Forecasting Summary

Good to Know

Important Calculus:

$$\emptyset_1 = \beta_1$$

Expected Value Properties

$$E(a) = a$$
$$E(aX) = a^{*}E(X)$$
$$E(X + a) = E(X) + a$$
$$E(X + Y) = E(X) + E(Y)$$
$$\text{If } X \text{ and } Y \text{ ind.:}$$
$$E(XY) = E(X)^{\star}E(Y)$$

Example:

$$y_t = \mu + \ \delta y_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$$
$$E[y_t] = \mu + \delta E[y_{t-1}] + 0 + 0$$
$$From\ stationary\ property:$$
$$|\delta| < 1 =>$$
$$E[y_t] = E[y_{t-s}] + s$$
$$E[y_t] = \frac{\mu}{1 - \delta}$$

Variance Properties

$$\text{Var}(X \pm a) = \text{Var}(X)$$
$$\text{Var(aX)} = \text{a}^{2\star}\,\text{Var(X)}$$
$$\text{Var}\ (X^2) = \text{Var}\ (X) + E(X)^2$$

If $X\&Y$ ind.

$$\text{Var}\ (X + Y) = \text{Var}\ (X) + \text{Var}\ (Y)$$

$$\text{Var}\ (X) = E(X^2) - (E(X))^2$$
$$E(X^2) = \text{Var}\ (X) + (E(X))^2$$

Example:

$$V(y_t) = \delta V(y_t - 1\ ) + \sigma^2 + \theta^2\sigma^2,$$
$$|\delta| < 1 => \ \ V(y_t) = V(y_{t+1})$$
$$V(y_t) = \frac{\sigma^2(1 + \delta)}{1 - \sigma^2}$$

Covariance Properties

$$\text{cov}\,(X, \text{a}) = 0$$
$$\text{cov}\,(X, X) = v\text{ar}\,(X)$$
$$\text{cov}\,(X, Y) = \text{cov}\,(Y, X)$$
$$\text{cov}\,(aX, bY) = ab\text{cov}\,(X, Y)$$
$$\text{cov}\,(X + a, Y + b) = \text{cov}\,(X, Y)$$
$$\text{cov}\,(aX + bY, cW) = ac\text{cov}\,(X, W) + bc\text{cov}\,(X, W)$$

Correlation Properties

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) * Var(Y)}}$$

Stationarity

- Series exists in a finite band and has a constant **mean and constant variance**. Plus, we have an AR process that's going to gradually fall towards the overall mean.
- Eigenvectors vs. Roots to test

Properties

Mean

$$\mathbb{E}[Y_t] = \mathbb{E}\left[\lim_{n\to\infty} \sum_{j=-n}^{n} \psi_j W_{t-j}\right] = \lim_{n\to\infty} \mathbb{E}\left[\sum_{j=-n}^{n} \psi_j W_{t-j}\right] = 0$$

Autocovariance [needs mean is zero]

$$\gamma(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|h|} \sigma_W^2$$

Lagged Values

- Past values

$$LY_t = Y_{t-1}$$

Properties:

$$L^2 Y_t = L(LY_t) = LY_{t-1} = Y_{t-2}$$
$$L^j L^i Y_t = L^{i+j} Y_t = Y_{t-i-j}$$
$$Lc = c$$
$$\left(L^i + L^j\right) Y_t = Y_{t-i} + Y_{t-j}$$

L to a negative power is a lead operator: $L^{-i} Y_t = Y_{t+i}$

For $|a| < 1$:

$$1 + aL + a^2 L^2 + \cdots + a^k L^k + \cdots = (1 - aL)^{-1} Y_t = \frac{Y_t}{1 - aL}$$

⇔ Backward shift operation B

$$x_t = x_{t-1} + w_t = Bx_t + w_t$$

⇔ Forward shift operation B

$$B^{-1}x_t = x_{t+1}$$

Non-Stationarity

- If series has a unit root
- ACF indicates non-stationarity
- Dickey-Fuller ~~or KPSS indicated root~~
- Differencing is required

White Noise

- Independent and identically distributed (i.i.d) of the series.
- Mean of Zero.
- No serial Correlation
- Residual Error Series: $x_t = y_y - \widehat{y_t}$
- Residual: i.i.d
- Gaussian White Noise: $w_t \sim N(0, \sigma^2)$

Properties

$$x_t = x_{t-1} + w_t$$

$$\mu_w = E[w_t] = 0$$

$$\gamma_k(t) = Cov(x_t, x_{t+k}) = t\sigma^2$$

Takeaway: In particular, the covariance is equal to the variance multiplied by the time. Hence, as time increases, so does the variance.

Loss Function Semantic

stimate.

symbol	name	equation
$\mathcal{L}_1$	L_1 loss	$\|\mathbf{y} - \mathbf{o}\|_1$
$\mathcal{L}_2$	L_2 loss	$\|\mathbf{y} - \mathbf{o}\|_2^2$
$\mathcal{L}_1 \circ \sigma$	expectation loss	$\|\mathbf{y} - \sigma(\mathbf{o})\|_1$
$\mathcal{L}_2 \circ \sigma$	regularised expectation loss[1]	$\|\mathbf{y} - \sigma(\mathbf{o})\|_2^2$
$\mathcal{L}_\infty \circ \sigma$	Chebyshev loss	$\max_j \lvert\sigma(\mathbf{o})^{(j)} - \mathbf{y}^{(j)}\rvert$
hinge	hinge [13] (margin) loss	$\sum_j \max(0, \frac{1}{2} - \hat{\mathbf{y}}^{(j)}\mathbf{o}^{(j)})$
hinge^2	squared hinge (margin) loss	$\sum_j \max(0, \frac{1}{2} - \hat{\mathbf{y}}^{(j)}\mathbf{o}^{(j)})^2$
hinge^3	cubed hinge (margin) loss	$\sum_j \max(0, \frac{1}{2} - \hat{\mathbf{y}}^{(j)}\mathbf{o}^{(j)})^3$
log	log (cross entropy) loss	$-\sum_j \mathbf{y}^{(j)} \log \sigma(\mathbf{o})^{(j)}$
$\log^2$	squared log loss	$-\sum_j [\mathbf{y}^{(j)} \log \sigma(\mathbf{o})^{(j)}]^2$
tan	Tanimoto loss	$\frac{-\sum_j \sigma(\mathbf{o})^{(j)}\mathbf{y}^{(j)}}{\|\sigma(\mathbf{o})\|_2^2 + \|\mathbf{y}\|_2^2 - \sum_j \sigma(\mathbf{o})^{(j)}\mathbf{y}^{(j)}}$
D_{CS}	Cauchy-Schwarz Divergence [3]	$-\log \frac{\sum_j \sigma(\mathbf{o})^{(j)}\mathbf{y}^{(j)}}{\|\sigma(\mathbf{o})\|_2 \|\mathbf{y}\|_2}$

- L1: Least Absolute Function
- L2: Least Square Error

AR: Autoregressive Model

<table>
<tr><th></th><th></th></tr>
<tr><td>

AR(1)

