# DOING HARD TIME

*Prepared for the*

Corrections Statistics Program Panel
Bureau of Justice Statistics

*October 6, 2004*

*Washington D.C.*

David P. Reilly
Automatic Forecasting Systems Inc.
P.O. Box 563
Hatboro, Pa 19040
215-675-0652

# DOING HARD TIME SERIES

*Prepared for the*

Corrections Statistics Program Panel
Bureau of Justice Statistics

*October 6, 2004*

*Washington D.C.*

David P. Reilly
Automatic Forecasting Systems Inc.
P.O. Box 563
Hatboro, Pa 19040
215-675-0652

# Contact Information for Dave Reilly

- Automatic Forecasting Systems, Inc. (AFS)

  P.O. Box 563

  Hatboro, PA 19040

  Phone: 215-675-0652

  Fax: 215-672-2534

  email: dave@Autobox.com

  Web Site: www.Autobox.com

# You Can Relax Help Is On The Way

# Forecasting History is Always Easier Than Forecasting The Future

# Abraham Lincoln (1809 - 1865) said:

"If we could first know where we are, then whither we are tending, we could then decide what to do and how to do it."

**Data recorded sequentially through time is called “Time Series Data”.**

**The analysis of time series data must use special statistical techniques, called “Time Series Techniques”.**

# If only we had known sooner….

# How Would This Be Accomplished ?

By computing the probability of observing what was observed !

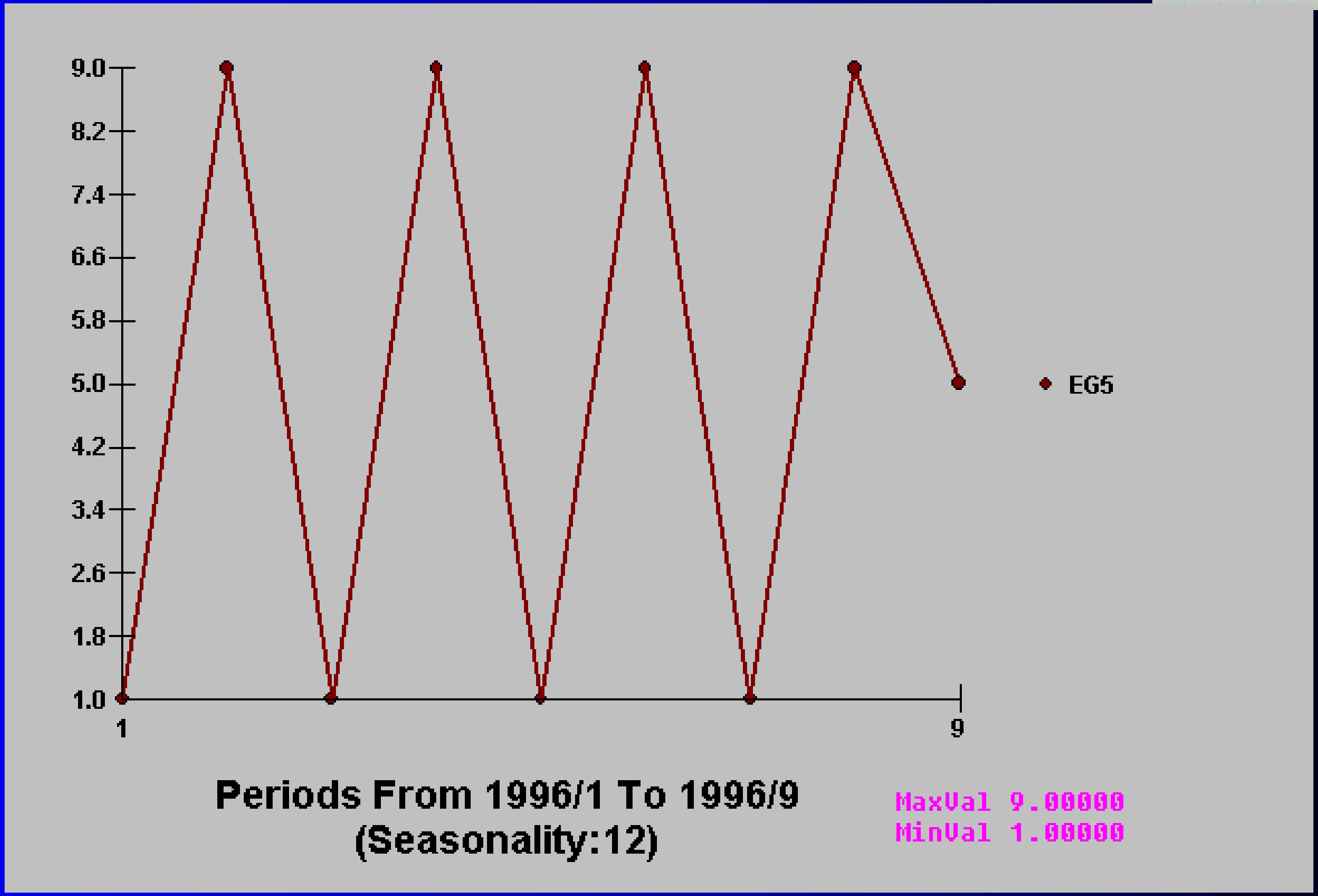
# Early Warning Systems

AUTOMATIC FORECASTING SYSTEMS

Find out why

Early warning systems should not simply detect high and low values, but should detect unusual activity inconsistent with expectations.

# The mean can be unusual



Statistical packages have enormous influence over analysis, especially over that of the less sophisticated user. There is a tendency for the user to do what is readily available in their software.

# Typical Hierarchy of Methods

- Qualitative
  - Judgmental
  - Analogical

# Typical Hierarchy of Methods

- Quantitative: Time Series Analysis
  - Smoothing
  - Trend Decomposition
  - Decomposition (e.g. Seasonal)
  - Box-Jenkins & Autoregressive Models

# Typical Hierarchy of Methods

- Quantitative: Causal Modeling
    - Linear Regression
    - Multiple Regression
    - Econometric Modeling

# A More Precise View of the Hierarchical Structure

- Qualitative
  - Judgmental
  - Analogical
- Quantitative: Time Series Analysis
  - Causal Modeling
  - Memory or Smoothing Models
  - Trend Decomposition

# A More Precise View of the Hierarchical Structure

- Qualitative
  - Judgmental
  - Analogical
- Quantitative: Time Series Analysis
  - Causal
  - Memory
  - Dummy

# State-of-the-Art Modeling Procedures Optimally Combine Three Kinds of Structures

$Y_t$ = Causal + Memory + Dummy

Future Value (at time t) of Variable of Interest

# CAUSAL


Using possible explanatory variables such as

- Temperature
- Unemployment Rates
- Labor Force Size etc.

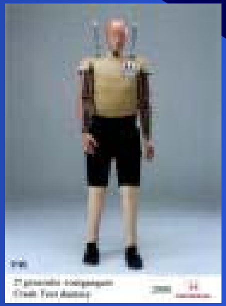
# MEMORY


Using historical values such as Intakes last month, a year ago at this time, perhaps a rate-of-change statistic embodying the auto-projective pattern.

Memory by itself is sometimes incorrectly referred to as Time Series Analysis, whereas TSA in it’s larger definition encompasses both Causals and Dummy Variables

# DUMMY

Using Month-of-the Year Profiles, Growth Patterns over Time (Level Shifts and/or Local Time Trends).



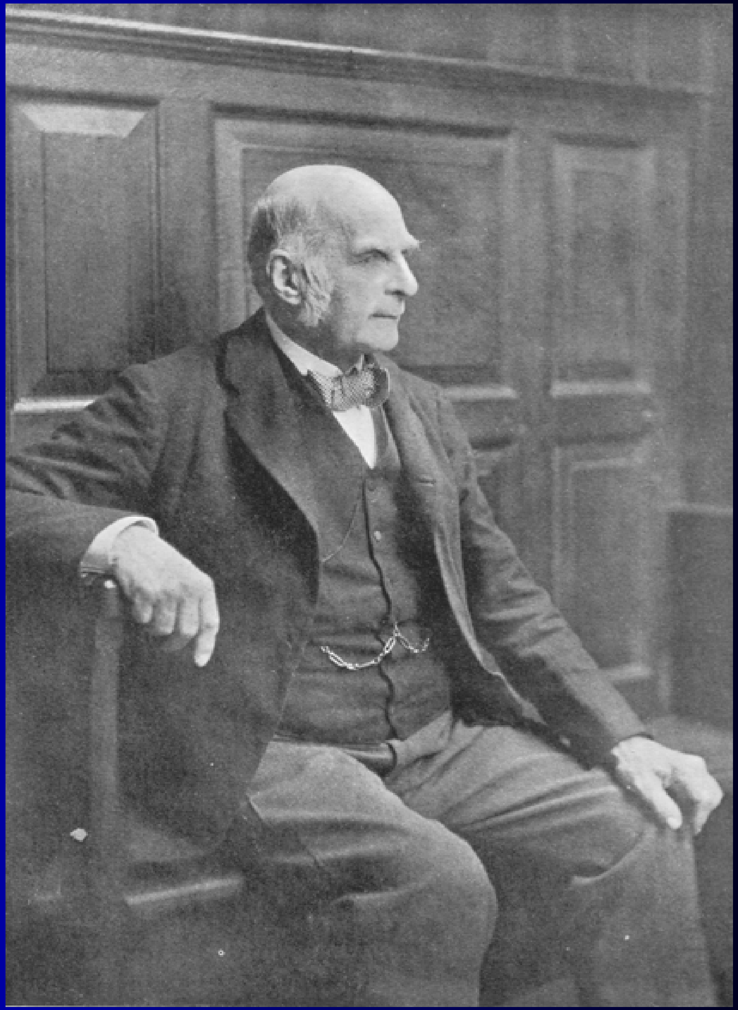
# Let’s Review How These Components Have Been Used

$Y_t$ = **Causal** + Memory + Dummy

# Forecasting History (CAUSAL )

# Sir Francis Galton

- Tropical Explorer
- Eugenicist
- Statistician
- Anthropologist
- Criminologist
- Hereditarian
- Half-cousin of Charles Darwin
- Psychologist



# Galton on Correlation

- December 1888, Galton's "Co-relations and their measurement, chiefly from anthropometric data"
- If both measurements (midparent and child's height) were expressed in terms of their probable errors, then both regression lines had same slope r (closeness of co-relation).
- In addition, "co-relation" was originally used because "correlation" was taken and had different meaning.

# Galton's Problem where Sample 1 might be Arkansas, Sample 2 New York etc.

## Cross-Sectional Data

| | | Characteristics (Measurements) | | |
|---|---|---|---|---|
| | | A | B ..... | Z |
| U | | | | |
| N | | | | |
| C | | | | |
| O | Independent Sample 1 | X1A | X1B | X1Z |
| R | Independent Sample 2 | X2A | X2B | X2Z |
| R | | | | |
| E | . | . | . | . |
| L | . | . | . | . |
| A | . | . | . | . |
| T | . | . | . | . |
| E | Independent Sample N | XNA | XNB | XNZ |
| D | | | | |

# Galton’s Solution

## Linear Regression Line : Concept

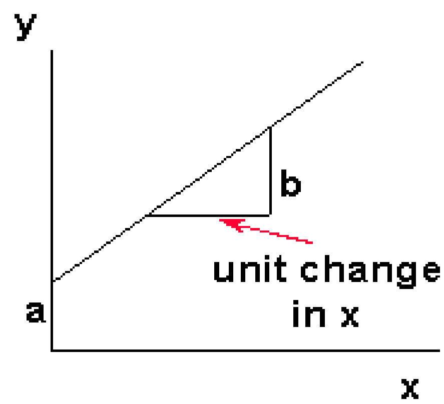
$$y = a + b\,x$$

Predicted Variable

y axis Intercept

Slope of Line

Predictor Variable




The linear regression line makes the sum of the squared residuals a minimum. Hence called the “least squares line”.

# A More Common Problem where Sample 1 is Jan, 2002, Sample 2 Feb, 2002, etc.

## Time Series Data

| | | Characteristics (Measurements) | | | |
|---|---|---|---|---|---|
| C | | A | B | ..... | Z |
| O | Correlated Sample 1 | X1A | X1B | | X1Z |
| R | Correlated Sample 2 | X2A | X2B | | X2Z |
| R | | | | | |
| E | . | . | . | | . |
| L | . | . | . | | . |
| A | . | . | . | | . |
| T | . | . | . | | . |
| E | Correlated Sample N | XNA | XNB | | XNZ |
| D | | | | | |

# Flawed When Applied To Time Series Data

A source for spurious correlation is a common cause acting on the variables.

In the recent *spurious regression* literature in time series econometrics (Granger & Newbold, *Journal of Econometrics,* 1974) the misleading inference comes about through applying the regression theory for stationary series to non-stationary series.

# Flawed When Applied To Time Series Data

The dangers of applying the regression theory for stationary series to non-stationary series were pointed out by G. U. Yule in his 1926 "Why Do We Sometimes Get Nonsense Correlations between Time-series? A Study in Sampling and the Nature of Time-series," *Journal of the Royal Statistical Society,* **89,** 1-69.

# Some Examples of the misuse of Regression/Correlation

- IQ and Foot Size seem to be related
- The more fireman at a fire, the more damage is reported
- The number of Churches in a town seem to be related to the number of Bars.
- Babies Per Capita seems to be related to Storks Per Capita.

# Flawed When Applied To Time Series Data (2)

More generally the misleading inference comes about through applying the regression theory for stationary series to series that have auto-regressive structure.

Recognizing this , early researchers attempted to extract the within relationship (autoregressive structure) and then proceed to examine cross-correlative (among) relationships.

# Flawed When Applied To Time Series Data (3)

Initial attempts to adjust for within relationships included “de-trending” and/or differencing.

