_Y(z) = G_X(z) \psi(z) \psi(z^{-1})$ and the spectral density of $\{Y_t\}$ is $f_Y(\omega) = f_X(\omega) |\psi(e^{i\omega})|^2, \omega \in [-\pi, \pi]$.

Example. WS ARMA $\phi(B) X_t = \theta(B) Z_t, \{Z_t\} \sim WN(0, \sigma^2), \phi(z) \ne 0$ for $|z| = 1$. Then $f_X(\omega) = \frac{|\theta(e^{i\omega})|^2}{|\phi(e^{i\omega})|^2} \frac{\sigma^2}{2\pi}, \omega \in [-\pi, \pi]$.

Power Transfer Function

Definition. For WS $\{X_t\}$ and real-coefficients ψ_j with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, consider the WS filtered process $Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} = \psi(B) X_t$, where $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$. The *power transfer function* of the filter $\{\psi_j\}$ is given by $|\psi(e^{i\omega})|^2, \omega \in [-\pi, \pi]$.

Estimating Spectral Densities

Result. For $\omega = 0$ and $EX_t = \mu$, it holds that $\lim_{n \to \infty} \mathrm{E}\left[I_n(0) - n\mu^2\right] = 2\pi f(0) = \sum_{k=-\infty}^{\infty} \gamma(k)$.

Under mild conditions, it also holds, for large n $\overline{X}_n \overset{\bullet}{\sim} N\left(\mu, \frac{2\pi f(0)}{n}\right)$ or $\overline{X}_n \overset{\bullet}{\sim} N\left(\mu, \frac{1}{n} \sum_{k=-\infty}^{\infty} \gamma(k)\right)$.

Result. Define the periodogram $I_n(\omega) = \frac{1}{n} |\sum_{t=1}^n X_t e^{-it\omega}|^2$ at any frequency $\omega \in [-\pi, \pi]$. For fixed $\omega \in [-\pi, \pi], \omega \ne 0$, it holds that $\lim_{n \to \infty} \mathrm{E}\left[\frac{I_n(\omega)}{2\pi}\right] = f(\omega)$.

Result. For $\{Z_t\} \sim IID(0, \sigma^2)$ and real-coefficients ψ_j with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, consider the linear process (WS) $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ with spectral density f. If $f(\omega) > 0$ for all $\omega \in [-\pi, \pi]$ and $0 < \lambda_1 < \cdots < \lambda_m < \pi$ denote a given set of fixed frequencies, then $\frac{1}{2\pi}[I_n(\lambda_1), \dots, I_n(\lambda_m)] \xrightarrow{d} (A_1, \dots, A_m)$ as $n \to \infty$, where $A_1, \dots, A_m$ are independent random variables and A_i has an exponential distribution with mean $f(\lambda_i)$. Equivalently, $\frac{2A_i}{f(\lambda_i)} \sim \chi_2^2, i = 1, \dots, m$.

Result. For $\{X_t\} \sim IIDN(0, \sigma^2)$, the periodogram values $\left\{\frac{I_n(\omega_j)}{2\pi} : \omega_j \in \mathcal{F}_n, \omega_j \notin \{0, \pi\}\right\}$ are IID Exponential$[\sigma^2/[2\pi]]$ random variables.

Note. The periodogram is not a consistent estimator of the spectral density: for a fixed $0 < \lambda < \pi$ as $n \to \infty$, $I_n(\lambda)/[2\pi] \xrightarrow{d} \text{Exp}[f(\lambda)]$, but $I_n(\lambda)/[2\pi] \overset{p}{\nrightarrow} f(\lambda)$.

Window estimator

- Let $W_n(\cdot)$ denote a weight function, m_n denote a bandwidth, $\omega_j \in \mathcal{F}_n$ be closest to λ
- Given $\lambda \in [0, \infty]$, define a window estimator of $f(\lambda)$ as $\hat{f}(\lambda) = \sum_{|k| \le m_n} W_n(k) \frac{I_n(w_j + k)}{2\pi}$ where the following hold:
  1. $m_n \to \infty$ as $n \to \infty$
  2. $m_n/n \to 0$ as $n \to \infty$
  3. $W_n(k) = W_n(-k)$
  4. $\sum_{|k| \le m_n} W_n(k) = 1$
  5. $\sum_{|k| \le m_n} [W_n(k)]^2 \to 0$ as $n \to \infty$
- the estimator $\hat{f}(\lambda)$ is MSE-consistent for $f(\lambda)$.
- Approximate $\hat{f} \overset{d}{=} cY$ where $Y \sim \chi_\nu^2$. Then choose c & ν to get ok distributional approximation by equating the first two moments to get $\nu = \frac{2}{\sum_{|k| \le m_n} [W_n(k)]^2}$ and $c = f/\nu$.
- An approximate $100(1 - \alpha)\%$ confidence interval for f is $\left[\frac{\nu \hat{f}}{\chi_\nu^2(1 - \alpha/2)}, \frac{\nu \hat{f}}{\chi_\nu^2(\alpha/2)}\right]$
- An approximate $100(1 - \alpha)\%$ confidence interval for $\ln f$ is $\left[\ln \hat{f} + \ln \nu - \ln \chi_\nu^2(1 - \alpha/2), \ln \hat{f} + \ln \nu - \ln \chi_\nu^2(\alpha/2)\right]$ give intervals with same lengths each λ.