First-order Autoregression Model

</td><td>

$$Y_t = \delta + \emptyset_1 Y_{t-1} + \epsilon$$

Where $\emptyset$ is the slope and δ is drift

Expectation

We have that Expectation should be the same for all timestamps

$$E[Y_t] = E[Y_{t-1}] = \mu$$

Thus,

$$E[Y_t] = E[\delta + \emptyset_1 Y_{t-1} + \epsilon]$$
$$\mu = \delta + \emptyset_1 E[Y_{t-1}] + E[\epsilon]$$
$$\mu = \delta + \emptyset_1 \mu + 0$$
$$\mu = \delta + \emptyset_1 \mu$$
$$(1 - \emptyset_1)\mu = \delta$$
$$\mu = \frac{\delta}{(1 - \emptyset_1)}$$

Variance

$\mathrm{Var}(y_t) = \mathrm{var}(\delta_t) + \mathrm{Var}(\emptyset_1 y_{t-1}) + \mathrm{Var}(\epsilon_t)$

$\mathrm{Var}(y_t) = 0 + \emptyset_1^2 \mathrm{Var}(y_{t-1}) + \sigma_\epsilon^2$

Stationary process:

$\mathrm{Var}(y_t) = \emptyset_1^2 \ \mathrm{var}(y_t) + \sigma_\epsilon^2$

$1 = \emptyset_1^2 * 1 + \frac{\sigma_\epsilon^2}{var(y_t)}$

$\mathrm{Var}\ (y_t) = \frac{\sigma_e^v}{1 - \phi_1^2}$

Stationarity

$$|\varphi| < 1$$

Or

$$|\varphi| - 1 < 0$$

Autocovariance function:

$$\gamma_k = \emptyset \gamma_{k-1} \ for\ k = 1,2,..and\ \ \gamma_0 = \sigma_z^2$$

It shows us that since $|\emptyset| < 1$ the dependence between observations decreases when the lag increases

AR(1) derivative:

</td></tr>
</table>

Subtract Y_t $from\ both\ sides$

We have

$$\Delta\, Y_t = Y_t - Y_{t-1}$$

$$\begin{aligned}
&Y_t = \ \text{Ø}_1 Y_{t-1} + \epsilon \\
&Y_t - Y_{t-1} = \ \text{Ø}_1 Y_{t-1} + \epsilon - Y_{t-1} \\
&Y_t - Y_{t-1} = \ \text{Ø}_1 Y_{t-1} - Y_{t-1} + \epsilon \\
&Y_t - Y_{t-1} = (\text{Ø}_1 - 1) Y_{t-1} + \epsilon \\
&\Delta\, Y_t = \ (\text{Ø}_1 - 1) Y_{t-1} + \epsilon
\end{aligned}$$

Autocorrelation Function (ACF) of a stationary AR(1)

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi \frac{\gamma_{j-1}}{\gamma_0} = \phi \rho_{j-1} \ \ j \geq 1$$

$$\rho_j = \phi^2 \rho_{j-2} = \phi^3 \rho_{j-3}$$

$$= \cdots = \phi^j \rho_0 = \phi^j$$

$$\begin{aligned}
&\gamma_0 = cov(x_t, x_t) = var(x_t) = \ \sigma^2_{x_t} \\
&\gamma_1 = cov(x_t, x_{t-1}) \\
&= cov(p_1 x_{t-1} \ + \epsilon, x_{t-1}) \\
&= p_1 \ cov(x_{t-1}, x_{t-1}) \\
&= p_1 * \gamma_0
\end{aligned}$$

$$\begin{aligned}
&\gamma_1 = p_1 * \gamma_0 \\
&p_1 = \frac{\gamma_1}{\gamma_0}
\end{aligned}$$

$$\begin{aligned}
&\gamma_2 = cov(x_t, x_{t-2}) \\
&= cov(p_1 x_{t-1} + \epsilon_t, x_{t-2}) \\
&= \ cov(p_1(p_1 x_{t-2} + \epsilon_{t-2}) + \epsilon_t, x_{t-2}) \\
&= cov(p_1^2 x_{t-2+} p_1 \epsilon_{t-2} + \epsilon_t, x_{t-2}) \\
&= p_1^2 cov(x_{t-2}, x_{t-2}) \\
&= p_1^2 \gamma_0
\end{aligned}$$

Autocovariance (ACVF) of a stationary AR(1)

Rewrite the process as $(Z_t - \mu) = \phi(Z_{t-1} - \mu) + a_t$

$$\begin{aligned}
\gamma_j &= E\big[(Z_t - \mu)\big(Z_{t-j} - \mu\big)\big] \\
&= E\big[(\phi(Z_{t-1} - \mu) + a_t)\big(Z_{t-j} - \mu\big)\big]
\end{aligned}$$

$$\begin{aligned}
&= \phi E\big[(Z_{t-1} - \mu)\big(Z_{t-j} - \mu\big) + a_t\big(Z_{t-j} - \mu\big)\big] \\
&= \phi \gamma_{j-1}
\end{aligned}$$

$$\gamma_j = \phi \gamma_{j-1} \, j \geq 1$$

AR(2) **Second order Autoregression Model**	$$Y_t = \delta + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \epsilon$$ **Stationarity** Second order polynomial Equation The characteristic equation: $$1 - \varphi_1 x - \varphi_2 x^2 = 0$$ Solution is $$x = \frac{\varphi_1 \sqrt{\varphi_1^2 + 4\varphi_2}}{-2\varphi_2}$$ Solutions referred w1 and w2. $$w1 < 1 \,\&\, w2 < 1$$ Or solution: $$\begin{gathered} \lvert\varphi_2\rvert < 1 \\ \varphi_2 + \varphi_1 < 1 \\ \varphi_2 - \varphi_1 < 1 \end{gathered}$$ Eigenvalue's method: $$\tilde{y}_t = \tilde{\mu} + \boldsymbol{A}\tilde{y}_{t-1} + \vec{\varepsilon}_t$$ $$\boldsymbol{A} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \Rightarrow \lvert\boldsymbol{A} - \lambda\boldsymbol{I}\rvert = \mathbf{det} \begin{bmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{bmatrix} = -(\phi_1 - \lambda)\lambda - \phi_2$$ $$\begin{gathered} \lambda_1\lambda_2 = \phi_2 \Rightarrow \lvert\lambda_1\lambda_2\rvert = \lvert\phi_2\rvert < 1 \\ \lambda_1 + \lambda_2 = \phi_1 \Rightarrow \lvert\lambda_1 + \lambda_2\rvert = \lvert\phi_1\rvert < 2 \end{gathered}$$
AR(p)	$$X_t - \phi_1 X_{t-1} + \phi_2 X_{t-1} + \cdots + \phi_p X_{t-j} + Z_t$$ **Autoregressive polynomial** $$\begin{gathered} \varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p \\ \varphi(B)(Y_t - \mu) = W_t \end{gathered}$$ **Autocorrelation Function (ACF)** Its recursive $$\rho_k = \phi_1\rho_{k-1} + \cdots + \phi_p\rho_{k-p}, k > q + 1$$ **Autocovariance Function (ACVF)** $$\gamma_k = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p}, k > q + 1$$ Partial Autocorrelation Function (PACF)

	$$\Phi_{kk} = \beta_k^{(k)} = \frac{\begin{vmatrix} 1 & \rho_x(1) & . & \rho_x(1) \\ \rho_x(1) & 1 & . & \rho_x(2) \\ . & . & . & . \\ \rho_x(k-1) & \rho_x(k-2) & . & \rho_x(k) \end{vmatrix}}{\begin{vmatrix} 1 & \rho_x(1) & . & \rho_x(k-1) \\ \rho_x(1) & 1 & . & \rho_x(k-2) \\ . & . & . & . \\ \rho_x(k-1) & \rho_x(k-2) & . & 1 \end{vmatrix}}$$ $$\Phi_{11} = \beta_1^{(1)} = \frac{\lvert\rho(1)\rvert}{\lvert 1\rvert} = \rho(1)$$ $$\Phi_{22} = \beta_2^{(2)} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)}$$