Both are usually presumptive and often lead to “Model Specification Bias”.

# Flawed When Applied To Time Series Data (3)

Box and Jenkins codified this process by recognizing that an ARIMA filter is the optimum transform to extract the “within structure” prior to identifying the “among structure”.

They pointed out that both “de-trending” and “differencing” are particular cases of a filter, whose optimized form is an ARMAX model potentially containing both ARIMA and Dummy Variables such as Trends.

# How to Identify the Relationship

The first step to this process is to develop an ARIMA model for each of the user-specified input time series in the equation.

Each series must then be made stationary by applying the appropriate differencing and transformation parameters from its ARIMA model.

# How to Identify the Relationship

Each input series is prewhitened by its own ARIMA model AR (autoregressive) and MA (moving average) factors.

The output series is filtered by the input series AR and MA factors.

The cross correlations between the prewhitened input and output reveal the extent of this interrelationship.

# Why We Filter to Identify

- $Y(_t\ ) = W(B)X(_t) + V(_t\ )$ (equation 1)
- Now if $X(_t) = [t(B)/p(B)]x(_t)$ then $x(_t) = [p(B)/t(B)]X(_t)$
- Using [p(B)/t(B)] on equation (1) we get
- $[p(B)/t(B)]\ Y(_t\ ) = W(B)\ [p(B)/t(B)]\ X(_t) + [p(B)/t(B)]\ V(_t\ )$ or
- $y(_t\ ) = W(B)\ x(_t) + W(_t\ )$ (equation 2)
- Enabling the identification of W(B) since $x(_t)$ is a white noise process and cross-correlations between $y(_t)$ and $x(_t)$ are meaningful as compared to the useless cross-correlations between $Y(_t\ )$ and $X(_t)$
- Note that [p(B)/t(B)] plays no role in W(B)

If $x(_t)$ and $y(_t)$ are bivariate normal ( implies no ARIMA structure within x and within y ) then

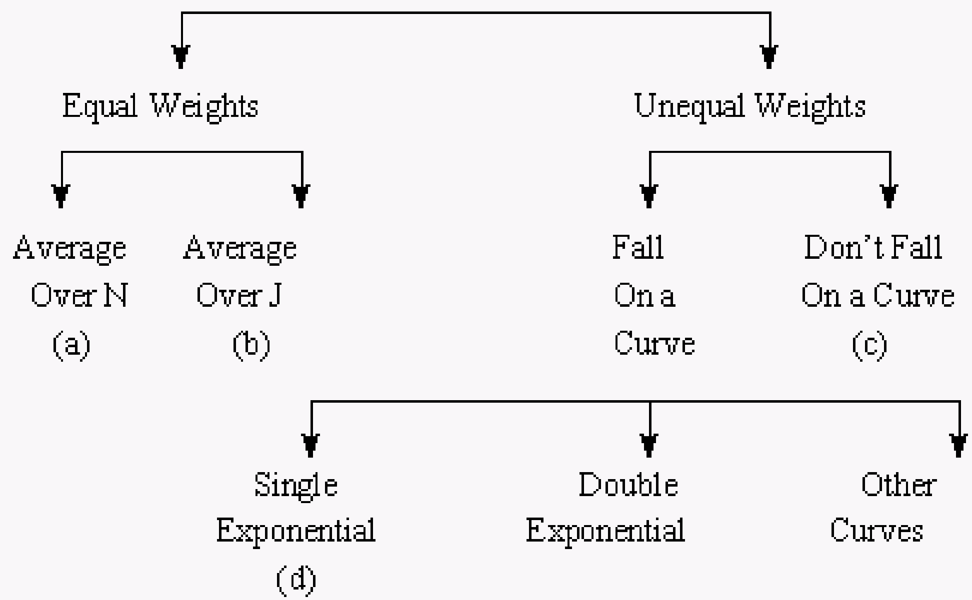
$$\rho = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

# Let's Review How These Components Have Been Used

$Y_t$ = Causal + Memory + Dummy

# Historical development of Memory

# Memory Model



# Auto-Projective Equations

a) $Y_{N+1} = (1/N)*Y_1 + (1/N)*Y_2 + (1/N)*Y_3 + \ldots\ldots\ldots\ldots (1/N)*Y_N$

b) $Y_{N+1} = (1/J)*Y_N + (1/J)*Y_{N-1} + (1/J)*Y_{N-2}$ where $J=3$

c) $Y_{N+1} = .6*Y_N + .3*Y_{N-1} + .1*Y_{N-2}$ where .6,.3,.1 are the weights

a) $Y_{N+1} = (1/N)*Y_1 + (1/N)*Y_2 + (1/N)*Y_3 + \ldots\ldots\ldots\ldots (1/N)*Y_N$

b) $Y_{N+1} = (1/J)*Y_N + (1/J)*Y_{N-1} + (1/J)*Y_{N-2}$ where $J=3$

c) $Y_{N+1} = .6*Y_N + .3*Y_{N-1} + .1*Y_{N-2}$ where .6,.3,.1 are the weights

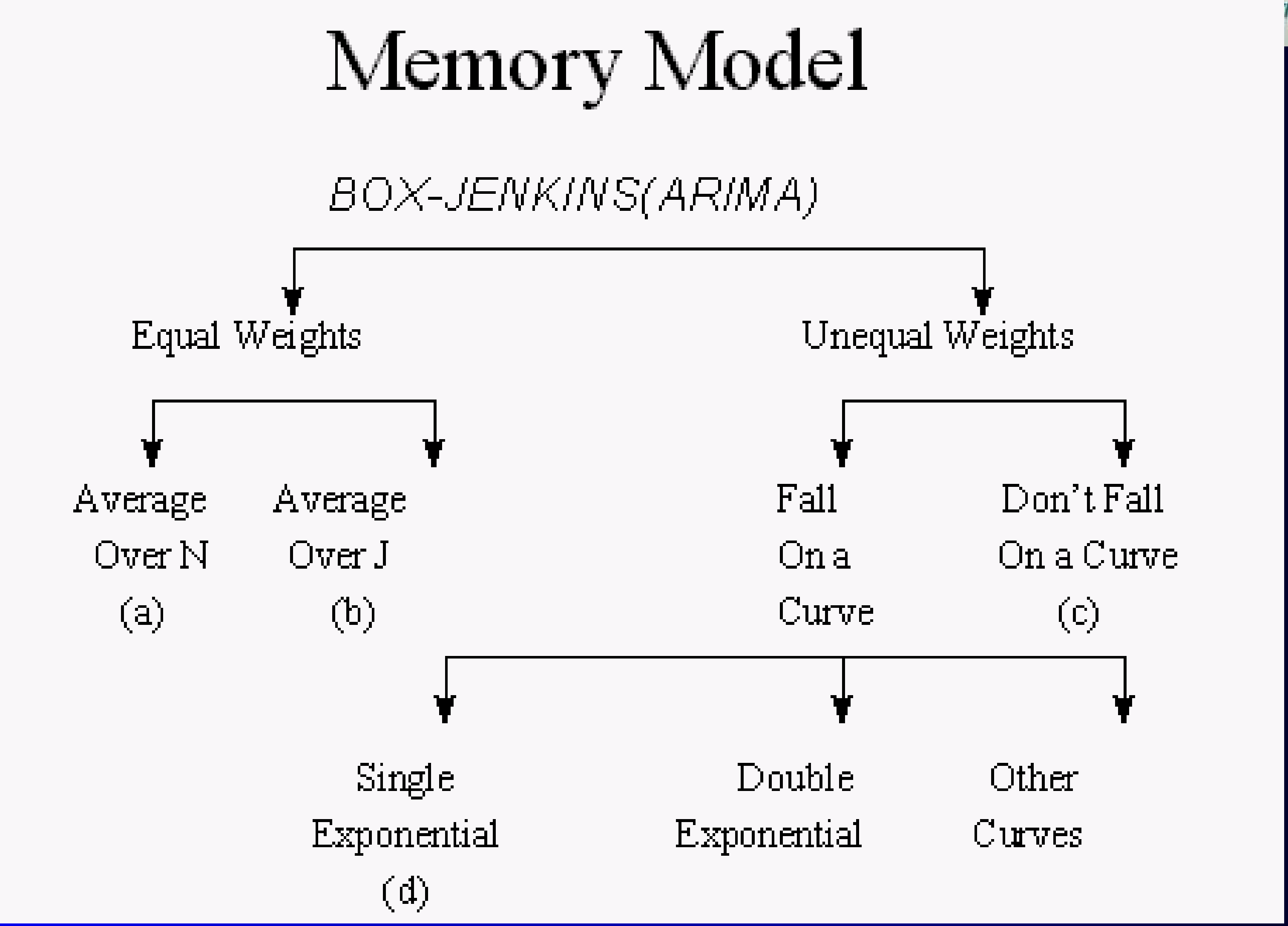
d) $Y_{N+1} = C1*Y_N + C2*Y_{N+1} + C3*Y_{N+2} + CK*Y_{N+K}$ where C1,C2,C3

are the weights for example:

$C1 = .8$, $C2 = .2*.8$, $C3 = .2*.2*.8$ , etc. $CK = .2^{(K-1)}.8$

$C1 = .8$ $C2 = .16$ $C3 = .032$

# The Family Tree



# Consider an "N Period" Equally Weighted Model

$$Y_{N+1} = (1/N)*Y_1 + (1/N)*Y_2 + (1/N)*Y_3 + \ldots\ldots\ldots\ldots (1/N)*Y_N$$

$$Y_{N+1} = (1/N)*Y_1 + (1/N)*Y_2 + (1/N)*Y_3 + \ldots\ldots\ldots\ldots (1/N)*Y_N$$

# The Mechanics of a 60 day Weighted Average

If you wished to use a 60 period equal weighted average you would need to have available the most recent 60 values. In the early days of computing storage was a major problem thus Statistical Innovation was in order.

# Relationship Between Number of Observations in an Equally Weighted Average and The Exponential Model Smoothing Coefficient in terms of Average Age of the Data

| Number of Observations | Variance of Estimate | Smoothing Constant |
|---|---|---|
| 3 | 0.333 | 0.5 |
| 4 | 0.25 | 0.4 |
| 5 | 0.2 | 0.333 |
| 5.67 | 0.177 | 0.3 |
| 6 | 0.167 | 0.286 |
| 9 | 0.111 | 0.2 |
| 12 | 0.083 | 0.154 |
| 18 | 0.056 | 0.105 |
| 19 | 0.053 | 0.1 |
| 24 | 0.042 | 0.08 |
| 39 | 0.026 | 0.05 |
| 52 | 0.019 | 0.038 |
| 199 | 0.005 | 0.01 |

R.G. Brown in 1961 developed the concept of capturing historical data in a forecast and then using that forecast and an adjustment for the last error to get a new forecast.

Y(new)=(1-a)*Y(old)+a*error

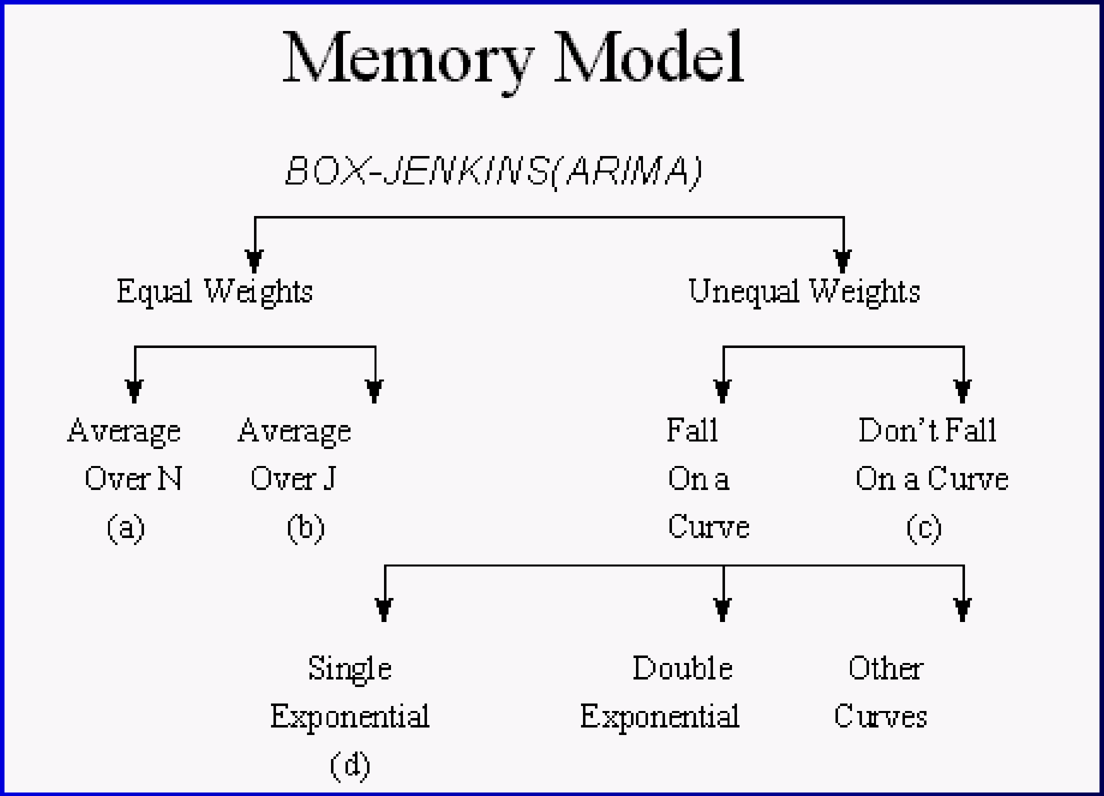
There was no theoretical development used just the idea that one could quickly compute an updated forecast and only two values were required to be stored.