Result (Consistency). Define $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ for $\{Z_t\} \sim IID(0, \sigma^2)$ with $\mathrm{E}|Z_t|^4 < \infty$ and $\sum_{j=-\infty}^{\infty} |\psi_j| \sqrt{|j|} < \infty$. Then, as $n \to \infty$, $\mathrm{E}\hat{f}(\lambda) \to f(\lambda)$ and

$$\frac{\mathrm{Cov}[\hat{f}(\lambda), \hat{f}(\omega)]}{\sum_{|k| \le m_n} [W_n(k)]^2} \to \begin{cases} 2[f(\lambda)]^2 & \lambda = \omega \in \{0, \pi\} \\ [f(\lambda)]^2 & \lambda = \omega \notin \{0, \pi\} \\ 0 & \lambda \ne \omega \end{cases}.$$

Hypothesis Testing

Test for hidden periodicity.

H_0 :$\{X_t\} IIDN(0, \sigma^2)$ so that $f(\omega) = \sigma^2/[2\pi]$ is constant

H_1 :some periodic component at an unspecified frequency

Compare largest periodogram ordinate to the average periodogram ordinate

$$K = \frac{\max_{1 \le j \le q} I_n(\omega_j)}{\frac{1}{q} \sum_{k=1}^{q} I_n(\omega_k)} = q \max_{1 \le j \le q} (y_j - y_{j-1}$$

where $y_0 = 0$ and $y_j = \sum_{k=1}^{j} I_n(\omega_k) / \sum_{k=1}^{q} I_n(\omega_k)$. If $u_1, \ldots, u_{q-1} \sim IIDUnif(0,1)$, then $y_1, \ldots, y_{q-1} \stackrel{d}{=} u_{(1)}, \ldots, u_{(q-1)}$. Reject H_0 for large values of K and use the following to compute a p-value: $P(K \le x) = 1 - \sum_{j=0}^{q} (-1)^j \binom{q}{j} \left[\left(1 - \frac{jx}{q}\right)_+ \right]^{q-1}$

Model Fitting for ARMA

Yule-Walker Estimation in AR(p)

- Substitute sample ACVF values $\hat{\gamma(0)}, \ldots, \hat{\gamma(p)}$ for $\gamma(0), \ldots, \gamma(p)$ in $\underset{\sim}{\phi}_p = \Gamma_p^{-1} \underset{\sim}{\gamma_p}$ then $\underset{\sim}{\phi}_{YW} = \hat{\Gamma}_p^{-1} \underset{\sim}{\hat{\gamma}_p}$ of $\underset{\sim}{\phi}_p = (\phi_1, \ldots, \phi_p)'$ and $\hat{\sigma}^2 = \hat{\gamma}(0) - \underset{\sim}{\hat{\phi}}' \underset{\sim}{\hat{\gamma}}_p$.
- For data $X_1, \ldots, X_n$ from AR(p) process, it holds that as $n \to \infty$, $\sqrt{n}\left(\underset{\sim}{\hat{\phi}}_{YW} - \underset{\sim}{\phi_p}\right) \xrightarrow{d} N\left(\underset{\sim}{0_p}, \sigma^2 \Gamma_p^{-1}\right)$.
- Special cases:

  AR(1) $\hat{\phi}_{YW} \stackrel{\bullet}{\sim} N\left(\phi, \frac{1-\phi^2}{n}\right)$

  AR(2) $\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} \stackrel{\bullet}{\sim} N\left[\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \frac{1}{n}\begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}\right]$
- YW estimator is the MoM estimator for AR(p).
- matches sample and theoretical ACVF, ACF, PACF at lags $h = 0, \ldots, p$.
- fitted model will be causal.

