Time Series Analysis Exam 1 Studying

Part 1

1) Stationary Processes

1a) Give a definition of a stationary process and explain how it helps in forecasting.

$\mu_t = E(X_t), \sigma_t^2 = Var(X_t), \gamma(t, t+h) = Cov(X_t, X_{t+h})$ are all independent of t.

A stationary process means that is has no long term trends. This helps in forecasting because on a normally yearly scale the process looks overall similar year to year.

1b) What is White Noise? Is is stationary? What is its auto-covariance function?

White noise is a function where $\mu_t = 0, \sigma_t^2 = \sigma^2, \gamma(t, t+h) = 0$ It is stationary because it is not correlated with time.

2-4) Show that the MA(x) process given is stationary. Also find its mean and autocorrelation function.

2) $X_t = Z_t + \theta Z_{t-1}$

$\mu_t^2 = E(X) = E(Z_t + \theta Z_{t-1}) = E(Z_t) + \theta E(Z_{t-1}) = 0$

$\sigma_t^2 = var(X_t) = var(Z_t + \theta Z_{t-1}) = var(Z_t) + 2\theta cov(Z_t, Z_{t-1}) + \theta var(Z_{t-1}) = \sigma^2 + \theta^2\sigma^2 = (1-\theta^2)\sigma^2$

$cov(X_t, X_{t+h}) = cov(Z_t + \theta Z_{t-1}, Z_{t+h} + \theta Z_{t+h-1})$

$= cov(Z_t, Z_{t+h}) + \theta cov(Z_t, Z_{t+h-1}) + \theta cov(Z_{t-1}, Z_{t+h}) + \theta^2 cov(Z_{t-1}, Z_{t+h-1})$

- $h = 0 : 1$
- $h = 1 : 0 + \theta\sigma^2 + 0 + 0 = \theta\sigma^2$
- $h >= 2 : 0 + 0 + 0 + 0 = 0$

Mean: $\mu = 0$

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

- $h = 0 : 1$
- $h = 1 : \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{1+\theta^2}$
- $h = 2 : 0$

3) $X_t = Z_t + \theta Z_{t-2}$

$\mu_t^2 = E(X) = E(Z_t + \theta Z_{t-2}) = E(Z_t) + \theta E(Z_{t-2}) = 0$

$\sigma_t^2 = var(X_t) = var(Z_t + \theta Z_{t-2}) = var(Z_t) + 2\theta cov(Z_t, Z_{t-2}) + \theta var(Z_{t-2}) = \sigma^2 + \theta^2\sigma^2 = (1-\theta^2)\sigma^2$

$cov(X_t, X_{t+h}) = cov(Z_t + \theta Z_{t-2}, Z_{t+h} + \theta Z_{t+h-2})$

$= cov(Z_t, Z_{t+h}) + \theta cov(Z_t, Z_{t+h-2}) + \theta cov(Z_{t-2}, Z_{t+h}) + \theta^2 cov(Z_{t-2}, Z_{t+h-2})$

- $h = 0 : 1$

- $h = 1 : 0 + 0 + 0 + 0 = 0$
- $h = 2 : 0 + \theta\sigma^2 + 0 + 0 = \theta\sigma^2$
- $h >= 3 : 0 + 0 + 0 + 0 = 0$

Mean: $\mu = 0$

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

- $h = 0 : 1$
- $h = 1 : 0$
- $h = 2 : \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{1+\theta^2}$
- $h = 3 : 0$

4) $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$

$\mu_t^2 = E(X) = E(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}) = E(Z_t) + \theta_1 E(Z_{t-1}) + \theta_2 E(Z_{t-2}) = 0$

$\sigma_t^2 = var(X_t) = var(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2})$

$= var(Z_t) + 2\theta_1 cov(Z_t, Z_{t-1}) + 2\theta_2 cov(Z_t, Z_{t-2}) + \theta_1^2 var(Z_{t-1}) + 2\theta_1\theta_2 cov(Z_{t-1}, Z_{t-2})\theta_2 var(Z_{t-2})$

$= \sigma^2 + 0 + 0 + \theta_1^2\sigma^2 + \theta_2^2\sigma^2 = \sigma^2 + \theta_1^2\sigma^2 + \theta_2^2\sigma^2$

$cov(X_t, X_{t+h}) = cov(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}, Z_{t+h} + \theta_1 Z_{t+h-1} + \theta_2 Z_{t+h-2})$

$= cov(Z_t, Z_{t+h}) + \theta_1 cov(Z_t, Z_{t+h-1}) + \theta_2 cov(Z_t, Z_{t+h-2}) + \theta_1 cov(Z_{t-1}, Z_{t+h})$