1. The Previous Forecast
2. The Smoothing Coefficient(a)

In terms of selecting the appropriate Smoothing Coefficient, one was told to try different values between 0. and 1.0 and see which one you like best. Failing that you could call NYC and find out what they liked !

This method had an intuitive appeal as it was equivalent to exponentially forgetting the past or equivalently equally weighting a recent set without having to store all the data. The IT folks just loved it as it was fast and efficient if not as accurate as could be developed

# The Family Tree



In 1957, Julius Shiskin developed an ad hoc approach to computing a Seasonal Factor. This was not based on mathematical/statistical theory but rather on an arithmetic/simple approach , full of assumptions, to computing a weighted average. His procedure was called X11 and is widely used.

AUTOMATIC FORECASTING SYSTEMS

In 1963 Box and Jenkins suggested using lagged correlation coefficients to IDENTIFY the nature of the required memory structure rather than assuming it as Brown and Shiskin had.

Box was quoted as saying that his method would have been more aptly named “X12” since X11 is a particular subset and he drew heavily on the inspiration of Shiskin while generalizing and objectifying the analysis.

The relatively intense pattern identification strategy suggested the need to mechanize the process.

This lead rather naturally into pattern recognition schemes to automatically identify the form of the models. AUTOBOX was introduced in the early 70's

The "technical approach method" popular in the financial markets is a form of ARIMA or Autoprojective Modelling.

Similarly "The Rate of Change" Procedures are also a form of ARIMA.

For example Smoothed Rate of Change (SROC) first calculates a 13-day exponential moving average of closing price. Then calculate a 21-day Rate of Change of the exponential moving average.

# A Memory Model is a "Poor Man's Causal Model"

If $Y(T) = f[X(T)]$ (1)

Then $Y(T-1) = f[X(T-1)]$ (1A)

and inverting we get

$X(T-1) = g[Y(T-1)]$ (1B)

Now, if $X(T) = h[X(T-1)]$ (2)

then using (1B) we get

$X(T) = i[Y(T-1)]$ (3)

Substituting (3) into Equation (1) for $X(T)$ yields

$Y(T) = j[Y(T-1)]$ (4)

Thus the History of a series can be a proxy for an omitted Causal Series

# Let's Review How These Components Have Been Used

$Y_t$ = Causal + Memory + Dummy

# Historical development of Dummy

Early researchers assumed Trend Models and Additive Seasonal Factors like the Holt-Winters Class of Models. Again identification was bypassed and Estimation was conducted based upon an assumed model.

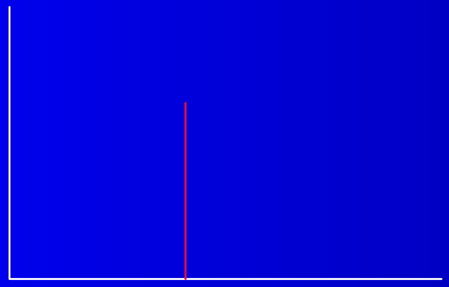
No thought was given to distinguishing between Level and Trend Changes or the detection of break points in trends. No consideration was given to detecting the onset of “seasonal factors”

Intervention Detection schemes introduced in the early 1980's suggested the empirical construct of Dummy Variables. Dummy Variables are related to Trends and Level Shifts

# The Family of Dummy Variables

Pulse $Z_t = 0,0,0,0,1,0,0,0$

Level Shift $Z_t = 0,0,0,0,1,1,1,1,1,,,,$



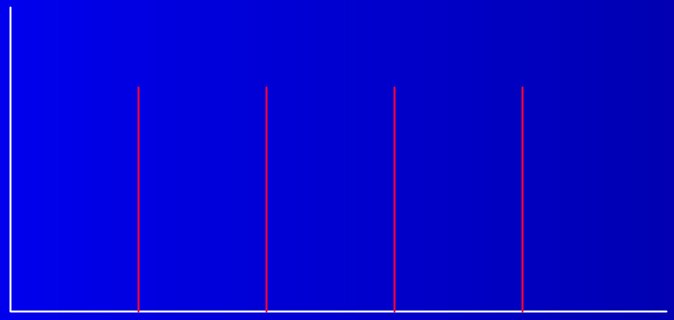
Pulse

Level Shift

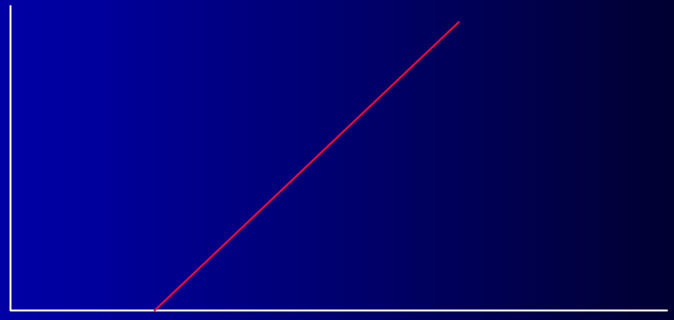
# The Family of Dummy Variables

Seasonal Pulse $Z_t = 0,1,0,0,0,1,0,0,0,1,,,,,$

Time Trend $Z_t = 0,0,0,0,1,2,3,4,5,,,,,$



Seasonal Pulse



Time Trend


Not adjusting for outliers?
Get outta' town!
AFS

# Outliers

- One time events that need to be “corrected for” in order to properly identify the general term or model
- Consistent events (i.e. holidays, events) that should be included in the model so that the future expected demand can be tweaked to anticipate a pre-spike, post spike or at the moment of the event spike.
- If you can’t identify the reason for the outlier than you will not get to the root of the process relationship and be relegated to the passenger instead of the driver

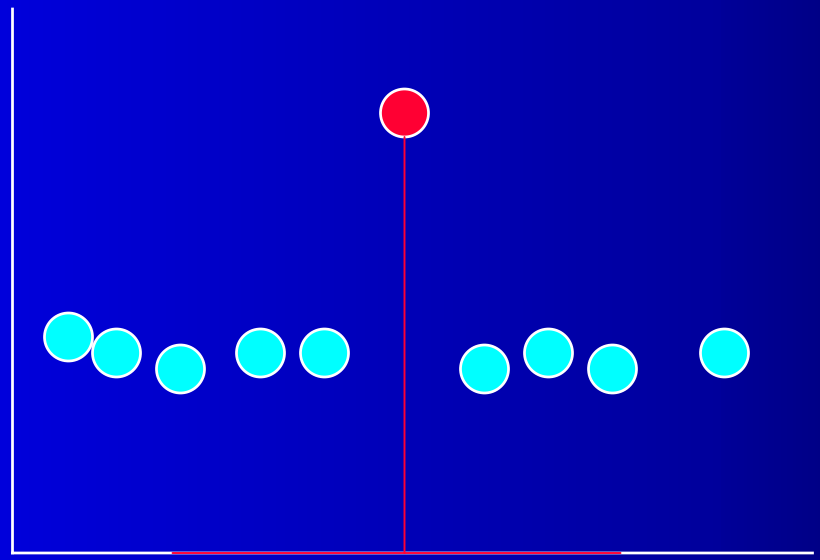
# OUTLIERS: WHAT TO DO ABOUT THEM?

- OLS procedures are INFLUENCED strongly by outliers. This means that a single observation can have excessive influence on the fitted model, the significance tests, the prediction intervals, etc.
- Outliers are troublesome because we want our statistical models to reflect the MAIN BODY of the data, not just single observations.

# Example of a Pulse Intervention

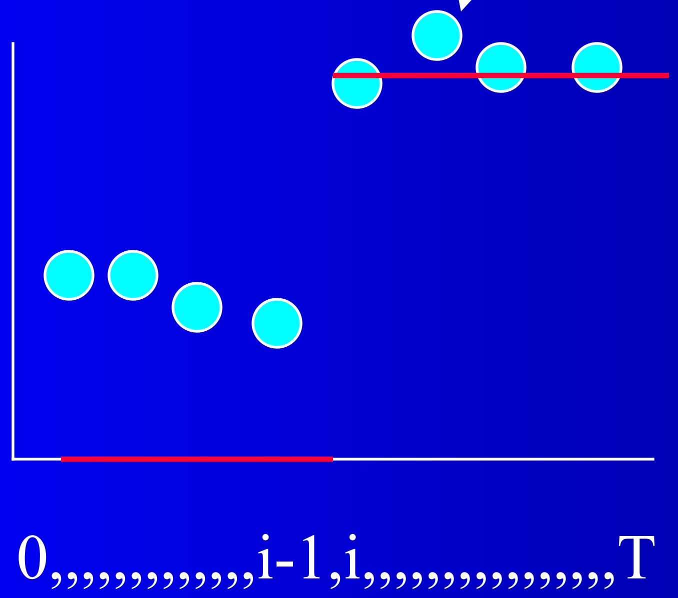
$Z_t$ represents a pulse or a one-time intervention at time period 6.

$Z_t = 0,0,0,0,0,1,0,0,0$



# Modeling Interventions - Level Shift

If there was a level shift and not a pulse then it is clear that a single pulse model would be inadequate thus $Y_t = BO + B3Z_t + U_t$




Assume the appropriate $Z_t$ is

$Z_t = 0,0,0,0,1,1,1,1,1,1,,,,,,,,T$

or $\left.\begin{array}{l} Z_t = 0 \quad t < i \\ Z_t = 1 \quad t > i-1 \end{array}\right\}$

# Modeling Interventions - Seasonal Pulses

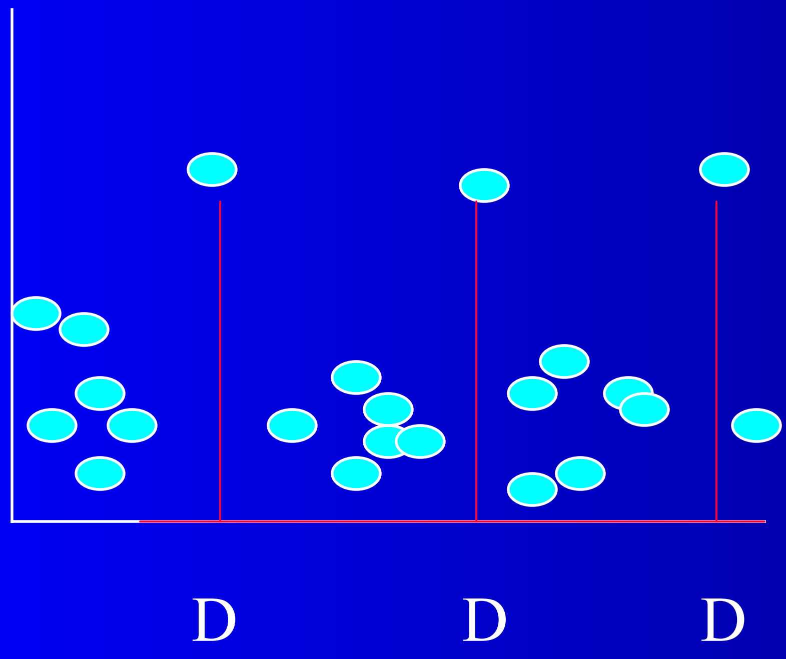
There are other kinds of pulses that might need to be considered otherwise our model may be insufficient. For example, December sales are high.

The data suggest this model

$$Y_t = BO + B3Z_t + U_t$$

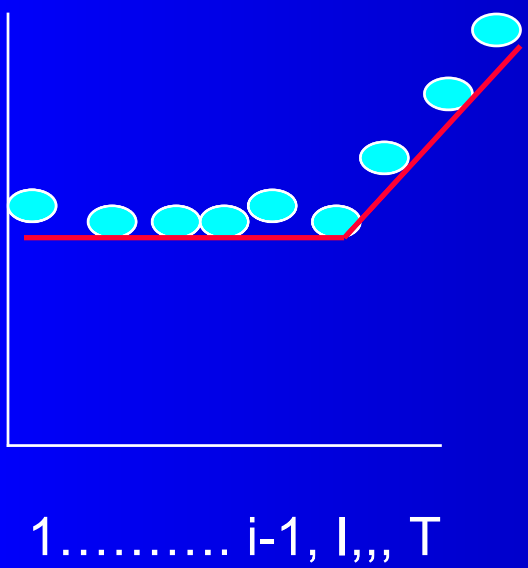
$Z_t = 0 \quad i <> 12,24,36,48,60$

$Z_t = 1 \quad i = 12,24,36,48,60$



# Modeling Interventions – Local Time Trend

The fourth and final form of a deterministic variable is the the local time trend.  For example,




The appropriate form of $Z_t$ is

$Z_t = 0 \quad t < i$

$Z_t = 1 \quad (t-(i-1)) * 1 >= i$

$Z_t = 0,0,0,0,0,0,1,2,3,4,5,,,,,$

# Intervention Model

**Response Function**
(Describes the timing and form of the intervention)

**Error Component**
(Accounts for underlying ARIMA structure of the time-series)

$$Z_t = \boxed{\frac{\omega(B)B^b}{\delta(B)} I_t} + \boxed{\frac{\theta(B)}{\phi(B)(1-B)^d} a_t}$$

where

$\omega(B) = \omega_0 - \omega_1 B - \ldots - \omega_s B^s$,

$\delta(B) = 1 - \delta_1 B - \ldots - \delta_r B^r$,

and $b$ is the time delay for the intervention effect.

$I_t$ is an intervention variable, and

$a_t$ is a zero mean white noise process.

And,

$\theta(B) = \left(1 - \theta_1 B - \ldots - \theta_q B^q\right)$,

$\phi(B) = \left(1 - \phi_1 B - \ldots - \phi_p B^p\right)$,

and $d$ is the level of differencing applied to the series.