MA: Moving Average Model

- For MA(1), Auto Correlation Function (ACF) and Autocovariance Functions are zero, for lag k >1
- For MA(2), Auto Correlation Function (ACF) and Autocovariance Functions are zero, for lag k >2

MA(1)	$$X_{\mathrm{t}} = \mu + \mathrm{a}_{\mathrm{t}} + \theta \mathrm{a}_{\mathrm{t-1}}$$ **Expectation** $E(\mathrm{X_t}) = \mu + E(a_t) + \theta E(a_{t-1}) = \mu$ **Variance** $$\mathrm{Var}\,(X_t) = E(X_t - \mu)^2 = E(a_t + \theta a_{t-1})^2 =$$ $$= E(a_t^2 + \theta^2 a_{t-1}^2 + 2\theta a_t a_{t-1}) = \sigma_a^2(1 + \theta^2)$$ **Auto Covariance Function (ACVF)** **Autocovariance 1st. order** $$E[(X_t - \mu)(X_{t-1} - \mu)] = E[(a_t + \theta a_{t-1})(a_{t-1} + \theta a_{t-2})] =$$ $$= E[(a_t a_{t-1} + \theta a_{t-1}^2 + \theta a_t a_{t-2} + \theta^2 a_{t-1} a_{t-2})] = \theta\sigma_a^2$$

MA(2)

$$X_t = \epsilon_t + a_1\epsilon_{t-1} + a_2\epsilon_{t-2}$$

Expectation:

$$E[X_t] = E[C_1] + a_1 E\left[\epsilon_{t,1}\right] + a_2 E[\epsilon_{t-2}]$$
$$E_t \sim N(0, \sigma^2) \text{ is } iid \text{ process}$$
$$E[G_t] = E[\epsilon_t,] = E[G_{t-2}] = 0$$
$$E[X_t] = 0$$

Variance:

$\mathrm{Var}(x_t) = E[x_k^2] - \left(E(x_t)\right)^2 = E[x_k^2] - 0^2 = E[x_k^2]$

$E[x_t^2] = E[\epsilon_2^2 + a_1^2\varepsilon_{t-1}^2 + a_2^2\epsilon_{t-2}^2 + 2a_1\epsilon_t\epsilon_{t-1} + a_2\epsilon_t\epsilon_{t-2} + a_1a_2\epsilon_t, \epsilon_{t-2}$
$+a_2\epsilon_t\epsilon_t - 2 \ + a_2 a\,\epsilon_{t-1}\epsilon_{t-2}]$

$E[x_k^2] = \sigma^2 + a_1^2\sigma^2 + a_2^2\sigma^2 = \sigma^2(1 + a_1^2 + a_2^2)$

$\mathrm{Var}\ (x_t) = E(x_t^2) = \sigma^2(1 + a_1^2 + a_2^2)$

Autocorrelation Function (ACF):

First Order Autocorrelation (p1):

$$S(1) = \frac{\mathrm{Cov}(x_t, x_{t-1})}{\mathrm{Var}(x_t)}$$

$$\mathrm{Cov}(x_t, x_{t-1}) \quad = E\left[\left(x_t - \mu_{x_t}\right)\left(x_{t-1} - \mu_{x_{t-1}}\right)\right]$$
$$= E[x_t x_{t-1}]\ as\ \mu\ is\ zero$$
$$= E[(a, e_{t-1} + a_2\epsilon_{t-2} + \epsilon_t)(E_{t-1} + a_1 c_{t-2} + a_2\epsilon_{2-3})]$$

$$\mathrm{Cov}(x_t, x_{t-1}) = a_1\sigma^2 + a_1a_2\sigma^2 = a_1\sigma^2(1 + a_2)$$

$$\rho(1) = \frac{a_1\sigma^2(1 + a_2)}{\sigma^2(1 + a_1^2 + a_2^2)} = \frac{a_1(1 + a_2)}{(1 + a_1^2 + a_2^2)}$$

3rd order Autocorrelation (p3):

$$S(3) = \frac{\mathrm{Cov}\ (x_t, x_{t-3})}{\mathrm{Var}(x_t)}$$

$$\mathrm{Cov}(x_t, x_{t-3}) = \ E[x_t x_{t-3}] = 0$$

$$\rho(3) = 0$$

MA(q)

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$$

Autoregressive polynomial

$$Y_t = \theta(B)W_t$$

Invertibility

We can invert MA(1) into AR(∞)

Inverting is basically act of expanding function in geometric series

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots \beta_q Z_{t-q}$$

MA(q) process is invertible if the roots of the polynomial

$$\beta(B) = \beta_0 + \beta_1 B + \cdots + \beta_q B^q$$

all lie outside the unit circle, where we regard B as a complex variable (not an operator).

Expectation

$$E[X_t] = 0$$

Variance

$$Var(X_t) = Var(Z_t) + Var(\theta Z_{t-1}) + \cdots + Var\big(\theta_2 Z_{t-q}\big)$$
$$= \sigma^2 + \sigma^2\theta_1^2 + \sigma^2\theta_2^2 + \cdots + \sigma^2\theta_k^2$$
$$= \sigma^2(1 + \sum_{i=0} \theta_i^2)$$

Autocorrelation Function (ACF)

$$\gamma_0 = \text{var}\,(X_t) = \big(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\big)\sigma_Z^2$$
$$\gamma_j = E\big[\big(Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}\big)\big(Z_{t-j} + \theta_1 Z_{t-j-1} + \cdots + \theta_q Z_{t-j-q}\big)\big]$$
$$\gamma_j = \begin{cases} \big(\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q\theta_{q-j}\big)\sigma^2 \text{ for } j \le q \\ 0 \text{ for } j > q \end{cases}$$

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q\theta_{q-j}}{\sum_{i=1}^{q}\theta_i^2}$$

Autocovariance Function (ACVF)

$$\gamma_k = \begin{cases} \sigma_a^2\big(-\theta_k + \theta_k\theta_1 + \cdots + \theta_{q-k}\theta_q\big) \;\; k = 1,2,\ldots,q \\ 0, |k| > q \end{cases}$$

Takeaway: An MA series is <u>always stationary</u>, but for an AR series to be stationary, all its characteristic roots must be less than 1 in modulus.