$+\theta_1^2 cov(Z_{t-1}, Z_{t+h-1}) + \theta_1\theta_2 cov(Z_{t-1}, Z_{t+h-2}) + \theta_2 cov(Z_{t-2}, Z_{t+h}) + \theta_1\theta_2 cov(Z_{t-2}, Z_{t+h-1}) + \theta_2^2 cov(Z_{t-2}, Z_{t+h-2})$

- $h = 0 : 1$
- $h = 1 : 0 + \theta_1\sigma^2 + 0 + 0 + 0 + \theta_1\theta_2\sigma^2 + 0 + 0 + 0 = (1 + \theta_2)\theta_1\sigma^2$
- $h = 2 : 0 + 0 + \theta_2\sigma^2 + 0 + 0 + 0 + 0 + 0 + 0 = \theta_2\sigma^2$
- $h = 3 : 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$

5) Assume that the AR(1) process given by $X_t = \phi X_{t-1} + Z_t$ where $|\phi| < 1$ and Z_t is uncorrelated with $X_s, s < t$ is stationary.

Find its mean and ACF.

Mean: $\mu = E(X_t) = E(\phi X_{t-1} + Z_t) = \phi E(X_{t-1}) + E(Z_t) = \phi\mu + 0 = \phi\mu$

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

$var(X_t) = var(\phi X_{t-1} + Z_t) = \phi^2 var(X_{t-1}) + 2\phi cov(X_{t-1}, Z_t) + var(Z_t) = \phi^2\sigma^2 + \sigma^2 = (1 - \phi^2)\sigma^2$

$\gamma(h) = cov(X_t + h, X_t) = cov(\phi X_{t+h-1} + Z_{t+h}, X_t) = cov(\phi X_{t+h-1}, X_t) + cov(Z_{t+h}, X_t)$

$= cov(\phi X_{t+h-1}, X_t) = \phi\gamma(h-1) = \phi^h\gamma(0)$

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

Why do you need $|\phi| < 1$?

If $|\phi| >= 1$ then the function would have infinite variance.

6) Show that X_t given by $X_t = X_{t-1} + Z_t$ when Z_t is uncorrelated with $X_s, s < t$ is non-stationary.

$X_t = X_{t-1} + Z_t = X_{t-2} + X_{t-1} + Z_t = X_0 + Z_1 + Z_2 + \ldots + Z_t$

$var(X_t) = var(X_0) + var(Z_1 + Z_2 + \ldots + Z_t) = var(X_0) + \sigma^2 + \sigma^2 + \ldots + \sigma^2 = var(X_0) + t\sigma^2$

$var(X_t)$ is dependent on t

7) Show that if Y_t is stationary, $(1 - B)Y_t = Y_t - Y_{t-1}$ is also stationary.

$X_t = Y_t - Y_{t-1}$

$\mu_t = E(X_t) = E(Y_t - Y_{t-1}) = E(Y_t) + E(-Y_{t-1}) = 0$

$\sigma^2 = var(Y_t - Y_{t-1}) = var(Y_t) + cov(Y_t, -Y_{t-1}) + var(-Y_{t-1}) = \sigma^2 - 2\gamma_y(1) + \sigma^2$

$\gamma_x(h) = cov(X_{t+h}, X_t) = cov(Y_{t+h} - Y_{t+h-1}, Y_t - Y_{t-1})$

$= cov(Y_{t+h}, Y_t) + cov(Y_{t+h}, -Y_{t-1}) + cov(Y_t, -Y_{t-1}) + cov(-Y_{t+h-1}, -Y_{t+h-1})$

$= \gamma_y(h) - \gamma_y(h+1) - \gamma_y(h-1) + \gamma_y(h)$

8) Suppose a time series data, $X_t, t = 1, 2, \ldots, n$ follows a quadratic trend.

Show that $(1 - B)^2 X_t$ eliminates the trend.

Quadratic Trend: $m_t = \beta_0 + \beta_1 t + \beta_2 t^2$

$(1 - B)^2 X_t = (1 - B)^2 m_t + (1 - B)^2 Z_t$

$(1 - B)^2 m_t = (1 - 2B + B^2) m_t = m_t - 2Bm_t + B^2 m_t = m_t - 2m_{t-1} + m_{t-2}$

$= \beta_0 + \beta_1 t + \beta_2 t^2 - 2(\beta_0 + \beta 1(t-1) + \beta_2 (t-1)^2) + \beta_0 + \beta_1 (t-2) + \beta_2 (t-2)^2 = 2\beta_2$

Thus $(1 - B)^2 X_t = 2\beta_2 + (1 - B)^2 Z_t$ eliminates the trend.

Also show using the result of 7 that the resulting process is stationary.

Since Z_t is stationary, $(1 - B)Z_t$ is stationary. Applying this to itself $(1 - B)^2 Z_t$ is also stationary.