The advantages of a time-series Box-Jenkins approach versus a classic multiple regression approach are:

# Advantages of Box-Jenkins

- Omitted stochastic series can be proxied with the ARIMA structure
- Omitted Deterministic series can be empirically identified (Intervention Detection)

# Advantages of Box-Jenkins

- The form of the seasonality can either be auto-projective
  ( i.e. project from seasonal lags) or use one or more Seasonal Dummies versus using them all.
- Furthermore the intensity of the seasonal factors may have changed over time.

# Advantages of Box-Jenkins

- The form of the non-stationarity can be one or more local trends and/or level shifts or differencing versus the assumption of one monotonic trend

AUTOMATIC FORECASTING SYSTEMS

# Advantages of Box-Jenkins

- The form of the relationship can be either fixed for a number of periods or dynamic (ripple effect)
- It can have a period of delay as compared to a pure fixed effect ( i.e. change in x immediately effects y but no other y)

# From 500 Miles High
# This is a Straight-forward Business Intelligence Problem

- We observe Intake Data for a particular geographical area for a number of months. We know what some other demographic variables were.
- We know what the weather was.

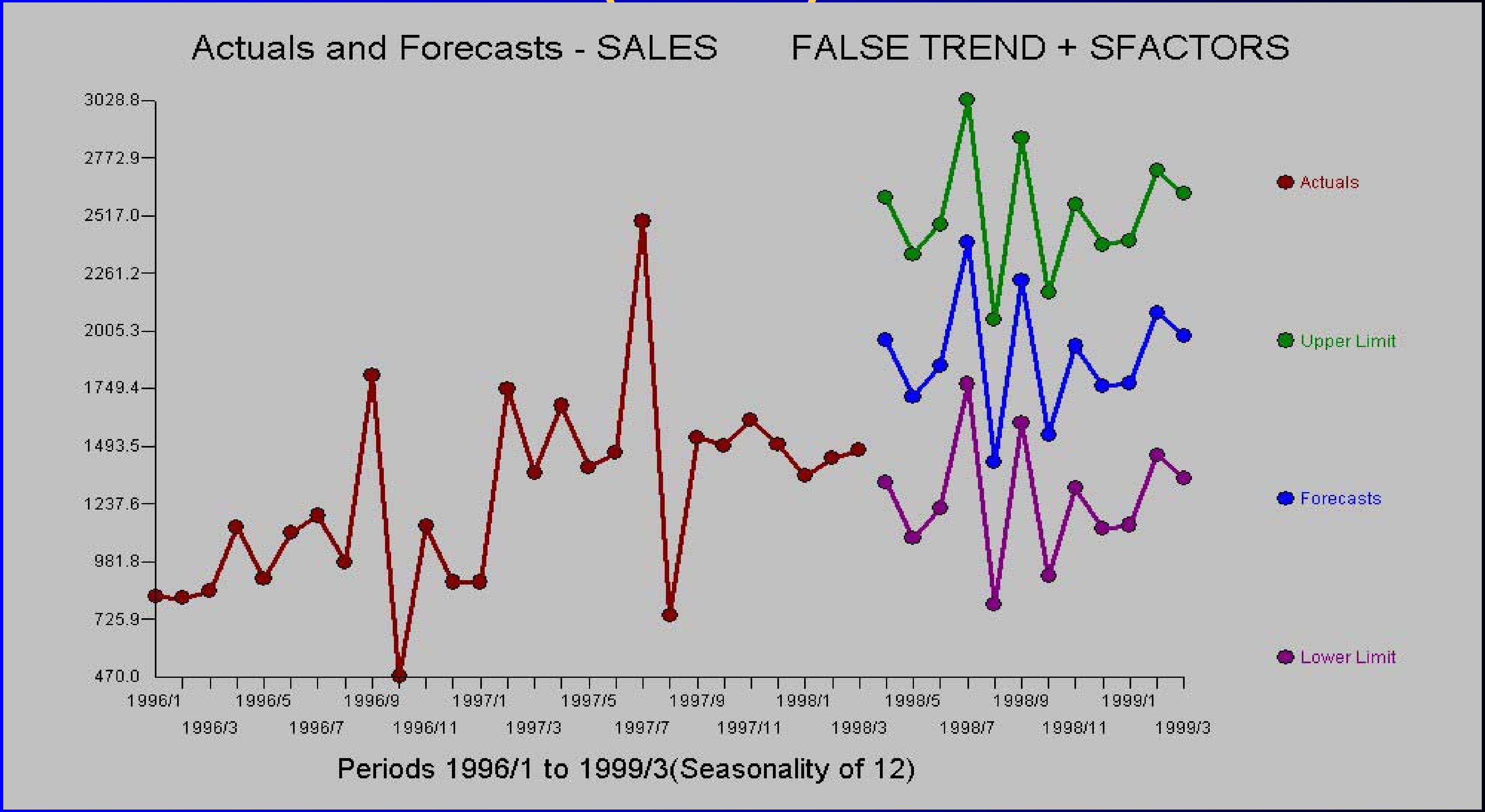
# From 0 Miles High This is a Difficult Statistical Modeling Problem !

- What we don’t know is which of the known variables have an effect and the temporal form of that effect.
- We don’t know how to use historical values of Intake , if at all.
- We don’t know if there is a month-of-the-year effect.
- We don’t know about the effect of unusual activity that may have occurred during the observed history.

- Determine which of the user-suggested input series are statistically significant and what lags are appropriate.
- Determine what lags are needed of the output series
- Determine how the variability changes over time
- Determine how the parameters change over time
- Determine if the model/parameters differ by geographical area.

Traditional techniques assumed a Model and then selected that model that was deemed the “best”.

# Textbook Example (Bad)



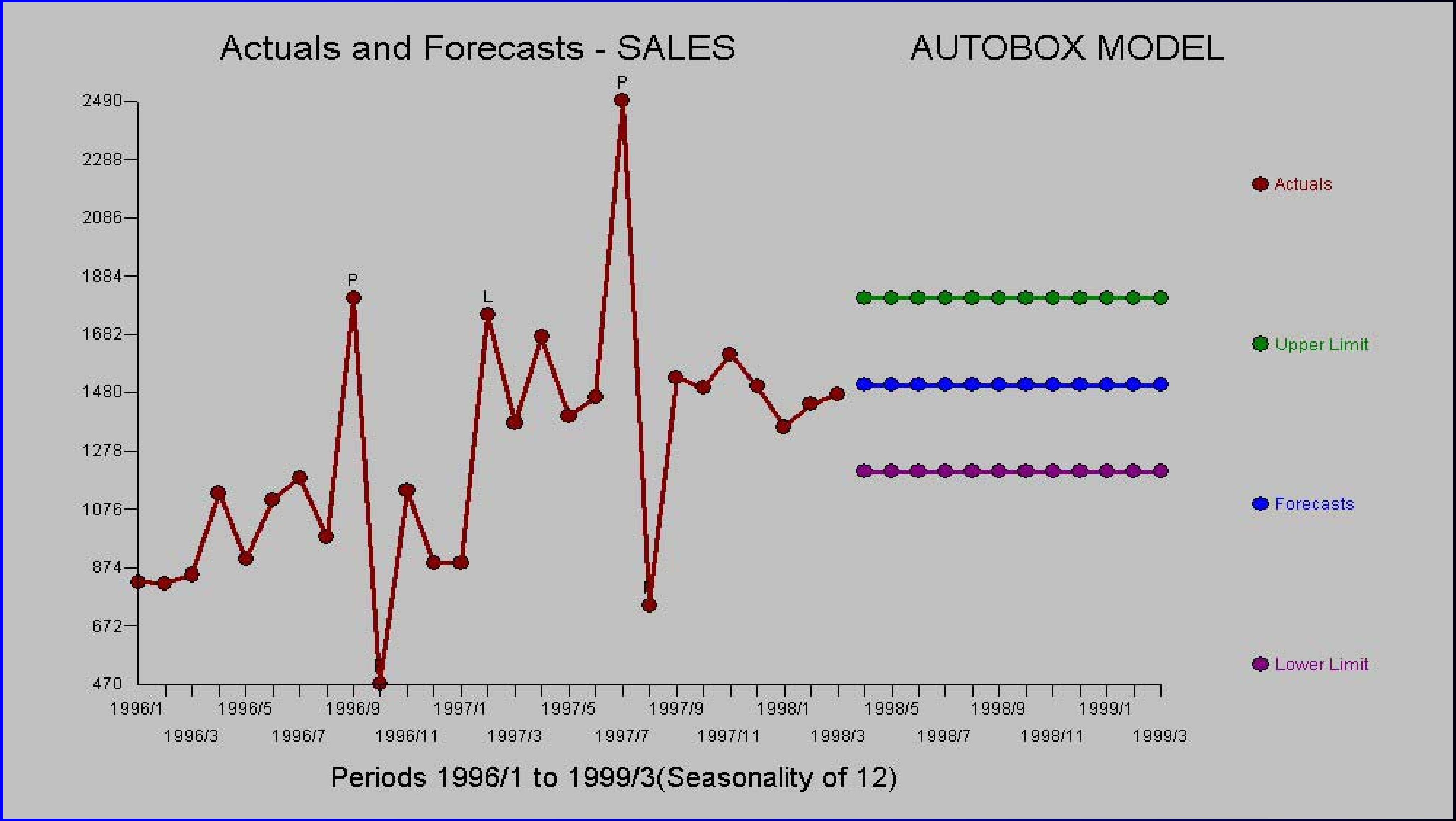
# How do Seasonal Factor models get Fooled?

- y(t) = b0 + b1*t where t is time one generates t residuals or errors

- a(1),a(2),a(3).....a(27) . If one were then to average a(1)+a(13)+a(25) to get a January effect and similarly for each of the other other 11 months, then one would get a “seasonal forecast” all without any formal test of seasonality

- Unusual values become part of the Seasonal Process rather than being isolated or identified as being exceptional.

# AUTOBOX MODEL



# An Illustrative Example From The DC Department of Corrections.

View Options Process Help

torical Data | Future Values | Forecast Data | Graph | Reports | WhatIf

| | TOTAL INTAKES | TOTALREPCRIM | TOTALCASES | LABOR FORCE | UNEMPLOYMEN | TEMP+DEF F | RI |
|---|---|---|---|---|---|---|---|
| 02/1 | 1419.00000000 | 3317.00000000 | 1528.00000000 | 302829.00000000 | 6.80000000 | 41.60000000 | 62.0 |
| 02/2 | 1380.00000000 | 3017.00000000 | 1399.00000000 | 302288.00000000 | 7.00000000 | 42.60000000 | 52.0 |
| 02/3 | 1255.00000000 | 3213.00000000 | 1605.00000000 | 303331.00000000 | 6.30000000 | 47.70000000 | 63.0 |
| 02/4 | 1309.00000000 | 3103.00000000 | 1498.00000000 | 303102.00000000 | 5.70000000 | 60.00000000 | 63.0 |
| 02/5 | 1380.00000000 | 3361.00000000 | 1750.00000000 | 301414.00000000 | 6.10000000 | 65.20000000 | 66.0 |
| 02/6 | 1377.00000000 | 3235.00000000 | 1711.00000000 | 307704.00000000 | 7.00000000 | 76.10000000 | 65.0 |
| 02/7 | 1541.00000000 | 3260.00000000 | 1818.00000000 | 313000.00000000 | 6.80000000 | 80.90000000 | 62.0 |
| 02/8 | 1577.00000000 | 3438.00000000 | 1896.00000000 | 305467.00000000 | 6.30000000 | 81.10000000 | 66.0 |
| 02/9 | 1317.00000000 | 3215.00000000 | 1363.00000000 | 299518.00000000 | 6.20000000 | 73.00000000 | 71.0 |
| 02/10 | 1453.00000000 | 3822.00000000 | 1607.00000000 | 299505.00000000 | 6.30000000 | 58.70000000 | 81.0 |
| 02/11 | 1425.00000000 | 3919.00000000 | 1491.00000000 | 300200.00000000 | 6.30000000 | 47.10000000 | 78.0 |
| 02/12 | 1430.00000000 | 3772.00000000 | 1440.00000000 | 298371.00000000 | 6.20000000 | 37.20000000 | 69.0 |
| 03/1 | 1432.00000000 | 3438.00000000 | 1528.00000000 | 296619.00000000 | 6.50000000 | 31.10000000 | 64.0 |
| 03/2 | 1125.00000000 | 2383.00000000 | 1399.00000000 | 299382.00000000 | 7.10000000 | 33.70000000 | 63.0 |
| 03/3 | 1375.00000000 | 3239.00000000 | 1605.00000000 | 301765.00000000 | 6.30000000 | 47.10000000 | 60.0 |
| 03/4 | 1330.00000000 | 3425.00000000 | 1498.00000000 | 300176.00000000 | 6.50000000 | 55.10000000 | 60.0 |
| 03/5 | 1396.00000000 | 3695.00000000 | 1750.00000000 | 300469.00000000 | 6.10000000 | 61.70000000 | 66.0 |
| 03/6 | 1455.00000000 | 3715.00000000 | 1711.00000000 | 308740.00000000 | 7.30000000 | 71.40000000 | 68.0 |
| 03/7 | 1505.00000000 | 3818.00000000 | 1809.00000000 | 314039.00000000 | 6.90000000 | 77.80000000 | 68.0 |
| 03/8 | 1488.00000000 | 3599.00000000 | 1896.00000000 | 306095.00000000 | 7.30000000 | 78.80000000 | 70.0 |
| 03/9 | 1393.00000000 | 3371.00000000 | 1363.00000000 | 300750.00000000 | 6.70000000 | 70.50000000 | 71.0 |
| 03/10 | 1516.00000000 | 3371.00000000 | 1607.00000000 | 302902.00000000 | 7.00000000 | 57.50000000 | 70.0 |
| 03/11 | 1337.00000000 | 3371.00000000 | 1491.00000000 | 300815.00000000 | 6.90000000 | 53.10000000 | 66.0 |
| 03/12 | 1290.00000000 | 3371.00000000 | 1440.00000000 | 295679.00000000 | 6.60000000 | 39.20000000 | 66.0 |
| 04/1 | 1504.00000000 | 3072.00000000 | 1528.00000000 | 298492.00000000 | 6.60000000 | 30.60000000 | 64.0 |
| 04/2 | 1549.00000000 | 2515.00000000 | 1399.00000000 | 302192.00000000 | 6.70000000 | 38.20000000 | 63.0 |
| 04/3 | 1668.00000000 | 2663.00000000 | 1605.00000000 | 302638.00000000 | 6.80000000 | 48.90000000 | 60.0 |
| 04/4 | 1624.00000000 | 2757.00000000 | 1498.00000000 | 301252.00000000 | 6.70000000 | 57.40000000 | 60.0 |
| 04/5 | 1629.00000000 | 2835.00000000 | 1750.00000000 | 296847.00000000 | 7.00000000 | 71.90000000 | 66.0 |
| 04/6 | 1566.00000000 | 3097.00000000 | 1711.00000000 | 303574.00000000 | 7.70000000 | 73.40000000 | 68.0 |
| 04/7 | 1670.00000000 | 3078.00000000 | 1809.00000000 | 307939.00000000 | 8.20000000 | 78.60000000 | 68.0 |
| 04/8 | 1701.00000000 | 2893.00000000 | 1896.00000000 | 300239.00000000 | 8.30000000 | 85.00000000 | 72.0 |