ARMA

ARMA(1, 1)	$(Y_t \cdot \mu) = \phi_1\big(Y_{(t-1)} - \mu\big) + \varepsilon_t + \theta_1\varepsilon_{(t-1)}$
ARMA(2, 1)	$(Y_t - \mu) = \phi_1\big(Y_{(t-1)} - \mu\big) + \phi_2\big(Y_{(t-2)} - \mu\big) + \varepsilon_t + \theta_1\varepsilon_{(t-1)}$
ARMA(1, 2)	$(Y_t - \mu) = \phi_1\big(Y_{(t-1)} - \mu\big) + \varepsilon_t + \theta_1\varepsilon_{(t-1)} + \theta_2\varepsilon_{(t-2)}$
ARMA(p, q)	

MA, AR & ARMA

	MA(q)	AR(p)	ARMA(p,q)
	Innovations prior to lag q have no effect on current observations	Current values are explained only by the past p observations (longer term dependence)	Combines the ideas of AR and MA models into a compact form to reduce the number of parameters
Order Selection	Use **ACF** plot	Use **PACF** plot	Use AIC, BIC, EACF
PACF	Decays to zero	Goes to zero after lag p	Decays to zero
ACF	Goes to zero after lag p	Decays to zero	Decays to zero
Stationarity	Always stationary	Only when characteristic roots < 1 in absolute value or eigenvalues	Only when characteristic roots of AR component < 1 in absolute value

The importance of stationarity

- Stationarity is crucial for modeling
- Non-stationary series are impossible to predict for example the random walk
- Stationary series are bounded by their overall variance
- AR(1) is easy to see stationary if phi sub 1 is less than 1 in absolute value

Covariance Stationarity

- Across all timestamps

Definition

A sequence of random variables is covariance stationary if and only if

$$\exists \mu \in \mathbb{R}: E[X_n] = \mu, \forall n > 0$$
$$\forall j \geq 0, \exists \gamma_j \in \mathbb{R}: Cov[X_n, X_{n-j}] = \gamma_j, \forall n > j$$

In other words:

1. All the terms of the sequence have mean mu
2. The covariance $Cov[X_n, X_{n-j}]$ depends only on the relative position i and not on the absolute position n.

Note that

$$Var[X_n] = Cov[X_n, X_n]$$

Implies that weakly stationary process has constant variance

Covariance stationary sequences are also called:

- Weakly stationary sequences
- Covariance stationary processes
- Weakly stationary processes

Unit Root
The problem of Unit Roots

- Stationarity is crucial for modeling
- Indicate explosive unpredictable behavior
- Can be indication something is missing, like accounting for inflation
- Unit roots are an important type of non-stationarity
- Decomposition int AR(1) processes has a parameter >= 1
- AR Root near 1 suggest data should be differenced

Shock
- Thus, the coefficients do not converge to zero the effect of any shock is theoretically permanent

Naïve Method
- Forecast is equal to the last observed value

Drift
- Forecast equal to last value **plus average change**

The Dickey-Fuller Unit Root Test [the one used in class]
- This test provides a statistical test for first differencing
- The null hypothesis is that first differencing is required
- Alternative hypothesis: the series is stationary
- No closed form for the test statistic, nor for the statistics distribution!
- The test has three forms: Zero Mean, Single Mean, Trend and variations for differences of higher order for seasonal differencing.
- Test for checking if we are not over-differencing.

τ & δ is zero:

$$\nabla Y_t = \phi Y_{t-1} + X_t$$
$$\nabla Y_t = (\rho - 1)Y_{t-1} + W_t$$

Or:

$$Y_t = \tau + \delta t + \rho Y_{t-1} + W_t$$
$$\nabla Y_t = \tau + \delta t + (\rho - 1)Y_{t-1} + W_t$$

F-test	(τ, δ, ρ)	$\lvert\rho\rvert < 1$	$\lvert\rho\rvert = 1$
	$(\tau \neq 0)$ $\delta \neq 0$	stationary around a linear trend	integrated and exhibits a quadratic trend
Yes	$\tau \neq 0,$ $\delta = 0$	stationary with a nonzero mean	integrated and exhibits a linear trend
Yes	$\tau = 0$ $\delta = 0$	stationary with zero mean	integrated without deterministic trend

$$P(\emptyset = 0.033|\theta = 0) < 0.05 => reject$$

Did we go through this?

Dickey-Fuller Zero Mean Test [used]:

- Forecast for an AR(1) model reverts back to the mean

Dickey-Fuller Single Mean Test [unused]:

- Rarely reports a difference with the zero-mean case
- Means in a stationary series just change the level, not stationarity.
- Adding a constant to a random walk still isn't stationary after the constant is removed

Dickey-Fuller Trend Test [unused]:

- This one will distinguish between linear time trend with stationary variance and non-stationary which either has a time trend or mean

Information Criteria

- Minimize the criteria
- Calculates the tradeoff of parameters needed
- Minimize the fit for the smallest number of parameters. Less risk of overfitting
- Take the test
    - ARIMA 1.2 vs ARIMA 2.1
    - ARIMA 1.1 vs AR 2
    - ARMA 1.1 vs AR 3
- AIC vs BIC
    - AIC higher chance of overfitting

Transformations

- Box Cox transformation
- Differencing

Smoothing

- Moving Average

Extract a Trend

- Hodrick-Prescott Filter [nope]
- Separates a time-series yt into a trend component τt and a cyclical component ct.
- t $y_t = \tau_t + c_t + \epsilon_t$.

$$\min_{\tau} \left(\sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right).$$

- First term: penalizes the cyclic component
- Second term: penalizes variations in the growth rate

State Space [SS]

- Depends on state in time
- $y_t = a_t + \epsilon_t, \epsilon_t \sim i.i.d\ N(0, W)$
- $a_{t+1} = a_t + n_t, n \sim i.i.d\ N(0, W_n)$
- 1. Measurement Equation: Relationship between the observed values $(Y_1, \ldots)$
- 2. State Equation: Dynamic of the unobserved state variables $((a_1, \ldots)$
- We want to infer properties of a_t from the observed Y_s
- One technique is Kalman Filter
- State Space Models: AIC, BIC

State Space Models

A *state space model* consists of a *state equation*,

$$\mathbf{Y}_t = \mathbf{\Phi}_t \mathbf{Y}_{t-1} + \boldsymbol{\nu}_t + \mathbf{a}_t,$$

and an *observation equation*,

$$z_t = \mathbf{H}_t \mathbf{Y}_t + \boldsymbol{\mu}_t + \mathbf{b}_t,$$

where $\mathbf{Y}_t$ is a state vector with a transition matrix $\mathbf{\Phi}_t$, $\mathbf{a}_t$ are independent shocks with covariance matrices $\mathbf{A}_t$, $\mathbf{H}_t$ is the observation matrix, and $\mathbf{b}_t$ are another set of shocks with covariance matrices $\mathbf{B}_t$, which are independent of $\mathbf{a}_t$. The arrays $\mathbf{\Phi}_t$, $\boldsymbol{\nu}_t$, $\mathbf{H}_t$, and $\boldsymbol{\mu}_t$ are deterministic, and often independent of t.

Through the construction of the state vector $\mathbf{Y}_t$, the AR(1) state equation is capable of representing higher order structures. The state space representation of a time series model is not unique.