9) For a time series data X_t with $X_t = m_t + s_t + Y_t$ where m_t is a linear trend and s_t is a seasonal effect of order 12 and Y_t is stationary.

Show that $V_t = (1 - B)(1 - B^1 2)X_t$ eliminates linear trend and seasonality.

$V_t = (1 - B)(1 - B^1 2)m_t + (1 - B)(1 - B^1 2)s_t + (1 - B)(1 - B^1 2)Y_t$

Since m_t is linear $\beta_0 + \beta_1 t$

$(1 - B)(1 - B^1 2)m_t = (1 - B^1 2)(1 - B)m_t = (1 - B^1 2)(m_t - m_{t-1}) = (1 - B^1 2)\beta_1 = 0$

Since s_t is a seasonal of order 12 $s_t = s_{t-12}$

$(1 - B)(1 - B^1 2)s_t = (1 - B)(s_t - s_{t-12}) = 0$

So $V_t = (1 - B)(1 - B^1 2)Y_t$

Also show using the result of 9 that the resulting process is stationary.

10) Show that the random walk defined by $X_t = Y_1 + Y_2 + ... + Y_t$ where Y_t are independent and identically distributed with $Y_t = 1P = 1/2, -1P = 1/2$ is non-stationary.

$E(Y_t) = 1 * 0.5 - 1 * 0.5 = 0$

$var(Y_t) = E(Y_t^2) - [E(Y_t)]^2 = 1^2 * 0.5 + 1^2 * 0.5 - 0 = 1$

$var(X_t) = var(Y_1 + Y_2 + ... + Y_t) = var(Y_1) + var(Y_2) + ... + var(Y_t) = 1 + 1 + ... + 1 = t$

The variation of X_t is dependent on t so the random walk is non-stationary.

Part 2

1) The graph below is the autocorrelation graphs of the data of 4 time series.

Which one of the 4 series are stationary processes and which are not?

Series 1 - 3 are stationary since the autocorrelation function decreases exponentially or cuts-off by h = 1. Series 4 is non-stationary since its autocorrelation function decreases linearly.

Which series is white noise?

Series 3 is white noise since all $\rho(h) = 0$ for $h \geq 1$.

Which series follows the model $x_t = \phi x_{t-1} + w_t$?

Series 1 follows the model since its autocorrelation function decreases exponentially.

Which series follows the model $x_t = w_t + \theta w_{t-1}$?

Series 2 follows the model since its autocorrelation function cuts off at $h = 1$.

2) Suppose a time series data set follows a quadratic trend where the residual is stationary, explain the 2 methods of eliminating the trend.

Method 1:

- Use least squares estimate to estimate β_{0to2} as for a linear regression model.
- Estimate the residual as $\hat{y}_t = x_t - \hat{\beta}_0 - \hat{\beta}_1 t - \hat{\beta}_2 t^2$.
- Use y_t as a stationary time series.

Method 2:

- Use second order differencing $(1 - B)^2 x_t$.
- Resulting series will be stationary and eliminate the trend.

3) Let x_t be a time series data set of monthly production index and u_t be the corresponding unemployment index. We fit the linear regression model $x_t = \beta_0 + \beta_1 u_t + y_t$. Based on the given plots explain why fitting a traditional linear regression model is not appropriate.

The linear regression assumes that the model is white noise, however, the autocorrelation plot shows that residuals are autocorrelated and it is not white noise. The scatter plot show that residuals cannot be assumed to be independent which is required for linear regression modeling.

4) Skip

5) Discuss any smoothing method.

Moving Average - Uses surrounding values on each side to compute an average to replace each value.

Regression Model - Uses previous values to compute an average to replace each value.

6) Consider fitting the time series model $x_t = \alpha cos(\frac{2\pi t}{50}) + \beta cos(\frac{2\pi t}{50}) + z_t$. Explain how to fit this model using a regression approach.

If $u_t = cos(\frac{2\pi t}{50})$ and $v_t = sin(\frac{2\pi t}{50})$, then the model is $x_t = \alpha u_t + \beta v_t + z_t$ which is a regression model.

7) Discuss a statistical hypothesis test for testing that a time series process is white noise.

$H_0 : \rho(h) = 0$ vs $H_a : \rho(h) \neq 0$

If $|\hat{\rho}(h)| < \frac{1.96}{n}$ for all $h > 0$

8) The following is a time series plot generates for a model as in question 6.

One is smoothed with regression modeling and the other by moving average. Which is which?

The first plot used regression because it is an even sine wave where as the second is much more uneven.

Discuss the pros and cons of the 2 smoothing approaches.

Regression

- Can be used to forecast future values assuming the same pattern of the regression continues (natural, engineering).
- Does not consider autocorrelation, assumes that the previous pattern will continue infinitely.

Moving Average

- Assumes nothing about the functional pattern of the data.
- If there is a pattern it cannot use that to forecast.