Hannan-Rissanen Estimation for preliminary estimates

1. Pick $m < n$ and run D-L algorithm with sample ACVF $\hat{\gamma}(\cdot)$ to get estimates $\hat{\phi}_{m,1}, \ldots, \hat{\phi}_{m,m}$. Then, obtain estimates of residuals as $\hat{Z}_t = X_t - \hat{\phi}_{m,1} X_{t-1} - \cdots - \hat{\phi}_{m,m} X_{t-m}, t = m+1, \ldots, n$.
2. Regress X_t on $X_{t-1}, \ldots, X_{t-p}, \ldots, \hat{Z}_{t-1}, \ldots \hat{Z}_{t-q}$ to produce parameter estimates for $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$.

Maximum Likelihood Estimation

Definition. Write ARMA model parameters as $\underset{\sim}{\Psi} = (\sigma^2, \underset{\sim}{\phi}, \underset{\sim}{\theta})$. The *likelihood* of the data $\underset{\sim}{X}_n$ is given by $L(\underset{\sim}{\Psi} | \underset{\sim}{X}_n) \equiv$ joint distribution of $X_1, \ldots, X_n$ regarded as a function of $\underset{\sim}{\Psi}$.
Note the "chain rule":
$L(\underset{\sim}{\Psi} | \underset{\sim}{X}_n) = P_{\underset{\sim}{\Psi}}(X_n | \underset{\sim}{X}_{n-1}) P_{\underset{\sim}{\Psi}}(X_{n-1} | \underset{\sim}{X}_{n-2}) \cdots P_{\underset{\sim}{\Psi}}(X_2 | \underset{\sim}{X}_1) P_{\underset{\sim}{\Psi}}(X_1)$.

To concentrate or profile σ^2 out of the log-likelihood, we want to maximize over σ^2 given the other parameter values $\underset{\sim}{\phi}, \underset{\sim}{\theta}$.

Definition. Equivalently, we can minimize the *reduced likelihood*: $\updownarrow(\underset{\sim}{\phi}, \underset{\sim}{\theta} | \underset{\sim}{X}_n) = \log(\hat{\sigma}^2) + \frac{1}{n}\sum_{t=1}^{n} \log(r_{t-1})$ where $\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \frac{(X_t - \hat{X}_t)^2}{r_{t-1}}$ and $r_{t-1} = \frac{\mathrm{E}(X_t - \hat{X}_t)^2)}{\sigma^2}$.

We minimize the reduced likelihood numerically to obtain MLEs $\underset{\sim}{\hat{\phi}}_{MLE}, \underset{\sim}{\hat{\theta}}_{MLE}$.
Then compute $\hat{X}_t(\underset{\sim}{\hat{\phi}}_{MLE}, \underset{\sim}{\hat{\theta}}_{MLE}), r_{t-1}^2(\underset{\sim}{\hat{\phi}}_{MLE}, \underset{\sim}{\hat{\theta}}_{MLE}), t = 1, \ldots, n$ and obtain $\hat{\sigma}^2_{MLE}$.

Asymptotic Distribution of $\underset{\sim}{\hat{\phi}}_{MLE}, \underset{\sim}{\hat{\theta}}_{MLE}$ in ARMA

ARMA process $\phi(B)X_t = \theta(B)Z_t, Z_t \sim WN(0, \sigma^2)$.
• Let $Z_t^* \sim WN(0, \sigma^2)$ and define two new AR processes U_t and V_t:

$$U_t = \frac{Z_t^*}{\phi(B)} \sim AR(p), \qquad V_t = \frac{Z_t^*}{\theta(B)} \sim AR(q)$$

• Then, for large n,

$$\begin{pmatrix} \underset{\sim}{\hat{\phi}}_{MLE} \\ \underset{\sim}{\hat{\theta}}_{MLE} \end{pmatrix} \stackrel{\bullet}{\sim} MVN\left[\begin{pmatrix} \underset{\sim}{\phi} \\ \underset{\sim}{\theta} \end{pmatrix}, \frac{\sigma^2}{n}\begin{pmatrix} \mathrm{E}\underset{\sim}{U}_p \underset{\sim}{U}_p' & \mathrm{E}\underset{\sim}{U}_p \underset{\sim}{V}_p' \\ \mathrm{E}\underset{\sim}{V}_p \underset{\sim}{U}_p' & \mathrm{E}\underset{\sim}{V}_p \underset{\sim}{V}_p' \end{pmatrix}\right]$$

Order Selection & Diagnostics

Rough rule of thumb: if sample ACF ≈ 0 for lags $> q$, try MA(q), if sample PACF ≈ 0 for lags $> p$, try AR(p). Doesn't work for ARMA(p,q).