ARIMA(p,d,q) in State Space Form

Consider an ARIMA(p,d,q) process in the generalized ARMA form $\varphi(B)z_t = \theta(B)a_t$. Let $m = \max(p+d, q+1)$, one has

$$z_t = \varphi_1 z_{t-1} + \cdots + \varphi_m z_{t-m} + a_t - \theta_1 a_{t-1} - \cdots - \theta_{m-1} a_{t-m+1}.$$

Let $y_t^{(m)} = \varphi_m z_{t-1} - \theta_{m-1} a_t$, and $y_t^{(j)} = \varphi_j z_{t-1} + y_{t-1}^{(j+1)} - \theta_{j-1} a_t$, $j < m$ ($\theta_0 = -1$). The state vector $\mathbf{Y}_t = (y_t^{(1)}, \ldots, y_t^{(m)})^T$ satisfies

$$\mathbf{Y}_t = \begin{pmatrix} \varphi_1 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{m-1} & 0 & \cdots & 1 \\ \varphi_m & 0 & \cdots & 0 \end{pmatrix} \mathbf{Y}_{t-1} + \begin{pmatrix} 1 \\ -\theta_1 \\ \vdots \\ -\theta_{m-1} \end{pmatrix} a_t.$$

It is easy to check that $z_t = y_t^{(1)}$, so the observation equation is simply $z_t = (1, 0, \ldots, 0)\mathbf{Y}_t$.

Another State Space Form for ARIMA

Recall the complementary function

$$C_t(l) = z_{t+l} - \textstyle\sum_{j=0}^{l-1} \psi_j a_{t+l-j} = C_{t-1}(l+1) + \psi_l a_t.$$

Set $\mathbf{Y}_t = (C_t(0), \ldots, C_t(m-1))^T$, one has

$$\mathbf{Y}_t = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \varphi_m & \varphi_{m-1} & \cdots & \varphi_1 \end{pmatrix} \mathbf{Y}_{t-1} + \begin{pmatrix} 1 \\ \psi_1 \\ \vdots \\ \psi_{m-1} \end{pmatrix} a_t,$$

where the last equation follows from $\varphi(B)C_{t-1}(m) = 0$. The observation equation is again $z_t = (1, 0, \ldots, 0)\mathbf{Y}_t$, as $C_t(0) = z_t$.

Remember that the ψ weights are determined from φ_j, θ_j via

$$\psi_j = \varphi_1 \psi_{j-1} + \cdots + \varphi_m \psi_{j-m} - \theta_j, \quad j > 0,$$

Example: ARIMA(1,1,1)

Consider $(1 - \phi B)(1 - B)z_t = (1 - \theta B)a_t$ with $\varphi_1 = 1 + \phi$, $\varphi_2 = -\phi$, $\theta_1 = \theta$, and $\psi_1 = 1 + \phi - \theta$.

With the first representation, $Y_t^{(2)} = -\phi z_{t-1} - \theta a_t$,

$$\mathbf{Y}_t = \begin{pmatrix} 1+\phi & 1 \\ -\phi & 0 \end{pmatrix} \mathbf{Y}_{t-1} + \begin{pmatrix} 1 \\ -\theta \end{pmatrix} a_t.$$

With the second representation, $Y_t^{(2)} = (1+\phi)z_t - \phi z_{t-1} - \theta a_t$,

$$\tilde{\mathbf{Y}}_t = \begin{pmatrix} 0 & 1 \\ -\phi & 1+\phi \end{pmatrix} \tilde{\mathbf{Y}}_{t-1} + \begin{pmatrix} 1 \\ 1+\phi-\theta \end{pmatrix} a_t.$$

It is seen that $\tilde{\mathbf{Y}}_t = \left(\begin{smallmatrix} 1 & 0 \\ 1+\phi & 1 \end{smallmatrix}\right)\mathbf{Y}_t$, as $\left(\begin{smallmatrix} 1 & 0 \\ 1+\phi & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 \\ -\theta \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 \\ 1+\phi-\theta \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 0 \\ 1+\phi & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1+\phi & 1 \\ -\phi & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 1 \\ -\phi & 1+\phi \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & 0 \\ 1+\phi & 1 \end{smallmatrix}\right)$.

Kalman Filter: Derivation

Consider Gaussian process with initial state $\mathbf{Y}_0 \sim N(\mathbf{y}_0, \mathbf{V}_0)$. The state at time 1, $\mathbf{Y}_1 = \mathbf{\Phi}_1 \mathbf{Y}_0 + \boldsymbol{\nu}_1 + \mathbf{a}_1$, has mean and covariance

$$\mathbf{y}_{1|0} = \mathbf{\Phi}_1 \mathbf{y}_0 + \boldsymbol{\nu}_1, \quad \mathbf{V}_{1|0} = \mathbf{\Phi}_1 \mathbf{V}_0 \mathbf{\Phi}_1^T + \mathbf{A}_1.$$

The joint distribution of $(\mathbf{Y}_1^T, \mathbf{z}_1^T)^T$ has mean and covariance

$$\begin{pmatrix} \mathbf{y}_{1|0} \\ \mathbf{H}_1 \mathbf{y}_{1|0} + \boldsymbol{\mu}_1 \end{pmatrix}, \quad \begin{pmatrix} \mathbf{V}_{1|0} & \mathbf{V}_{1|0}\mathbf{H}_1^T \\ \mathbf{H}_1 \mathbf{V}_{1|0} & \mathbf{H}_1 \mathbf{V}_{1|0} \mathbf{H}_1^T + \mathbf{B}_1 \end{pmatrix}.$$

The conditional distribution of $\mathbf{Y}_1|\mathbf{z}_1$ thus has the mean

$$\mathbf{y}_1 = \mathbf{y}_{1|0} + \mathbf{V}_{1|0}\mathbf{H}_1^T(\mathbf{H}_1\mathbf{V}_{1|0}\mathbf{H}_1^T + \mathbf{B}_1)^{-1}(\mathbf{z}_1 - \mathbf{H}_1\mathbf{y}_{1|0} - \boldsymbol{\mu}_1),$$

and the covariance

$$\mathbf{V}_1 = \mathbf{V}_{1|0} - \mathbf{V}_{1|0}\mathbf{H}_1^T(\mathbf{H}_1\mathbf{V}_{1|0}\mathbf{H}_1^T + \mathbf{B}_1)^{-1}\mathbf{H}_1\mathbf{V}_{1|0}.$$

Replacing 1 by t and 0 by $t-1$, one obtains the Kalman filter.

Kalman Filter: Prediction and Updating

At time $t-1$, the *prediction equations*

$$\mathbf{y}_{t|t-1} = \mathbf{\Phi}_t \mathbf{y}_{t-1} + \boldsymbol{\nu}_t, \quad \mathbf{V}_{t|t-1} = \mathbf{\Phi}_t \mathbf{V}_{t-1}\mathbf{\Phi}_t^T + \mathbf{A}_t,$$

give the optimal estimator of $\mathbf{Y}_t$ and its error covariance. For the prediction $\hat{\mathbf{z}}_{t|t-1} = \mathbf{H}_t \mathbf{y}_{t|t-1} + \boldsymbol{\mu}_t$ of $\mathbf{z}_t$, one has the *innovation* $\mathbf{e}_t = \mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}$ with the covariance $\mathbf{\Sigma}_t = \mathbf{H}_t \mathbf{V}_{t|t-1}\mathbf{H}_t^T + \mathbf{B}_t$.