**Series Properties**

Observations: 32

Forecasts: 12

Series: 8

Major Period: 2002

Minor Period: 1

Frequency: 12

Apply Cancel

nt Status | Engine = M | 9/28/2004 | 8:39 AM

View Options Process Help

orical Data | Future Values | Forecast Data | Graph | Reports | WhatIf

| | TOTALCASES | LABOR FORCE | UNEMPLOYMEN | TEMP+DEF F | RH | TOTAL RELEASES |
|---|---|---|---|---|---|---|
| 02/1 | 1528.00000000 | 302829.00000000 | 6.80000000 | 41.60000000 | 62.00000000 | 1386.00000000 |
| 02/2 | 1399.00000000 | 302288.00000000 | 7.00000000 | 42.60000000 | 52.00000000 | 1292.00000000 |
| 02/3 | 1605.00000000 | 303331.00000000 | 6.30000000 | 47.70000000 | 63.00000000 | 1371.00000000 |
| 02/4 | 1498.00000000 | 303102.00000000 | 5.70000000 | 60.00000000 | 63.00000000 | 1281.00000000 |
| 02/5 | 1750.00000000 | 301414.00000000 | 6.10000000 | 65.20000000 | 66.00000000 | 1392.00000000 |
| 02/6 | 1711.00000000 | 307704.00000000 | 7.00000000 | 76.10000000 | 65.00000000 | 1274.00000000 |
| 02/7 | 1818.00000000 | 313000.00000000 | 6.80000000 | 80.90000000 | 62.00000000 | 1419.00000000 |
| 02/8 | 1896.00000000 | 305467.00000000 | 6.30000000 | 81.10000000 | 66.00000000 | 1501.00000000 |
| 02/9 | 1363.00000000 | 299518.00000000 | 6.20000000 | 73.00000000 | 71.00000000 | 1300.00000000 |
| 02/10 | 1607.00000000 | 299505.00000000 | 6.30000000 | 58.70000000 | 81.00000000 | 1365.00000000 |
| 02/11 | 1491.00000000 | 300200.00000000 | 6.30000000 | 47.10000000 | 78.00000000 | 1294.00000000 |
| 02/12 | 1440.00000000 | 298371.00000000 | 6.20000000 | 37.20000000 | 69.00000000 | 1466.00000000 |
| 03/1 | 1528.00000000 | 296619.00000000 | 6.50000000 | 31.10000000 | 64.00000000 | 1476.00000000 |
| 03/2 | 1399.00000000 | 299382.00000000 | 7.10000000 | 33.70000000 | 63.00000000 | 1106.00000000 |
| 03/3 | 1605.00000000 | 301765.00000000 | 6.30000000 | 47.10000000 | 60.00000000 | 1495.00000000 |
| 03/4 | 1498.00000000 | 300176.00000000 | 6.50000000 | 55.10000000 | 60.00000000 | 1408.00000000 |
| 03/5 | 1750.00000000 | 300469.00000000 | 6.10000000 | 61.70000000 | 66.00000000 | 1292.00000000 |
| 03/6 | 1711.00000000 | 308740.00000000 | 7.30000000 | 71.40000000 | 68.00000000 | 1374.00000000 |
| 03/7 | 1809.00000000 | 314039.00000000 | 6.90000000 | 77.80000000 | 68.00000000 | 1423.00000000 |
| 03/8 | 1896.00000000 | 306095.00000000 | 7.30000000 | 78.80000000 | 70.00000000 | 1360.00000000 |
| 03/9 | 1363.00000000 | 300750.00000000 | 6.70000000 | 70.50000000 | 71.00000000 | 1266.00000000 |
| 03/10 | 1607.00000000 | 302902.00000000 | 7.00000000 | 57.50000000 | 70.00000000 | 1621.00000000 |
| 03/11 | 1491.00000000 | 300815.00000000 | 6.90000000 | 53.10000000 | 66.00000000 | 1398.00000000 |
| 03/12 | 1440.00000000 | 295679.00000000 | 6.60000000 | 39.20000000 | 66.00000000 | 1403.00000000 |
| 04/1 | 1528.00000000 | 298492.00000000 | 6.60000000 | 30.60000000 | 64.00000000 | 1242.00000000 |
| 04/2 | 1399.00000000 | 302192.00000000 | 6.70000000 | 38.20000000 | 63.00000000 | 1457.00000000 |
| 04/3 | 1605.00000000 | 302638.00000000 | 6.80000000 | 48.90000000 | 60.00000000 | 1600.00000000 |
| 04/4 | 1498.00000000 | 301252.00000000 | 6.70000000 | 57.40000000 | 60.00000000 | 1722.00000000 |
| 04/5 | 1750.00000000 | 296847.00000000 | 7.00000000 | 71.90000000 | 66.00000000 | 1713.00000000 |
| 04/6 | 1711.00000000 | 303574.00000000 | 7.70000000 | 73.40000000 | 68.00000000 | 1741.00000000 |
| 04/7 | 1809.00000000 | 307939.00000000 | 8.20000000 | 78.60000000 | 68.00000000 | 1620.00000000 |
| 04/8 | 1896.00000000 | 300239.00000000 | 8.30000000 | 85.00000000 | 72.00000000 | 1588.00000000 |

