PREDICTING TRAFFIC CONGESTION WITH THE AUTO-REGRESSIVE MOVING AVERAGE (ARMA) MODEL

A Presentation by Stephanie Boggs
Home University- Jackson State University
Program Site- Salisbury University
August 5th, 2019

Explore Emerging Computing In Science and Engineering REU
Mixed Auto-Regressive Moving Average Model – ARMA(p, q)

**ARMA Model** consists of two closely related polynomials used for understanding and perhaps predicting future values in a series. Before considering how the models may be combined they must be examined separately.

\[
X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \Theta_i \varepsilon_{t-i}
\]
AR Model (AR(p)) perspective

Predicts future behavior based on the correlation of past values and the values that succeed them.

\[ X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t \]

Specific lagged values of \( x_t \) are used as predictor variables. Lags are where results from one time period affect following periods.
Variables at time \( (t) \) & order \((i)\)

\[ X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t \]

- **Current time step** is dependent on **previous time steps**
- **Model parameter**
Order of AR model

\[ X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t \]

The number of immediately preceding values in the series that are used to predict the value at the present time

Determining this order may be done by plotting a partial autocorrelation function of data.

At lag \( k \), this is the correlation between series values that are \( k \) intervals apart, accounting for the values of the intervals between.
Error (White Noise)

Randomly Determined (Stochastic)

Characterized by independent normal random variables sampled by Gaussian Distribution (mean = 0)

\[ X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t \]
MA Model (MA(q)) perspective

The error of any period, \( t \), is linearly dependent on previous errors.

\[
X_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}
\]

Fitting the MA estimates is more complicated than it is in autoregressive models (AR models), because the lagged error terms are not observable.
Variables at time (t) and order (i)

\[ X_t = \sum_{i=1}^{q} \theta_i \hat{\epsilon}_{t-i} \]

- This model’s parameter accounts for the weights applied to prior values in the time series.

- Errors for previous periods (Residuals)
Order of MA Model

\[ X_t = \sum_{i=1}^{q} \Theta_i \varepsilon_{t-i} \]

Determining the order of this model may be done by plotting an autocorrelation function of data.

- At lag \( k \), this is the correlation between series values that are \( k \) intervals apart.
ARMA Model is atheoretic and used for prediction of random processes.

Traffic is difficult to model & predict because it lacks inherent symmetry

This model has not been used in traffic prediction
Methodology & Tools

- MATLAB – Econometrics Toolbox
  - Econometric Modeler
  - Correlation functions
- Box-Jenkins Method for ARMA prediction
- Compare experimental results to real-life values
Data Overview

Chicago, Illinois: Kennedy Expressway
O'Hare Int'l Airport ↔ I:290 J.B. Interchange
Approximately: 18 miles
Speed Limit: 70 mph

Observed Variable:
Traffic Congestion Levels:

\[ TCL = \frac{(V_{max} - V_i)}{V_{max}} \]
Traffic Congestion Levels for Kennedy Expressway

INBOUND

OUTBOUND

Traffic Congestion Level

Time
Box Jenkins Method for ARMA Prediction

1. Is the time series stationary?
   - Yes
   - No → Make Time Series Stationary by “Differencing”

2. Model Identification
   - Yes
   - No → Model Estimation

3. Model Estimation
   - Yes → Is Model Adequate?
   - No

4. Is Model Adequate?
   - Yes → Model Forecasting
   - No → Model Estimation

Model Forecasting
Box Jenkins Method for ARMA Prediction

Is the time series stationary?

Yes

Model Identification

Model Estimation

Is Model Adequate?

Yes

Model Forecasting

No

Make Time Series Stationary by “Differencing”
Time series is non-stationary

- Autocorrelation pattern signifies a trend.
- Dependence on time
Box Jenkins Method for ARMA Prediction

Is the time series stationary?

Yes

Model Identification

Model Estimation

Is Model Adequate?

