### <u>An Introduction to</u> <u>AutoRegressive Integrated</u> <u>Moving Average (ARIMA) Models</u>

By Paul Bakker info@cathexisconsulting.ca



124 Merton St., Suite 502 Toronto, Ontario M4S 2Z2 Telephone: (416) 469-9954 Fax: (416) 469-8487 E-mail: info@cathexisconsulting.ca

### Introduction

- ARIMA models can help you determine attribution with high level of confidence
- We'll cover:
  - When to use ARIMA
  - Long vs. short time series designs
  - ARIMA vs. linear regression
  - Weaknesses, challenges, and going for gold



# When to use ARIMA



### When to use ARIMA

- 1. Random assignment is not possible or appropriate
- 2. There is no appropriate comparison group
- 3. Want to determine attribution with high level of confidence



### When to use ARIMA

### Example:

- Cathexis posted a job posting on CES's website on January 22, 2009.
- We want to find out if the posting was an effective way of attracting candidates to Cathexis.

How do we measure the post's effectiveness?







### The Issue of Maturation (a series normal trend)

Were visits normally increasing or decreasing before CES post?





### Begin to Consider Trend

But Periodic Changes in Trend May be Normal





### **1** Additional Pre-Observation

Clear idea of trend and possible effect

Q's remain: Dip normal? Do prepost trend continue?





Clear difference between pre and post trends

can take 100 observations to properly model all patterns in a series





1 Month Pre & 1 Month Post

### ARIMA's Improvements on Linear Regression





### ARIMA's Improvements on Linear Regression

### The Problem of Autocorrelation

**Example: Outliers** 

- Sales in Dec. > 2 SD from the regression line.
- Considered outliers.
- But normal part of series's trend.





## Weaknesses, Challenges, and Going for Gold



# Weaknesses, Challenges, and Going for Gold

### **Threats to Internal Validity**

- Major Threat is History
- Instrumentation (can be ruled out)
- Selection
- Threats can be limited to gain strong confidence in attribution
  - Consider our CES posting example





# Weaknesses, Challenges, and Going for Gold

Practice challenges:

- Collecting all that data!!!!
- Ensuring timely implementation.
- Ability to conduct ARIMA analysis





### ARIMA Analysis Using SPSS's Expert Modeler.

Back to our website example

An ARIMA model will allows to know when and by how much visits increased with given level (95%) of confidence.





### •Use SPSS's Expert Modeler



| del Outliers  |  |  |
|---|--|--|
|   |  |  |
| odel Type   |  |  |
| ○ <u>A</u> ll models  |  |  |
| ○ Exponential smooth  | ning models only   |  |
| ARIMA models only   | /  |  |
| Expert Modeler cor  | nsiders seasonal models  |  |
|   | Current periodicity  | : 7  |
| onts  |  |  |
| ndenendent Variables:   |  |  |
| Event   | Туре   | Variable   |
|   |  |  |
|   |  |  |
|   |  |  |
|   |  |  |
|   |  |  |
|   |  |  |
| Event variables are sp  | ecial independent variab   | es that are used to model  |
| Event variables are spe<br>effects of external occ  | ecial independent variab<br>urrences such as a floo  | es that are used to model<br>d, strike, or introduction of a   |
| Event variables are spo<br>effects of external occ<br>new product line.   | ecial independent variab<br>urrences such as a floo  | es that are used to model<br>d, strike, or introduction of a   |
| Event variables are spe<br>effects of external occ<br>new product line.<br>Check all variables you<br>coded such that 1 indic                   | ecial independent variab<br>urrences such as a floo<br>want to treat as event where                        | es that are used to model<br>d, strike, or introduction of a<br>variables. Each should be                                |
| Event variables are spe<br>effects of external occ<br>new product line.<br>Check all variables you<br>coded such that 1 indic<br>had an effect. | ecial independent variab<br>urrences such as a floo<br>want to treat as event v<br>ates a time point where | es that are used to model<br>d, strike, or introduction of a<br>variables. Each should be<br>an event is thought to have |
| Event variables are spo<br>effects of external occ<br>new product line.<br>Check all variables you<br>coded such that 1 indic<br>nad an effect. | ecial independent variab<br>urrences such as a floo<br>want to treat as event v<br>ates a time point where | es that are used to model<br>d, strike, or introduction of a<br>variables. Each should be<br>an event is thought to have |

- Expert Modeler to only consider ARIMA.
- Expert Modeler to consider seasonality



| Time Series Modeler: Expert Modeler Criteria |   |  |  |
|--|---|--|--|
| Model Outliers                               |   |  |  |
| ✓ Detect outliers automatically              |   |  |  |
| Type of Outliers to Detect                   |   |  |  |
| 1 Additive                                   |   |  |  |
| 2 vel shift                                  |   |  |  |
| Innovational                                 |   |  |  |
| 3 Iransient                                  | 1 |  |  |
| 4  |   |  |  |
| Lo <u>c</u> al trend                         |   |  |  |
| Additive patch                               |   |  |  |
|  | 2 |  |  |
|  |   |  |  |
|  |   |  |  |
|  | 2 |  |  |
|  | J |  |  |
|  |   |  |  |
|  |   |  |  |
|  | 4 |  |  |
|  |   |  |  |
| Continue Cancel Help                         |   |  |  |

Automatically detect outliers.

I.e. significant changes in the series.

<u>Cathexis</u>



### ARIMA Analysis Using SPSS's Expert Modeler.

### Overall Fit of the model

Stationary R-squared: The higher the better.
Ljung-Box Q: If non-significant then model's error is random.

For our example: Stationary R-squared=.695 & Ljung-Box Q = .162

So far, a good model





### ARIMA Analysis Using SPSS's Expert Modeler.

The model determines Jan. 22 (day of CES post) to be a transient outlier with a magnitude of 39.4 with a decay factor of 96%





### Summary

- We'll covered:
  - When to use ARIMA
  - Long vs. short time series designs
  - ARIMA vs. linear regression
  - Weaknesses, challenges, and going for gold
- ARIMA models make it is possible to make claims of attribution with a high level of confidence without random assignment or comparison groups