**Series Properties**

Observations: 32

Forecasts: 12

Series: 8

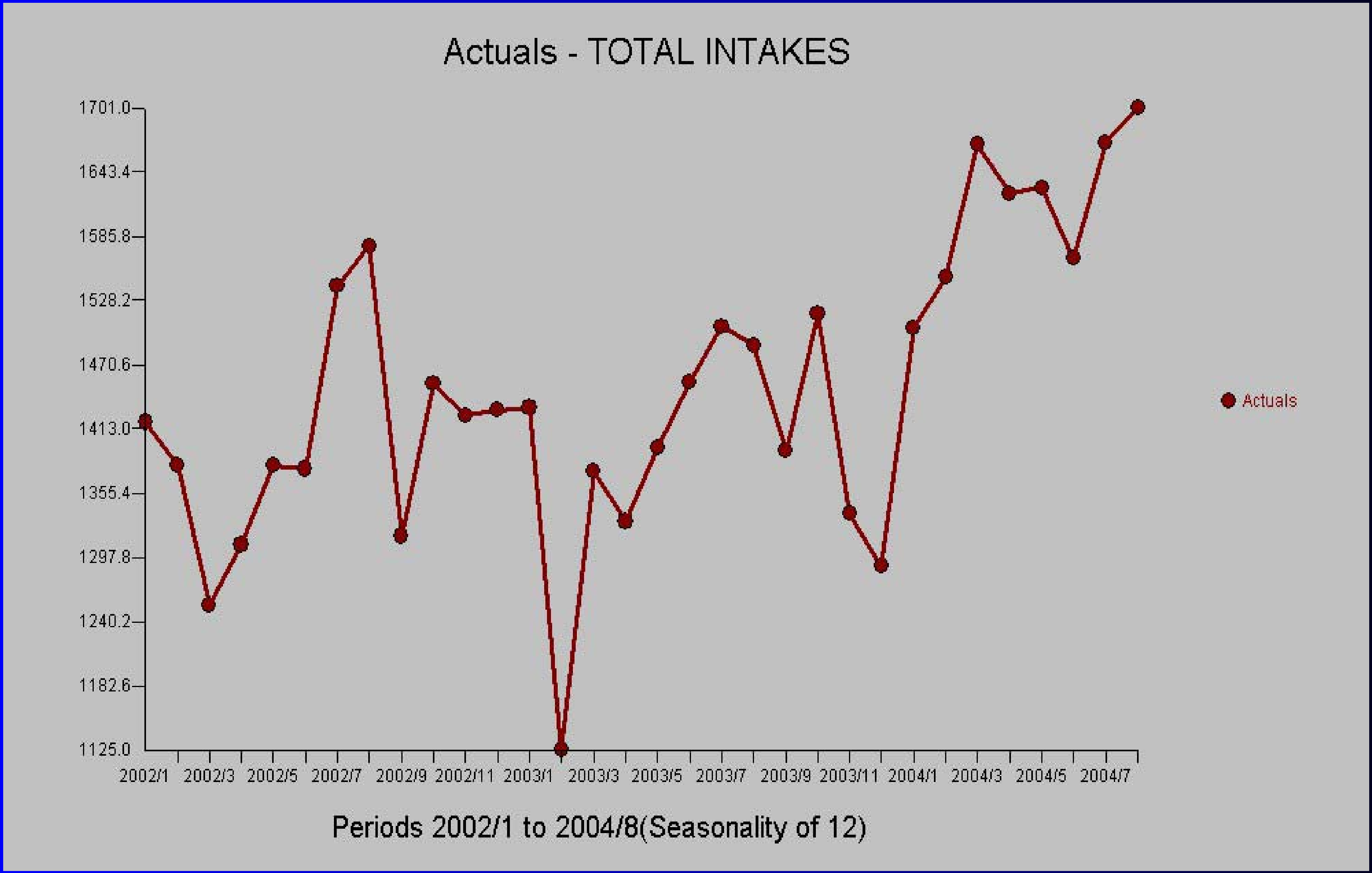
Major Period: 2002

Minor Period: 1

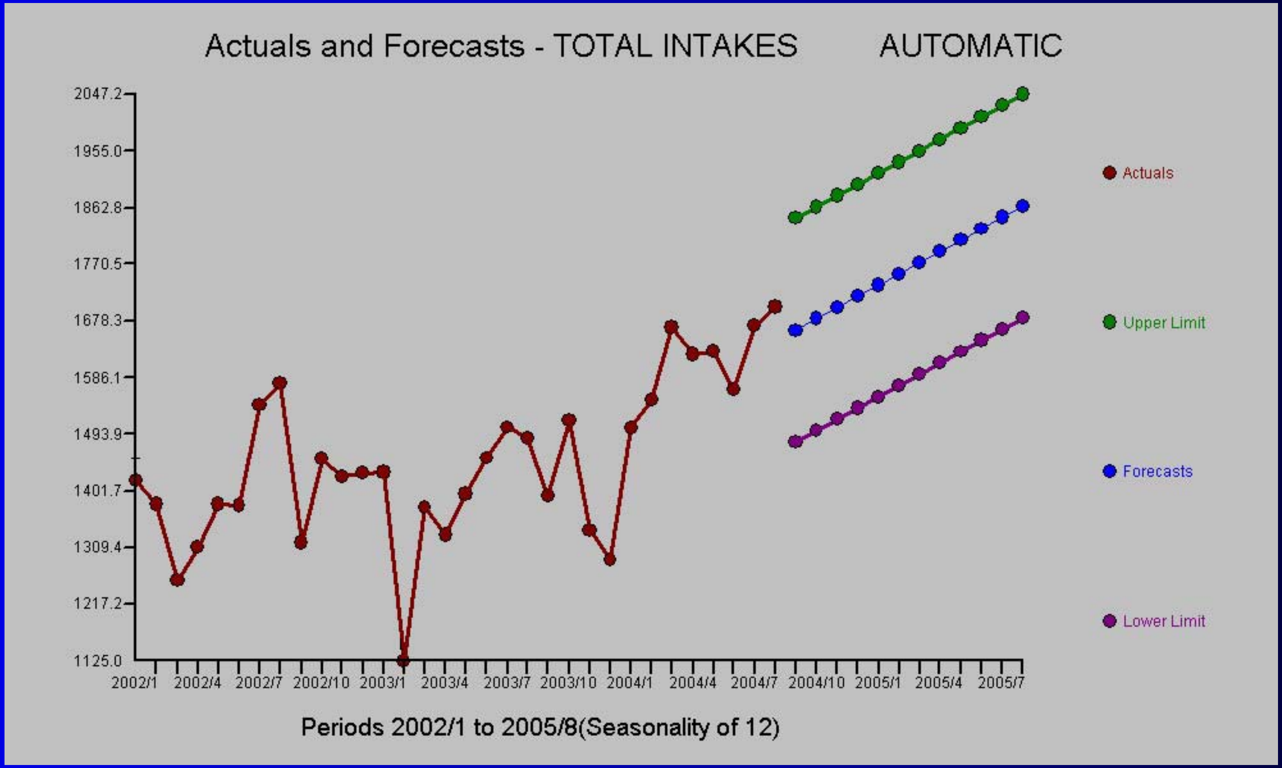
Frequency: 12

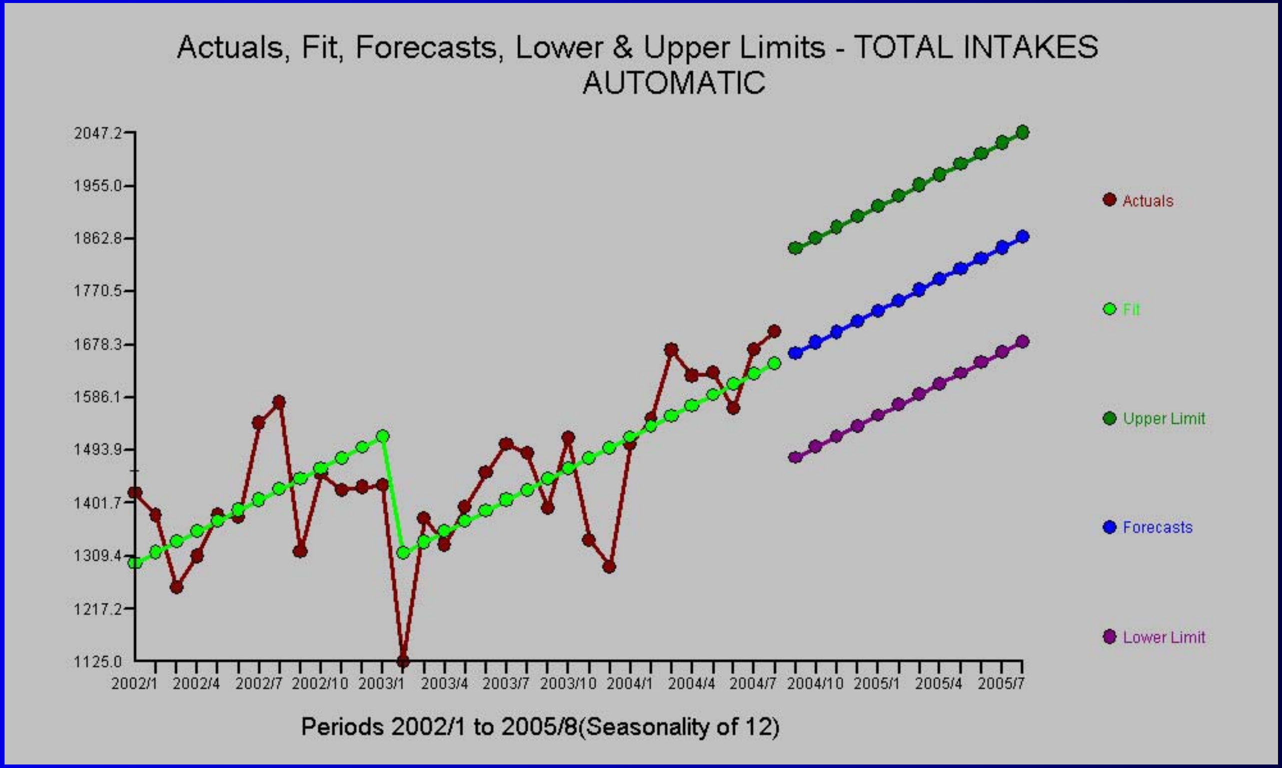
Apply Cancel

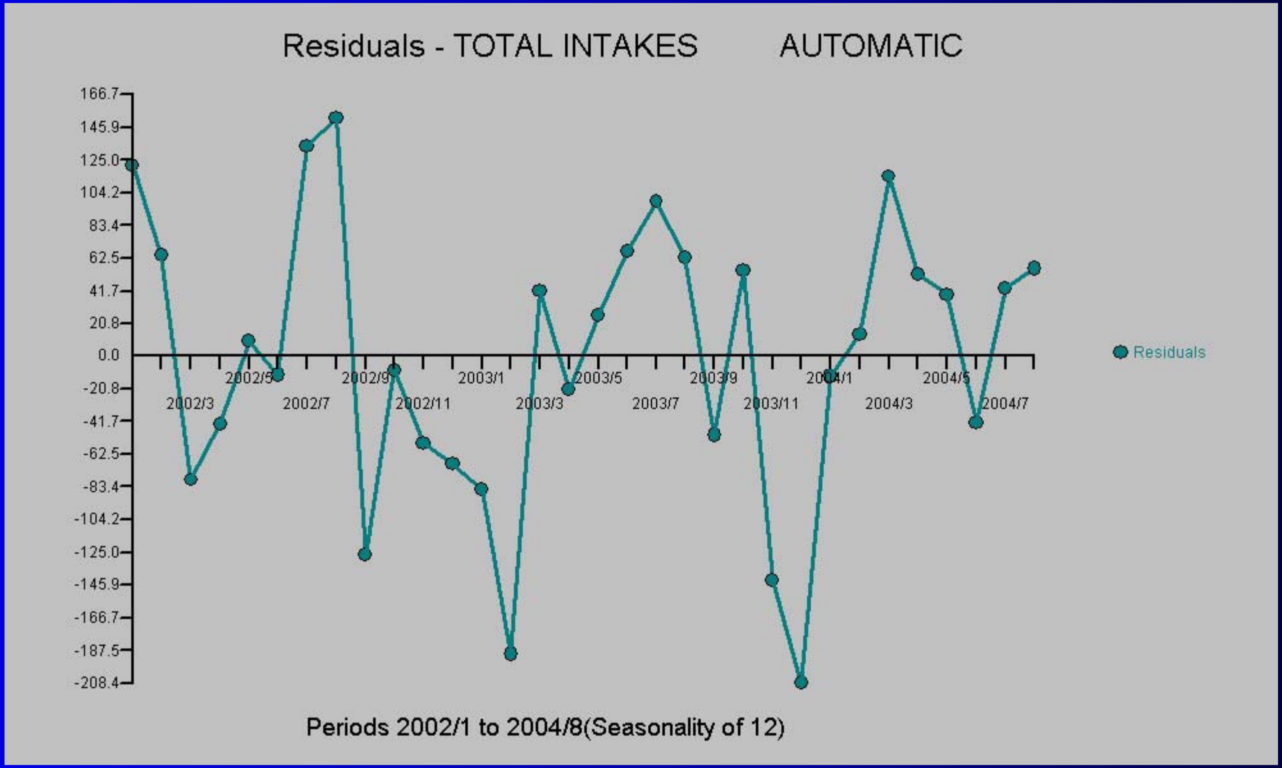
nt Status | Engine = M | 9/28/2004 | 8:39 AM



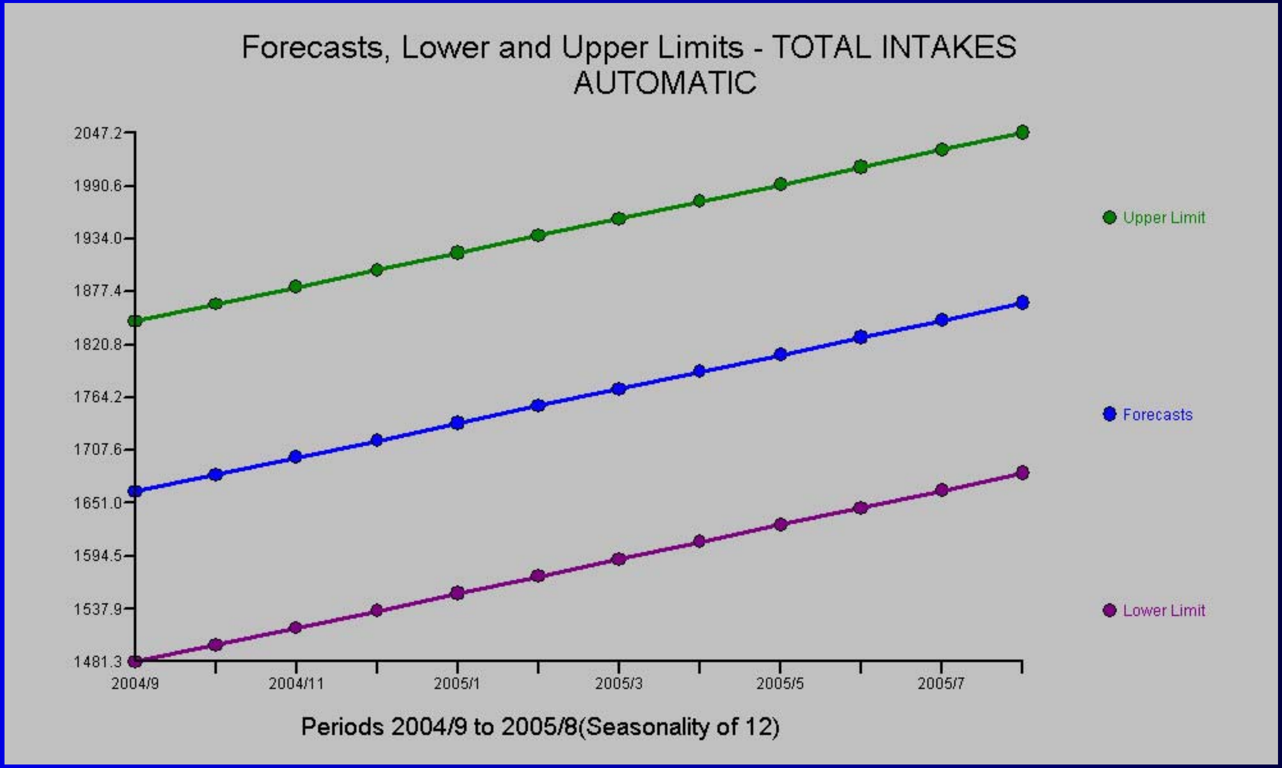
Actuals - TOTAL INTAKES
1701.0
1643.4
1585.8
1528.2
1470.6
1413.0
1355.4
1297.8
1240.2
1182.6
1125.0
2002/1
2002/3
2002/5
2002/7
2002/9
2002/11
2003/1
2003/3
2003/5
2003/7
2003/9
2003/11
2004/1
2004/3
2004/5
2004/7
Periods 2002/1 to 2004/8(Seasonality of 12)
Actuals


Actuals and Forecasts - TOTAL INTAKES AUTOMATIC
2047.2
1955.0
1862.8
1770.5
1678.3
1586.1
1493.9
1401.7
1309.4
1217.2
1125.0
2002/1 2002/4 2002/7 2002/10 2003/1 2003/4 2003/7 2003/10 2004/1 2004/4 2004/7 2004/10 2005/1 2005/4 2005/7
Periods 2002/1 to 2005/8(Seasonality of 12)
Actuals
Upper Limit
Forecasts
Lower Limit


Actuals, Fit, Forecasts, Lower & Upper Limits - TOTAL INTAKES
AUTOMATIC
2047.2
1955.0
1862.8
1770.5
1678.3
1586.1
1493.9
1401.7
1309.4
1217.2
1125.0
2002/1 2002/4 2002/7 2002/10 2003/1 2003/4 2003/7 2003/10 2004/1 2004/4 2004/7 2004/10 2005/1 2005/4 2005/7
Periods 2002/1 to 2005/8(Seasonality of 12)
Actuals
Fit
Upper Limit
Forecasts
Lower Limit

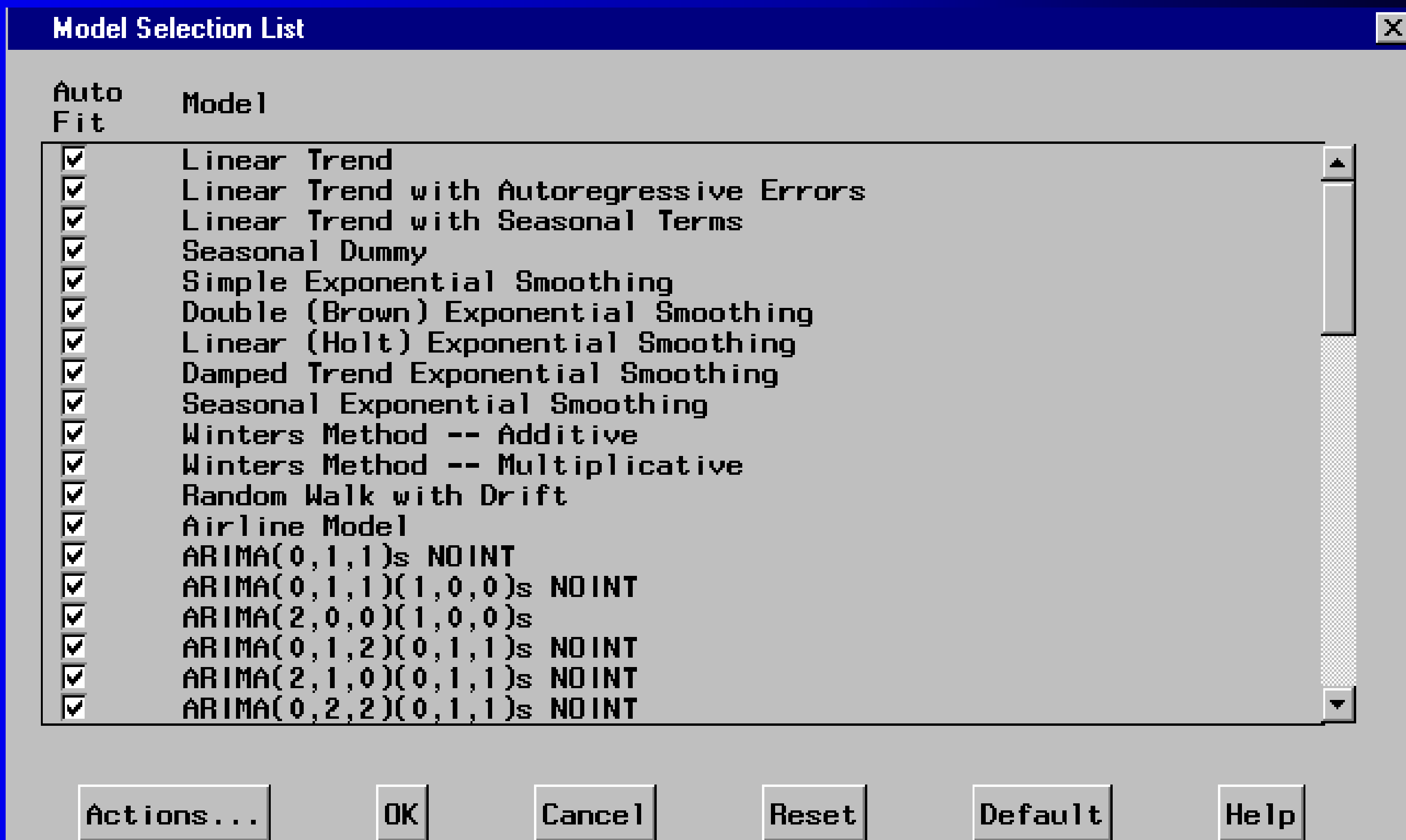


Residuals - TOTAL INTAKES AUTOMATIC
166.7
145.9
125.0
104.2
83.4
62.5
41.7
20.8
0.0
-20.8
-41.7
-62.5
-83.4
-104.2
-125.0
-145.9
-166.7
-187.5
-208.4
2002/3
2002/5
2002/7
2002/9
2002/11
2003/1
2003/3
2003/5
2003/7
2003/9
2003/11
2004/1
2004/3
2004/5
2004/7
Residuals
Periods 2002/1 to 2004/8(Seasonality of 12)