Once $\mathbf{z}_t$ becomes available, the estimator of $\mathbf{Y}_t$ is updated through

$$\mathbf{y}_t = \mathbf{y}_{t|t-1} + \mathbf{V}_{t|t-1}\mathbf{H}_t^T \mathbf{\Sigma}_t^{-1}\mathbf{e}_t,$$

which has a smaller covariance

$$\mathbf{V}_t = \mathbf{V}_{t|t-1} - \mathbf{V}_{t|t-1}\mathbf{H}_t^T\mathbf{\Sigma}_t^{-1}\mathbf{H}_t\mathbf{V}_{t|t-1}.$$

The matrix $K_t = \mathbf{V}_{t|t-1}\mathbf{H}_t^T\mathbf{\Sigma}_t^{-1}$ is the Kalman gain matrix.

Example: ARMA(1,1)

A state space representation of an ARMA(1,1) model is given by

$$\mathbf{Y}_t = \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix} \mathbf{Y}_{t-1} + \begin{pmatrix} 1 \\ -\theta \end{pmatrix} a_t,$$

where $Y_t^{(1)} = z_t$ and $Y_t^{(2)} = -\theta a_t$. Set $\mathbf{y}_0 = \mathbf{0}$ and

$$\mathbf{V}_0 = \sigma_a^2 \begin{pmatrix} (1+\theta^2-2\phi\theta)/(1-\phi^2) & -\theta \\ -\theta & \theta^2 \end{pmatrix} = \sigma_a^2 \begin{pmatrix} 1+v_0 & -\theta \\ -\theta & \theta^2 \end{pmatrix},$$

where $v_0 = (\phi-\theta)^2/(1-\phi^2)$. The updating equations give

$$\mathbf{y}_{1|0} = \mathbf{0}, \quad \mathbf{V}_{1|0} = \mathbf{\Phi}\mathbf{V}_0\mathbf{\Phi}^T + \mathbf{A} = \mathbf{V}_0.$$

The innovation is $e_1 = z_1$ and the Kalman gain matrix is $K_1 = \mathbf{V}_{1|0}(1/\sigma_a^2(1+v_0), 0)^T = \left(\begin{smallmatrix} 1 \\ -\theta/(1+v_0) \end{smallmatrix}\right)$, so $\mathbf{y}_1 = \left(\begin{smallmatrix} z_1 \\ -\theta z_1/(1+v_0) \end{smallmatrix}\right)$ and $\mathbf{V}_1 = (I - K_1(1,0))\mathbf{V}_{1|0} = \sigma_a^2 v_1 \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, where $v_1 = \theta^2 v_0/(1+v_0)$. Note that $a_1|z_1$ is *not* degenerate.

Example: ARMA(1,1)

At time $t-1>0$, let $V_{t-1} = \sigma_a^2 v_{t-1}\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. The updating equations give $y_{t|t-1}^{(1)} = \phi z_{t-1} - \theta\tilde{a}_{t-1}$, where $\tilde{a}_{t-1} = E[a_{t-1}|z_{t-1}, \ldots, z_1]$, $y_{t|t-1}^{(2)} = 0$, and

$$\mathbf{V}_{t|t-1} = \mathbf{\Phi}\mathbf{V}_{t-1}\mathbf{\Phi}^T + \mathbf{A} = \sigma_a^2 \begin{pmatrix} 1+v_{t-1} & -\theta \\ -\theta & \theta^2 \end{pmatrix}$$

The innovation $e_t = z_t - \phi z_{t-1} + \theta\tilde{a}_{t-1} = a_t - \theta(a_{t-1} - \tilde{a}_{t-1})$ has variance $\sigma_a^2(1+v_{t-1})$ and the Kalman gain matrix is given by $K_t = \mathbf{V}_{t|t-1}(1/\sigma_a^2(1+v_{t-1}), 0)^T = \left(\begin{smallmatrix} 1 \\ -\theta/(1+v_{t-1}) \end{smallmatrix}\right)$. One has $\mathbf{y}_t = \left(\begin{smallmatrix} z_t \\ -\theta e_t/(1+v_{t-1}) \end{smallmatrix}\right)$ and $\mathbf{V}_t = (I - K_t(1,0))\mathbf{V}_{t|t-1} = \sigma_a^2 v_t \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, where $v_t = \theta^2 v_{t-1}/(1+v_{t-1})$. Note that for $|\theta|<1$, $v_t \to 0$ at an exponential rate.

Kalman Filter: Multiple Steps Ahead

To predict more than one step ahead based on information at time T, one simply bypass the updating step,

$$\mathbf{y}_{T+l|T} = \mathbf{\Phi}_{T+l}\mathbf{y}_{T+l-1|T} + \boldsymbol{\nu}_{T+l}, \quad l = 1, 2, \ldots,$$

where $\mathbf{y}_{T|T} = \mathbf{y}_T$. The covariance of the prediction error is given by

$$\mathbf{V}_{T+l|T} = \mathbf{\Phi}_{T+l}\mathbf{V}_{T+l-1|T}\mathbf{\Phi}_{T+l}^T + \mathbf{A}_{T+l}, \quad l = 1, 2, \ldots,$$

where $\mathbf{V}_{T|T} = \mathbf{V}_T$. The predictor of $\mathbf{z}_{T+l}$ is

$$\hat{\mathbf{z}}_{T+l|T} = \mathbf{H}_{T+l}\mathbf{y}_{T+l|T} + \boldsymbol{\mu}_{T+l},$$

with error covariance $\mathbf{E}_{T+l|T} = \mathbf{H}_{T+l}\mathbf{V}_{T+l|T}\mathbf{H}_{T+l}^T + \mathbf{B}_{T+l}$.

For ARMA(1,1) with $T>0$, $\mathbf{y}_{T+l|T} = \mathbf{\Phi}^l \mathbf{y}_T = \phi^{l-1}\mathbf{y}_{T+1|T}$, and $\mathbf{V}_{T+l|T} = \sigma_a^2 v_{T+l|T}\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) + \mathbf{A}$, where

$$v_{T+1|T} = v_T, \quad v_{T+l|T} = \phi^2 v_{T+l-1|T} + (\phi-\theta)^2 \to v_0.$$

Kalman Filter: Maximum Likelihood

Recall that the joint likelihood of $\mathbf{z}_1, \ldots, \mathbf{z}_N$ can be factored as $L(Z_N) = \prod_{t=1}^N p(\mathbf{z}_t|Z_{t-1})$, where $Z_{t-1} = \{\mathbf{z}_1, \ldots, \mathbf{z}_{t-1}\}$. For Gaussian processes, $p(\mathbf{z}_t|Z_{t-1})$ is normal with mean $\hat{\mathbf{z}}_{t|t-1}$ and covariance $\mathbf{\Sigma}_t$. For $\mathbf{z}_t$ univariate, drop boldface and write ς_t for $\mathbf{\Sigma}_t$, one has the *prediction error decomposition* form of the likelihood,

$$\log L(Z_N) = -\tfrac{N}{2}\log 2\pi - \tfrac{1}{2}\textstyle\sum_{t=1}^N \log\varsigma_t - \tfrac{1}{2}\sum_{t=1}^N \frac{e_t^2}{\varsigma_t}.$$

As a function of model parameters, $\log L(Z_N)$ can be maximized using optimization tools to yield the MLE of parameters.