Definition. Final Prediction Error (FPE) for selecting AR(p): $\mathrm{FPE}(p) = \left[\frac{\hat{\sigma}^2 n}{n-p}\left(1 + \frac{p}{n}\right)\right]$ where $\hat{sigma}^2$ is the MLE for AR(p). Try to minimize FPE as a function of p, or minimize $\log(\mathrm{FPE}(p)) \approx \log(\hat{\sigma}^2) + \frac{2p}{n}$. The FPE estimates $\mathrm{E}_{\theta_{p0}}\left[\left\{Y_{n+1} - \hat{Y}_{n+1}(\underset{\sim}{\hat{\phi}}_p)\right\}^2\right]$

Definition. Select model by minimizing Akaike's Information Criterion $\mathrm{AIC} = -2\log L(\underset{\sim}{\hat{\psi}}_m | \underset{\sim}{X}_n)) + 2m$ is approximately unbiased estimator of expected out-of-sample likelihood $\mathrm{E}_{\psi_{m0}}\left[-2\log L(\underset{\sim}{\hat{\psi}}_m | \underset{\sim}{X}_n))\right]$

For pure ARMA(p,q), log[criterion] $\approx \log(\hat{\sigma}^2)$ + penalty

Criterion	Penalty
FPE (AR only)	$\frac{2p}{n}$
AIC	$2(p+q+1)$
BIC	$(p+q+1)log(n)$
AICC	$\frac{2n(p+q+1)}{n-p-q-2}$

Models within AIC/AICC of 2 are equivalent.

Diagnostics

Get MLEs $\underset{\sim}{\hat{\psi}}_{MLE}, \underset{\sim}{\hat{\theta}}_{MLE}$

Form predictions $\hat{X}_t \equiv \hat{X}_t(\underset{\sim}{\hat{\psi}}_{MLE}, \underset{\sim}{\hat{\theta}}_{MLE}), r_t \equiv r_t(\underset{\sim}{\hat{\psi}}_{MLE}, \underset{\sim}{\hat{\theta}}_{MLE})$.

Get residuals $W_t = \frac{X_t - \hat{X}_t}{\sqrt{r_{t-1}}} \approx Z_t$

Residuals $\{W_t\}$ should look like WN. Can conduct tests for WN: sample ACF (Bartlett bounds), PACF, Portmanteau, rank-based.
Comments
1. $\{W_t\}$ are not iid/WN, they depend on the MLEs.
2. sample ACF/PACF of residuals $\{W_t\}$ can suggest model improvement. ACF – increase MA order if see residual outside Bartlett bounds, PACF for AR order.

State Space Models

Models in *state space* allow unified treatment, recursive prediction (Kalman Filter), likelihood via prediction error/innovation decomp, missing values, extensions to non-linear/non-Gaussian models.
State Space Form Observation Equation: $\boldsymbol{Y}_t = \boldsymbol{G}_t \boldsymbol{X}_t + \boldsymbol{W}_t$ (wx1, (wxv)(vx1), wx1)
State Equation: $\boldsymbol{X}_{t+1} = \boldsymbol{F}_t \boldsymbol{X}_t + \boldsymbol{V}_t$ (vx1, (vxv)(vx1), vx1)
Assume for the error terms $\boldsymbol{W}_t, \boldsymbol{V}_t$, uncorrelated random vectors, expected values 0, and

$$\mathrm{Var}\begin{pmatrix} \boldsymbol{W}_t \\ \boldsymbol{V}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_t & \boldsymbol{S}_t \\ \boldsymbol{S}_t' & \boldsymbol{R}_t \end{pmatrix}$$

For each t, $\boldsymbol{W}_t$ & $\boldsymbol{V}_t$ uncorrelated with $\{\boldsymbol{X}_s : s \le t\}$.