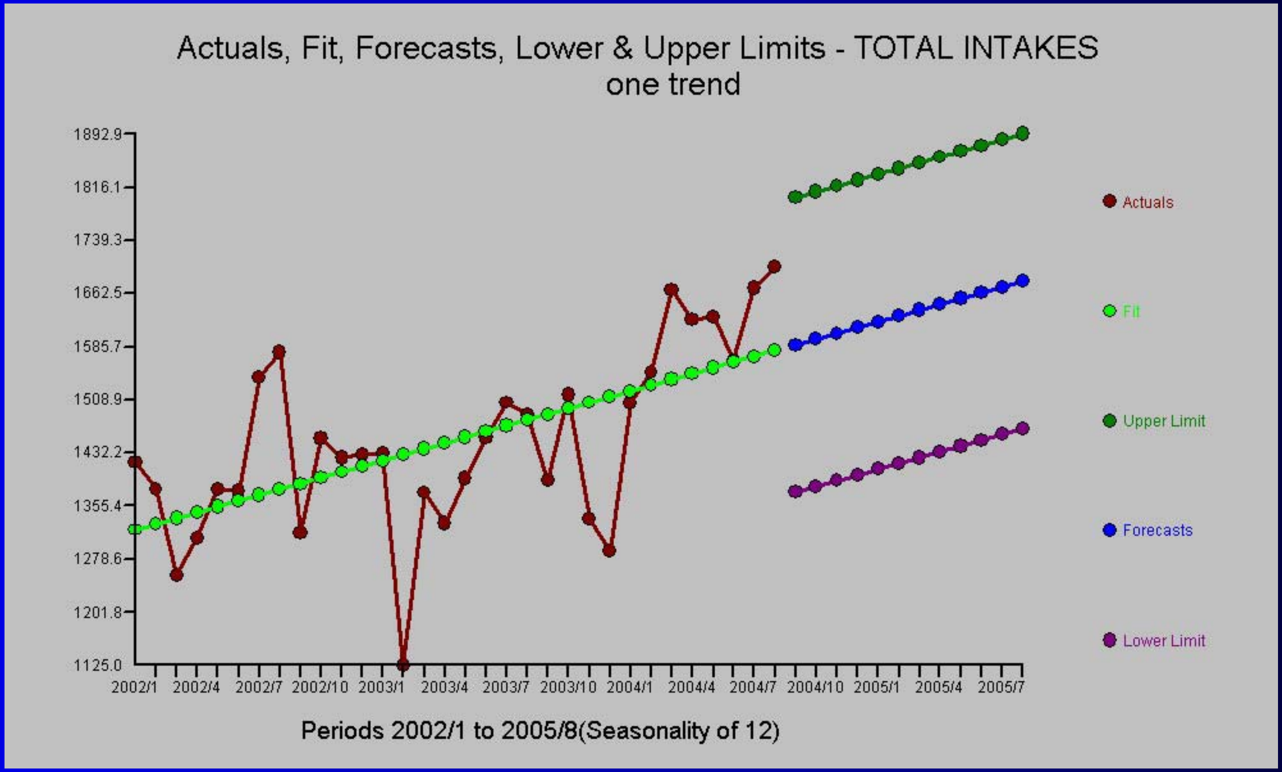

Forecasts, Lower and Upper Limits - TOTAL INTAKES
AUTOMATIC
2047.2
1990.6
1934.0
1877.4
1820.8
1764.2
1707.6
1651.0
1594.5
1537.9
1481.3
2004/9
2004/11
2005/1
2005/3
2005/5
2005/7
Periods 2004/9 to 2005/8(Seasonality of 12)
Upper Limit
Forecasts
Lower Limit

Traditional techniques assumed a “set of models” and then selected that model that was deemed the “best” based .

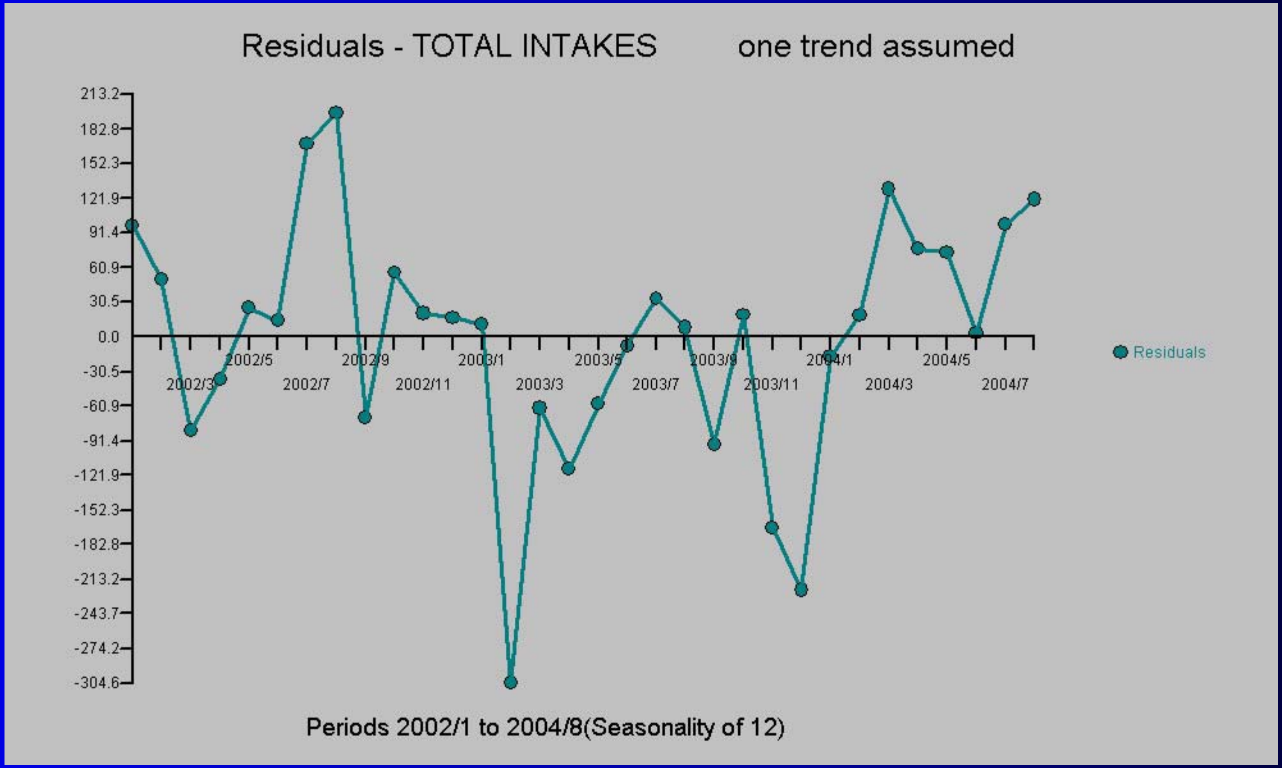
# The "Pick Best" Approached Favored by Some Software Packages