For ARMA(1,1), one has $\varsigma_t = \sigma_a^2(1+v_{t-1})$. Dropping the constant,

$$\log L(Z_N) = -\tfrac{1}{2}\textstyle\sum_{t=1}^N \log(1+v_{t-1}) - \tfrac{N}{2}\log\sigma_a^2 - \tfrac{1}{2\sigma_a^2}\sum_{t=1}^N \frac{e_t^2}{1+v_{t-1}}.$$

One can "profile" out σ_a^2, then work on the profile likelihood.

State Space Representation of AR(2)

- From AR(2) find the state space model
- Literally just sets up the model from

State var (unknown)

input(exog)

$$X_t = \Phi X_{t-1} + \Gamma U_t + B W_t, \quad \text{— input State eq-n}$$

$$Y_t = A X_t + \Lambda U_t + V_t, \quad \text{— Output Observation eq-n}$$

- Then goes backward afterwards to show its correct

AR(2)

1) $y_{t+1} = \theta_1 y_t + \theta_2 y_{t-1} + \varepsilon_t$ — SS?

$$\underbrace{\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix}}_{X_t} = \underbrace{\begin{pmatrix} \theta_1 & 1 \\ \theta_2 & 0 \end{pmatrix}}_{\Phi} \underbrace{\begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix}}_{X_{t-1}} + \underbrace{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{B} \underbrace{\varepsilon_t}_{\varepsilon_t}$$

$$\underbrace{y_t}_{y_t} = \underbrace{\begin{pmatrix} 1 & 0 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix}}_{X_t} + \underbrace{0}_{V_t}$$

$$\begin{cases} x_{1,t} = \theta_1 x_{1,t-1} + x_{2,t-1} + \varepsilon_t \\ x_{2,t} = \theta_2 x_{1,t-1} + 0 \\ x_{1,t} = \theta_1 x_{1,t-1} + \theta_2 x_{1,t-2} + \varepsilon_t \\ y_t = x_{1,t} \end{cases} \Rightarrow \underline{y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t}$$

$$\Phi = \begin{pmatrix} \theta_1 & 1 \\ \theta_2 & 0 \end{pmatrix} \qquad \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} y_t \\ \theta_2 y_{t-1} \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$A = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

State Space Representation of MA(2)

2) MA(1) $y_t = \varepsilon_t + \theta \varepsilon_{t-1}$

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}_{\Phi} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \underbrace{\begin{pmatrix} 1 \\ \theta \end{pmatrix}}_{\Gamma} \underline{\underline{\varepsilon_t}}$$

$$y_t = \underbrace{\begin{pmatrix} 1 & 0 \end{pmatrix}}_{A} \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} \qquad \Phi = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$$x_{1,t} = 0 + x_{2,t-1} + \varepsilon_t \qquad \Gamma = \begin{pmatrix} 1 \\ \theta \end{pmatrix}$$

$$x_{2,t} = 0 + 0 + \theta \varepsilon_t \qquad A = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

$$x_{1,t} = \theta \varepsilon_{t-1} + \varepsilon_t \qquad \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} y_t \\ \theta \varepsilon_t \end{pmatrix}$$

$$y_t = x_{1,t} = \varepsilon_t + \theta \varepsilon_{t-1}$$

State Space Representation of ARIMA(p,q) [BOOK]

Consider that a time series $\{y_t\}$ is generated from an ARIMA(1,1,1) model, so that

$$y_t - y_{t-1} = \alpha(y_{t-1} - y_{t-2}) + \epsilon_t + \gamma\epsilon_{t-1},$$

where α is the AR parameter, γ is the MA parameter and $\{\epsilon_t\}$ is a Gaussian white noise sequence with variance equal to 1.

Define the state vector

$$\beta_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \epsilon_t \end{bmatrix}.$$

Write down a state space representation for y_t, i.e. express y_t as a state space model:

$$y_t = x^\top \beta_t + \delta_t,$$
$$\beta_t = F\beta_{t-1} + \zeta_t.$$

In your answer you should:

(a) specify the components x, F, δ_t and ζ_t;

(b) write down the distributions of δ_t and ζ_t.

First write $$y_t = (\alpha - 1)y_{t-1} - \alpha y_{t-2} + \epsilon_t + \gamma\epsilon_{t-1}$$

With the definition of β_t we have

$$\beta_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \epsilon_t \end{bmatrix} = \begin{bmatrix} \alpha - 1 & -\alpha & \gamma \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \epsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ \epsilon_t \end{bmatrix}$$

$$\beta_t = F\beta_{t-1} + \zeta_t$$

This is the transition model.

The observation model is

$$y_t = [1, 0, 0] \begin{bmatrix} y_t \\ y_{t-1} \\ \epsilon_t \end{bmatrix} = x^\top \beta_t$$

where $\delta_t = 0$ with probability one.

The distributions of δ_t and ζ_t are:

$$\zeta_t = \begin{bmatrix} \epsilon_t \\ 0 \\ \epsilon_t \end{bmatrix} \sim N\left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \right\} = N(0, Z)$$

and δ_t does not have a distribution, it is equal to 0 with probability one.

State Space Representation of ARIMA(p,q)

The question:
Write the SS representation of ARMA (p,q) (ARMA to SS)
The asnwer:
1. find the max(p,q+1) = number of state space variables
2. write the SS representation (wirite state equation and observation equation) using the matrices from theory
3. Prove:
    3.1 Write the state equation as system of equations
    3.2 Make 1 equation for X1,t out of your d equations (bottom to top: from xd,t to xd-1,t; ... from x3,t to x2,t; from x2,t to x1,t)
    3.3 plug x1,t into yt
    3.4 come to original ARMA form
4. write all the matrcies as the answer (matrcies A,B,C + X,t)

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t + \rho_1 \varepsilon_{t-1} + \dots + \rho_q \varepsilon_{t-q}$$

1) $\max(p; q+1) = d$

2) $$\begin{pmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{d,t} \end{pmatrix} = \underbrace{\begin{pmatrix} \theta_1 & 1 & 0 & \cdots & 0 \\ \theta_2 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ \theta_{d-1} & 0 & 0 & \cdots & 1 \\ \theta_d & 0 & 0 & \cdots & 0 \end{pmatrix}}_{A} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \\ \vdots \\ x_{d,t-1} \end{pmatrix} + \underbrace{\begin{pmatrix} 1 \\ \rho_1 \\ \vdots \\ \rho_d \end{pmatrix}}_{B} \varepsilon_t$$

size $= d$

$$y_t = \underbrace{\begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}}_{C} X_t$$

3)

State Space Representation of ARIMA(2,1)

3) ARMA(2,1): $y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t + \rho \varepsilon_{t-1}$

$p = 2$

$q+1 = 1+1 = 2 \Rightarrow \max = 2$

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} \theta_1 & 1 \\ \theta_2 & 0 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ \rho \end{pmatrix} \varepsilon_t$$

$$y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix}$$

– SS form

$$x_{1,t} = \theta_1 x_{1,t-1} + x_{2,t-1} + \varepsilon_t$$

$$x_{2,t} = \theta_2 x_{1,t-1} + 0 + \rho \varepsilon_t$$

$$x_{1,t} = \theta_1 x_{1,t-1} + \theta_2 x_{1,t-2} + \rho \varepsilon_{t-1} + \varepsilon_t$$

$$y_t = x_{1,t} = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t + \rho \varepsilon_{t-1}$$