Example (Random Walk + Noise)**.**

obs eqn $: Y_t = X_t + W_t$

state eqn $: X_{t+1} = X_t + V_t$

for $\{W_t\} \sim WN(0, \sigma_w^2, \{V_t\} \sim WN(0, \sigma_v^2$, uncorrelated. Here $F_t = 1, G_t = 1, Q_t = \sigma_v^2, R_t = \sigma_w^2, S_t = 0$ time invariant, but NOT WS

Example (Co-integration)**.** Real-values $\{Y_t\} \sim I(d)$ if $\{(1-B)^d Y_t\}$ is WS, but $\{(1-B)^{d-1} Y_t\}$ is NOT WS. Random vector is $I(d)$ if each component is. $\{\boldsymbol{Y}_t\} \sim I(d)$ is cointegrated with cointegrating factor $\boldsymbol{\alpha}$ if $\{\boldsymbol{\alpha}' \boldsymbol{Y}_t\} \sim I(k)$ for some $k < d$

obs eqn $: \boldsymbol{Y}_t = \begin{pmatrix} 1 \\ \gamma \end{pmatrix} d_t + \begin{pmatrix} 0 \\ w_t \end{pmatrix}$

state eqn $: d_{t+1} = d_t + v_t$

Example (AR(p))**.**
$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \{Z_t\} \sim WN(0, sigma^2)$

$$\text{state eqn }: \begin{pmatrix} X_{t+1} \\ X_t \\ \vdots \\ X_{t+2-p} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t+1-p} \end{pmatrix} + \begin{pmatrix} Z_{t+1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

obs eqn $: Y_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{X}_t + 0 = X_t$

Example (MA(1))**.** $X_t = \phi_1 Z_t + \phi Z_{t-1}, \{Z_t\} \sim WN(0, sigma^2)$
Let (2×1) $\boldsymbol{X}_t = \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}$ and write

state eqn $: \boldsymbol{X}_{t+1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \boldsymbol{X}_t + \begin{pmatrix} Z_{t+1} \\ \theta Z_{t+1} \end{pmatrix}$

obs eqn $: Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{X}_t + 0 = X_{1,t} = \theta Z_{t-1} + Z_t$

Example (Linear Model)**.** $Y_t = \boldsymbol{Z}_t' \boldsymbol{\beta} + W_t$ with regressors $\boldsymbol{Z}_t$ & $\{W_t\} \sim WN(0, \sigma^2)$

obs eqn $: Y_t = \boldsymbol{Z}_t' \boldsymbol{\beta} + W_t$

state eqn $: \boldsymbol{X}_{t+1} = \boldsymbol{\beta} = \boldsymbol{X}_t + \boldsymbol{0}$

Kalman Filter

1. *Startup* $\hat{\boldsymbol{X}}_1 = \boldsymbol{X}_{1|0}$ = projection of $\boldsymbol{X}_1$ onto $\boldsymbol{Y}_0$ (some startup values) often $\hat{\boldsymbol{X}}_1 = \mathrm{E}(\boldsymbol{X}_1)$,
   $\boldsymbol{\Omega}_1 = \mathrm{E}\left[(\boldsymbol{X}_1 - \hat{\boldsymbol{X}}_1)(\boldsymbol{X}_1 - \hat{\boldsymbol{X}}_1)'\right]$
2. *Innovation* at $t \ge 1$ (new $\boldsymbol{Y}_t$ becomes available in addition to $\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_{t-1}$):

   $\boldsymbol{I}_t = \boldsymbol{G}_t(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t) + \boldsymbol{W}_t$

   $\boldsymbol{\Delta}_t = \mathrm{Var}(\boldsymbol{I}_t) = \boldsymbol{G}_t \boldsymbol{\Omega}_t \boldsymbol{G}_t' + \boldsymbol{R}_t$
3. *Filter or Update* at $t \ge 1$

   $\boldsymbol{X}_{t|t} = \hat{\boldsymbol{X}} + \boldsymbol{\Omega}_t \boldsymbol{G}_t' \boldsymbol{\Delta}_t^{-1} \boldsymbol{I}_t$

   $\boldsymbol{\Omega}_{t|t} = \boldsymbol{\Omega}_t - \boldsymbol{\Omega}_t \boldsymbol{G}_t' \boldsymbol{\Delta}_t^{-1} \boldsymbol{G}_t \boldsymbol{\Omega}_t$

   Note $\boldsymbol{X}_{t|t}$ is $\mathrm{E}(\boldsymbol{X}_t | \boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_t)$ assuming Gausian processes, or the best linear predictor of $\boldsymbol{X}_t$ given $\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_t$.
4. *Predict* at $t \ge 1$

   $\hat{\boldsymbol{X}}_{t+1} = \boldsymbol{F}_t \boldsymbol{X}_{t|t}$

   $\boldsymbol{\Omega}_{t+1} = \boldsymbol{F}_t \boldsymbol{\Omega}_{t|t} \boldsymbol{F}_t' + \boldsymbol{Q}_t$