Yes

No

Make Time Series Stationary by “Differencing”

No

Model Forecasting
ARMA become AR(1)MA

- Notation - ARIMA(p,d,q)

- (I) – Integrate
  Eliminated non-stationarity by a process known as “differencing”

- Differencing removes random trends by correlating raw values with a gradient
Non-Stationary to Stationary Time Series

Mean = 0
Constant Variance
Box Jenkins Method for ARMA Prediction

- Is the time series stationary?
  - Yes → Model Identification
  - No → Make Time Series Stationary by “Differencing”

- Model Identification

- Model Estimation
  - No → Is Model Adequate?
  - Yes → Model Forecasting

- Is Model Adequate?
  - Yes → Model Forecasting
  - No → Model Estimation
Correlograms

Inbound Autocorrelation Function

- 95% Confidence interval
- Significant correlations at the first couple of lags, followed by correlations that are not significant.
- 3 significant correlations
ARIMA (3,1,3)
Box Jenkins Method for ARMA Prediction

Is the time series stationary?

Model Identification

Model Estimation

Is Model Adequate?

Make Time Series Stationary by “Differencing”

Model Forecasting
Inbound TCL Model Estimation

![Model Fit Graph]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Standard Error</th>
<th>t Statistic</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.0010351</td>
<td>0.0071638</td>
<td>-0.14448</td>
<td>0.88512</td>
</tr>
<tr>
<td>AR{1}</td>
<td>-0.091047</td>
<td>0.34733</td>
<td>-0.26213</td>
<td>0.79322</td>
</tr>
<tr>
<td>AR{2}</td>
<td>0.39037</td>
<td>0.21629</td>
<td>1.8048</td>
<td>0.071104</td>
</tr>
<tr>
<td>AR{3}</td>
<td>-0.51859</td>
<td>0.22857</td>
<td>-2.2688</td>
<td>0.023281</td>
</tr>
<tr>
<td>MA{1}</td>
<td>0.26441</td>
<td>0.34906</td>
<td>0.75747</td>
<td>0.44877</td>
</tr>
<tr>
<td>MA{2}</td>
<td>-0.16467</td>
<td>0.2933</td>
<td>-0.56145</td>
<td>0.57449</td>
</tr>
<tr>
<td>MA{3}</td>
<td>0.53425</td>
<td>0.19214</td>
<td>2.7805</td>
<td>0.0054269</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0038219</td>
<td>0.0002559</td>
<td>14.9353</td>
<td>1.9427e-50</td>
</tr>
</tbody>
</table>
Outbound TCL Model Estimation

**Model Fit**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Standard Error</th>
<th>t Statistic</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.0010851</td>
<td>0.011272</td>
<td>0.09626</td>
<td>0.92331</td>
</tr>
<tr>
<td>AR{1}</td>
<td>-0.70176</td>
<td>0.25896</td>
<td>-2.7099</td>
<td>0.006731</td>
</tr>
<tr>
<td>AR{2}</td>
<td>-0.43793</td>
<td>0.20949</td>
<td>-2.0905</td>
<td>0.036577</td>
</tr>
<tr>
<td>AR{3}</td>
<td>0.25024</td>
<td>0.24757</td>
<td>1.0108</td>
<td>0.31213</td>
</tr>
<tr>
<td>MA{1}</td>
<td>0.88162</td>
<td>0.2619</td>
<td>3.3663</td>
<td>0.00076193</td>
</tr>
<tr>
<td>MA{2}</td>
<td>0.72256</td>
<td>0.20824</td>
<td>3.4699</td>
<td>0.00052071</td>
</tr>
<tr>
<td>MA{3}</td>
<td>0.0072267</td>
<td>0.26196</td>
<td>0.027587</td>
<td>0.97799</td>
</tr>
<tr>
<td>Variance</td>
<td>0.004298</td>
<td>0.00032998</td>
<td>13.0253</td>
<td>8.7895e-39</td>
</tr>
</tbody>
</table>
Box Jenkins Method for ARMA Prediction

1. Is the time series stationary?
   - Yes: Model Identification
   - No: Make Time Series Stationary by "Differencing"

2. Model Identification

3. Model Estimation
   - No: Is Model Adequate?
     - Yes: Model Forecasting
     - No: Model Estimation

4. Is Model Adequate?
   - Yes: Model Forecasting
   - No: Model Estimation
Future Work

- Model Forecasting
  Forecast function

- Alternate method for determining p & q orders

  **Akaike Information Criterion (AIC)**- It associates the number of model parameters and the goodness of fit. It also associates a penalty factor to avoid over-fitting
Reflection:

What did I learn this summer?

MATLAB
Statistical Analysis
Wallops Flight Facility
Planning
Organization
Volleyball


Questions ?