$\Rightarrow$

$$\Phi = \begin{pmatrix} \theta_1 & 1 \\ \theta_2 & 0 \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} 1 \\ \rho \end{pmatrix}$$

$$A = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

$$X_t = \begin{pmatrix} y_t \\ \theta_2 y_{t-1} + \rho \varepsilon_t \end{pmatrix}$$

$$X_t = \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} y_t \\ \theta_2 x_{1,t-1} + \rho \varepsilon_t \end{pmatrix} = \begin{pmatrix} y_t \\ \theta_2 y_{t-1} + \rho \varepsilon_t \end{pmatrix}$$

State Space Representation of ARIMA(1, 2) with drift

4) ARIMA(1,2) with drift

$$y_t = c + \theta y_{t-1} + \varepsilon_t + \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2}$$

$$\max(1, 2+1) = 3$$

I

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \end{pmatrix} = \underbrace{\begin{pmatrix} \theta & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}}_{A} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \\ X_{3,t-1} \end{pmatrix} + \underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}_{B} c + \underbrace{\begin{pmatrix} 1 \\ \rho_1 \\ \rho_2 \end{pmatrix}}_{C} \varepsilon_t$$

zeros

AR coef

$$y_t = \underbrace{\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}}_{D} \begin{pmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \end{pmatrix}$$

$$\begin{cases} X_{1,t} = \theta X_{1,t-1} + X_{2,t-1} + 0 + c + \varepsilon_t \\ X_{2,t} = X_{3,t-1} + 0 + \rho_1 \varepsilon_t \\ X_{3,t} = 0 + 0 + \rho_2 \varepsilon_t \end{cases}$$

1. *Inserting X3,t in X2,t.*
2. *Inserting the X2,t in X1t*
3. *Reconstructing the matrices for proof*

$$X_{2,t} = \rho_2 \varepsilon_{t-1} + \rho_1 \varepsilon_t$$

$$X_{1,t} = \theta X_{1,t-1} + \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + 0 + c + \varepsilon_t$$

$$y_t = X_{1,t} = c + \theta y_{t-1} + \varepsilon_t + \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2}$$

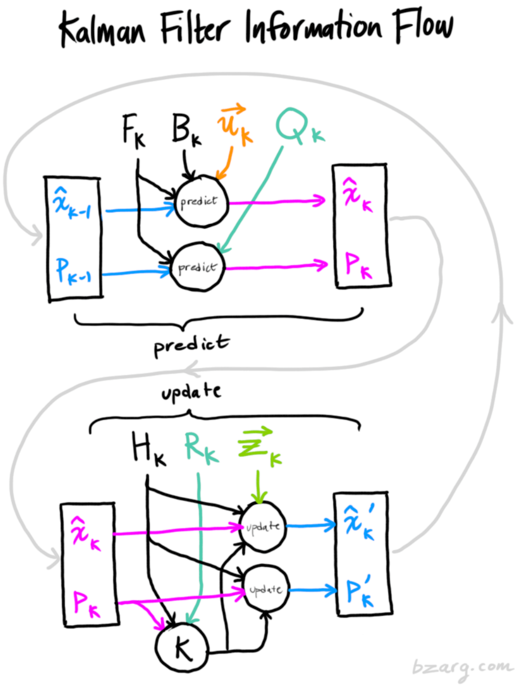
$$A = \begin{pmatrix} \theta & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad C = \begin{pmatrix} 1 \\ \rho_1 \\ \rho_2 \end{pmatrix} \quad D = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$$

$$X_t = \begin{pmatrix} y_t \\ \rho_1 \varepsilon_t + \rho_2 \varepsilon_{t-1} \\ \rho_2 \varepsilon_t \end{pmatrix} = \begin{pmatrix} y_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix}$$

Kalman Filter?????

- With Kalman filters, we can **mitigate the uncertainty by combining the information we do have with a distribution that we feel more confident in**.
- Forward Pass -> Find the filtered state. Recursive.
- Backward Pass -> Output of the Kalman filter.
- Relates the current state to the current observation

- Nowcasting, using current values to correct the “error”.



Vector autoregression [VAR]

- Statistical model to capture the relationship between multiple quantities as they change over time.
- Each variable has an equation modeling its evolution time.
- 5th order VAR: VAR(5): linear combination of the last 5 periods.

Granger Causality

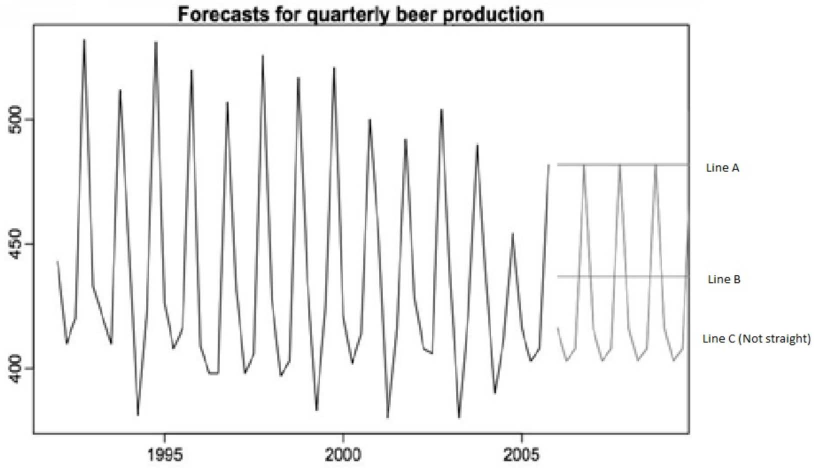
- Way of testing if one variable is useful at forecasting another.
- Concept: Cause cannot come after effect
- Feedback system: If it helps goes both ways

Box-Jenkins VS Vector Autoregression (VAR)

- The difference between box-jenkins and vector autoregression approach to economic forecasting: box-jenkins approach is used for a univariate time series data where vector autoregression approach is used for multivariate time series data. The variables used in a standard vector autoregression are stationary, but the variable in a box-jenkins process may have different stationary point. box-jenkins processes provide better forecasts of the variable in the model, where vector autoregression forecasts models sometime “over-specified” and don't forecast very well.

Exercises

Forecasting Method



Line A is which simple forecasting method?
Line B is which simple forecasting method?
Line C is which simple forecasting method?

Options:

- average method
- naive method
- seasonal naive method
- drift method

ANSWER:
Line A --Naive Method
Line B --Average method
Line C--Seasonal naive method
Naive method just uses the last periods actual to predict the current forecast, this can be seen in the line A that the last period levels are only used and the predictions are given accordingly.
Average method calculaes the overall moving average of the data and gives the forecast accordingly, as it can be seen that line B is nothing but the average of the overall series.
Seasonal naive method just forecast the value to be equal to the last observed value from the same season, so it can be seen that a proper seasonality is observed in data and hence the forecasts are following seasonal pattern.

Drift Method

Which simple forecasting method says the forecast is equal to the last value plus the average change?

- ○ None of these
- ○ Naive Method
- ○ Drift Method
- ○ Average Method

Which simple forecasting method says the forcast is equal to the last value plus average change?

Ans: ◉ Drift Method

Explanation:

Drift method

- Forcasts equal to last value plus average change

i.e $\hat{y}_{T+h/T} = y_T + \frac{h}{T-1}(y_T - y_1)$

This is simple forcasting method.