STAT 436 / 536 - Lecture 12: Key

Stationary Models

Time series data sets often have components that can be identified and modeled in a deterministic fashion. In particular, using regression we can fit models with:

- trends

- seasonal patterns

- autoregressive components of the series itself

- covariate information (potentially of an autoregressive fashion too.)

- As a result, with a well fit regression model, there should not be remaining determinstic features (such as trends or seasonal components) in the residuals of the model. However, the residuals will often be correlated in time.

- Two examples referenced in the textbook would be:

1. monthly values of the southern oscillation index tend to change slowly and may give rise to persistant weather patterns.

2. Negative correlation can also occur, such as an unusually high value (in monthly sales) followed by an unusually low value the previous month.

- We have seen problems related to inference when serial correlation is present in the data; hence, our goal is fit a model to address the correlation in the residuals for valid inference.

- One approach that we have seen is to use an autoregressive process, where AR(p) denotes an autoregressive process of order p
$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p}$$

- AR processes can be used for modeling the time series itself, with covariates, or for the residuals.

Stationary Processes

We have talked about stationarity, but now we focus on more details.

- A time series model $\{x_t\}$ is *strictly stationary* if the joint distribution of $(x_{t_1}, \ldots, x_{t_n})$ is the same as that of $(x_{t_1+m}, \ldots, x_{t_n+m})$. In other words, if the distribution is the same after a shift of m time units.

- Strict stationarity implies that the mean and variance are constant.

- Strict stationarity also implies that the autocorrelation between x_t and x_{t+k} only depends on k and not t.

- It is possible for a series to have constant mean and variance, and also have an autocorrelation function that depends only on the lag k, but not be strictly stationary. These series are known as *second-order stationary* processes.

- Stationary is a desirable property of modeling time series data. Fitting a model with stationarity assumptions, implies that the time series (could be a residual one) is indeed a realization from a stationary process.

- Therefore, the series should be checked to determine if evidence of a trend or seasonal effects persist. Linear modeling (regression) can be used fit the trend or seasonal patterns.

- Thus, after checking residual diagnostics, it is reasonable to treat the residual time series as a stationary process.

Moving Average Models

- We have seen AR processes, but another way to handle serial correlation is with a moving average (MA) process.

- Formally, a MA process of order q is a linear combination of the current white noise term and the q most recent white noise terms
$$x_t = w_t + \beta_1 w_{t-1} + \dots \beta_p w_{t-p}$$
where w_t is a zero mean white noise term and the β values form the linear combination.

- Recall and AR process of order p (with white noise) can be written as
$$x_t = w_t + \alpha_1 x_{t-1} + \dots \alpha_p x_{t-p}$$
. Note that the AR process is a linear combination of past realizations, whereas the MA process is a linear combination of the noise terms.

Moving Average Computation

1. Simulate a moving average process of order 2, finish code below.

```
## SETUP PARAMETERS
q <- 2
beta <- c(.6, .1)
sigma <- 1
num.time <- 100

# INITILIZATION
x <-  rep(0, num.time)
w <- rnorm(num.time , mean = 0, sd = sigma )

# SIMULATE TIME PTS 1 AND 2
x[1] <- w[1]
x[2] <- w[2] + beta[1] * w[1]

# NOW SIMULATE TIME PTS 3 - NUM.TIME
for (time.pt in 3:num.time){
  x[time.pt] <- w[time.pt] + beta[1] * w[time.pt - 1] + beta[2] * w[time.pt - 2]
}
```

2. Simulate several realizations from this MA process and track the mean and variances. Can you determine what the mean and variance of the process should be? (Either mathematically or computationally (hint β matters for one of these))

3. Select several different values of beta and plot the resultant process. Can you summarize the impacts of the β values?

4. Use the ACF and PACF to look the serial correlation in the series.

- The moving average process can also be written using the backshift operator **B**.

$$x_t = (1 + \beta_1 B + \cdots + \beta_q B^q) w_t = \phi_q(B) w_t$$

- The MA process consists of a linear combination of a finite sum of stationary white noise terms, hence the MA process is stationary with time-invariant mean and autocorrelation.

- The means can be derived as:

$$E[x_t] = E[w_t + \beta_1 w_{t-1} + \dots \beta_p w_{t-p}] = 0$$

as $E[w_t] = 0 \forall t$

- Similarly the variance can be computed as:

$Var[x_t] = Var[w_t + \beta_1 w_{t-1} + \dots \beta_q w_{t-q}] = Var(w_t) + \beta_1^2 Var(w_{t-1}) + \dots + beta_q Var(w_{t-q}) = (1 + \beta_1^2 + \dots + \beta_q^2)\sigma^2$

as $Var(w_t) = \sigma^2$

- The autocorrelation function for lag k has three possible values:

1. when $k = 0$, $\gamma_k = 1$

2. when $k > q$, $\gamma_k = 0$

3. for $k = 1, \dots, q$, $\rho_k = \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{q} \beta_i^2}$, where β_0 is defined as 1.

- Consider two MA(1) processes with coefficients equal to $\beta = .5$ and $\beta = 2$ and calculate the correlation ρ_1. *we get $\frac{\beta_1}{1+\beta_1^2}$, which results in $\frac{1/2}{1+1/4}$ and $\frac{2}{1+4}$ these values give the same serial correlation.

- The idea of *invertibility* is focused on a unique representation of an MA process.

- An MA process in *invertible* if it can be expressed as a stationary AR process of infinite order without an error term. In other words, using the MA process $x_t = (1 + \beta B)w_t$, then $x_t = w_t + \beta w_{t-1}$ and $w_{t-1} = x_{t-1} - \beta w_{t-2}$. Using this idea in a recursive manner

$$x_t = w_t + \beta x_{t-1} - \beta^2 x_{t-2} + \beta^3 x_{t-3} - \beta^4 x_{t-4} + \dots$$

  with $|\beta| < 1$.

- Recall the characteristic equation was used to determine stationarity of an AR process. Hence, an MA(q) process is invertible when the roots of $\phi_q(B)$ all exceed 1.

- Most importantly, a MA(q) process is unique only if invertibility is imposed.

Fitting MA processes

- The `arima` function in r can be used to fit an MA process.

- The `arima` function has an order argument that is : "A specification of the non-seasonal part of the ARIMA model: the three integer components (p, d, q) are the AR order, the degree of differencing, and the MA order."

- Furthermore, "The exact likelihood is computed via a state-space representation of the ARIMA process, and the innovations and their variance found by a Kalman filter. The initialization of the differenced ARMA process uses stationarity and is based on Gardner et al (1980)."

- Finally return to your simulation and fit the simulated MA models.