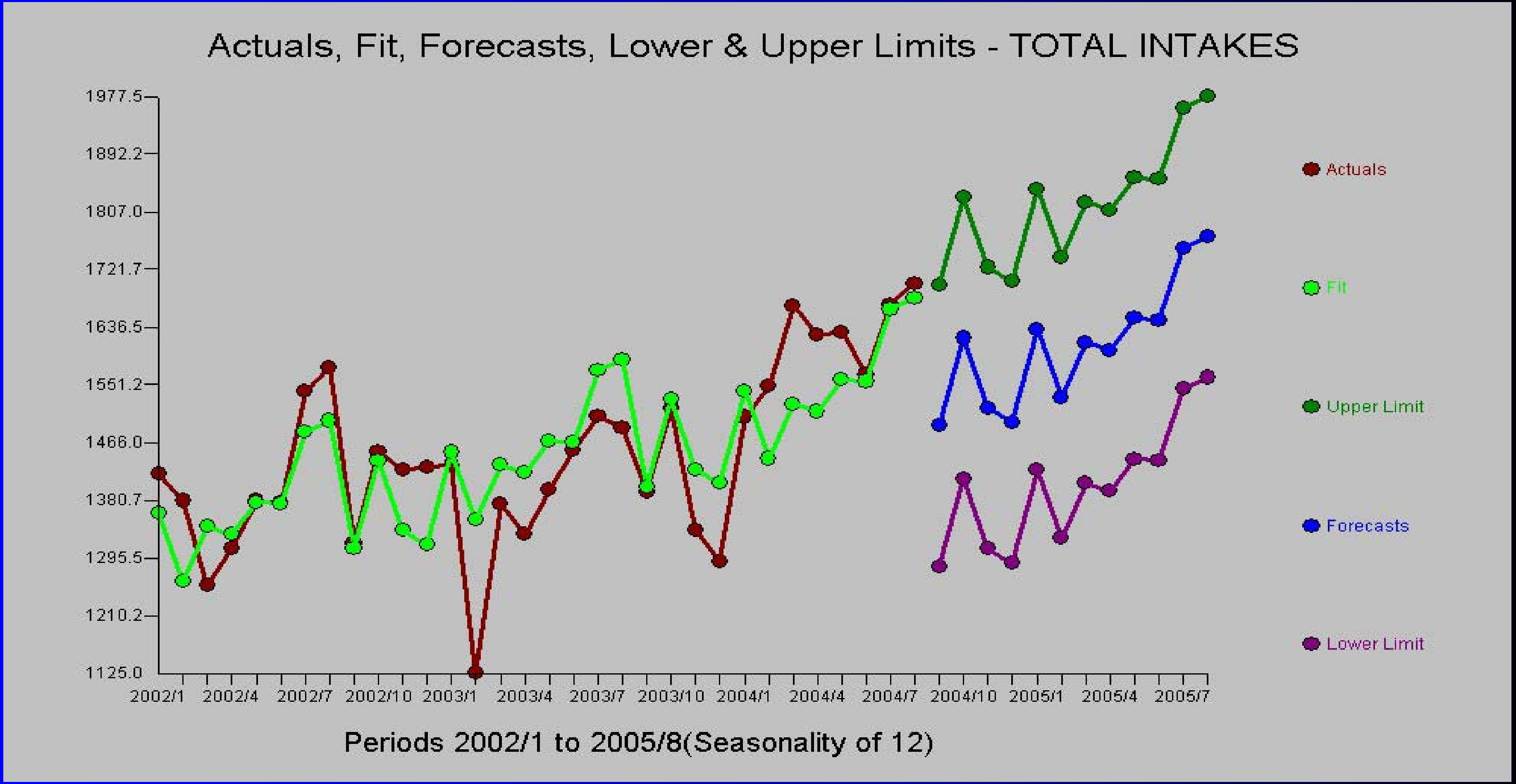
# Yields Rather Uninteresting Results



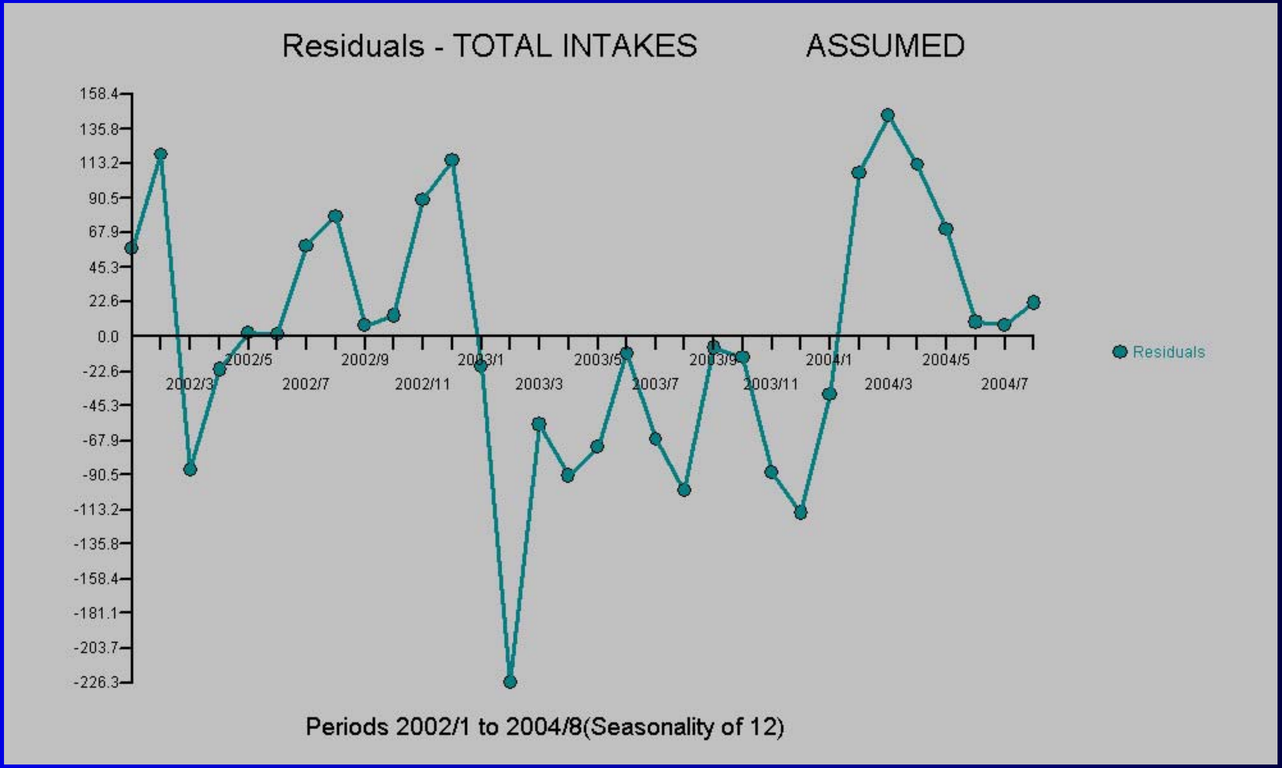
# Residuals Suggest a Poor Model



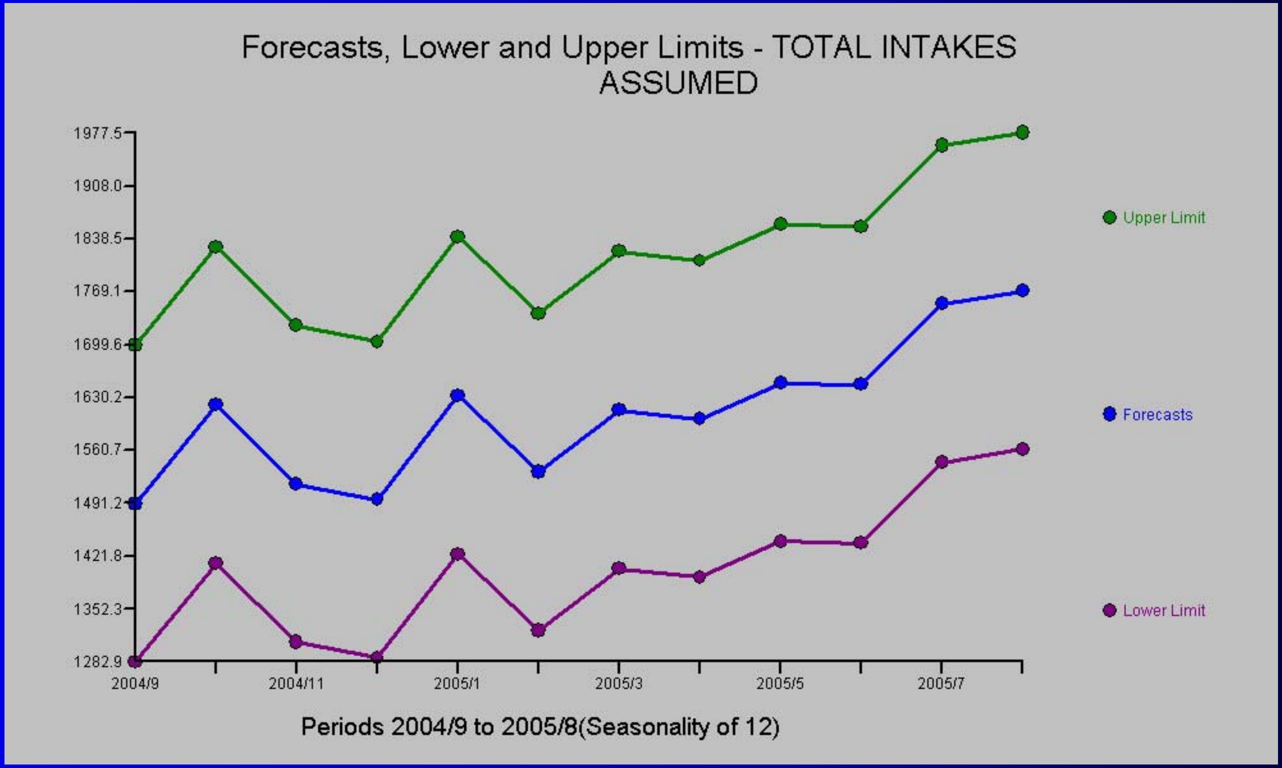
# If You Assume a Holt-Winters Model



# Residuals Suggest A Poor Model



# Forecasts Institutionalize a False Seasonality



# We wish to Model/Predict Total Intakes Using As Possible Predictor Variables:

Analysis for Variable Y = TOTAL_INTAKES
X1 = TOTALREPCRIM
X2 = TOTALCASES
X3 = LABOR_FORCE
X4 = UNEMPLOYMENT
X5 = TEMP+DEF_F
X6 = RH
X7 = TOTAL_RELEASES

# We wish to Model/Predict Total Intakes Using As Possible Predictor Variables:

Analysis for Variable Y = TOTAL_INTAKES

| | |
|---|---|
| X1 = TOTALREPCRIM | usao |
| X2 = TOTALCASES | usao |
| X3 = LABOR_FORCE | dol |
| X4 = UNEMPLOYMENT | dol |
| X5 = TEMP+DEF_F | www.noaa.gov |
| X6 = RH | www.noaa.gov |
| X7 = TOTAL_RELEASES | dc doc |

# In attempting to formulate the model Y=Xb + V, classic multiple regression assumes the following:

.V is uncorrelated

.V has no outliers

.Relationship is contemporary

$$Y(T) = 254.94 + X1(T)(.244) + X2(T)(.563) + V(T)$$

Y = TOTAL_INTAKES

X1 = TOTALCASES

X2 = TOTAL_RELEASES

R Square = .719382

# While Conducting Model Diagnostic Checking of the Assumed Model
# Y=Xb + V

## AUTOBOX found:

- .V is significantly autocorrelated
- .V has an inlier at 2003/4 ( time point 16 is 71.9 higher than it should have been)
- .There is a deterministic seasonal component for August (- 79.4)
- .Relationship is dynamic (multiple-lagged)

# Residuals Suggest a Poor Model

V(T)= A(t)+Omitted Lags of the Causals + A Seasonal Pulse + An Inlier + Autocorrelation

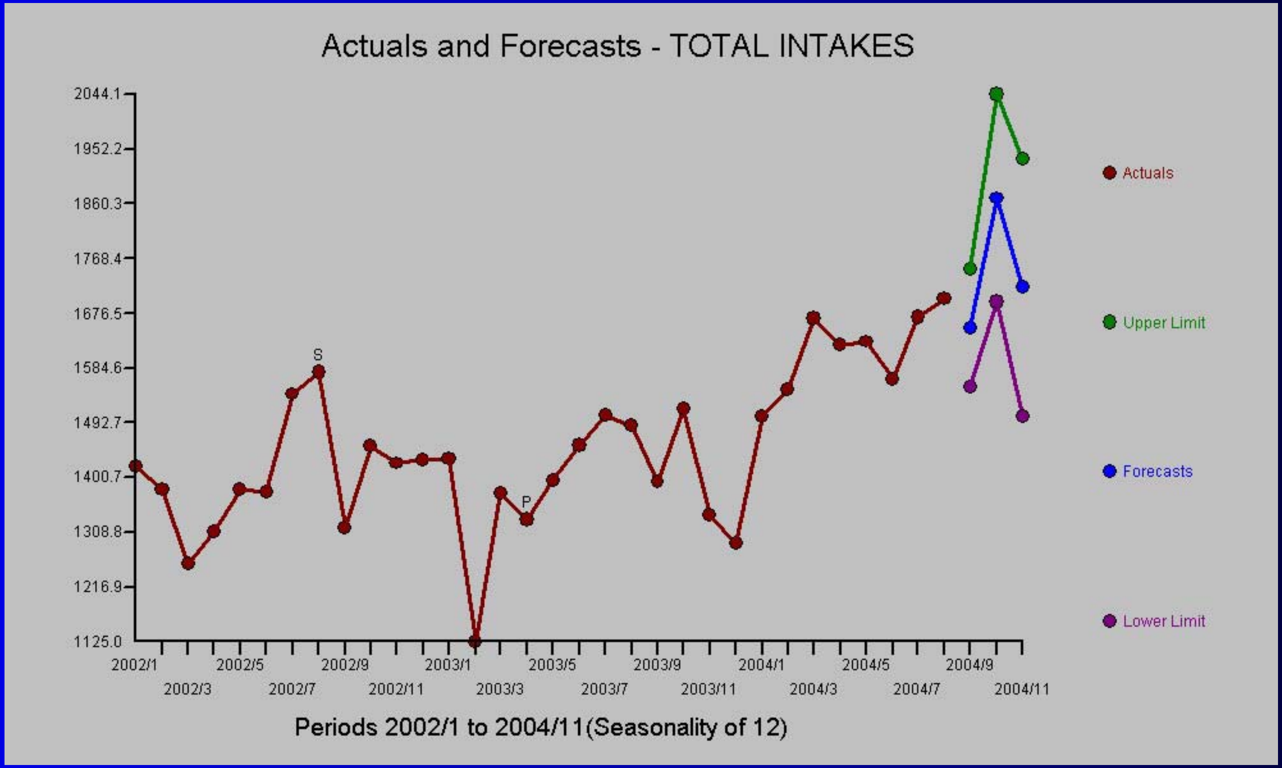
We have a case where causals were incorrectly rejected due to the inflated variance of the errors.

A Box-Jenkins approach is really a regression model with built-in diagnostic checking culminating in an efficient model.

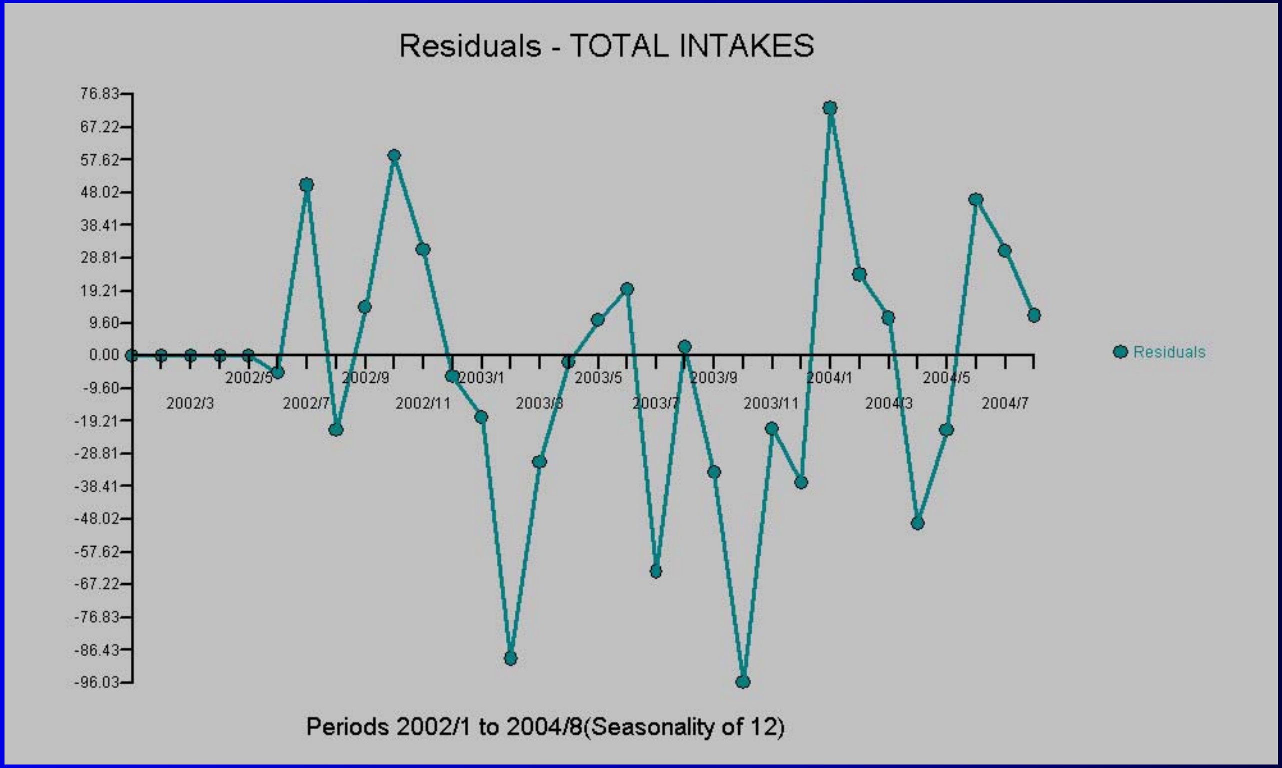
A Transfer Function Model is also called Generalized Least Squares as it incorporates both non-constant variance and correlated data opportunities.

# Combining Causals, Memory and Dummy Variables

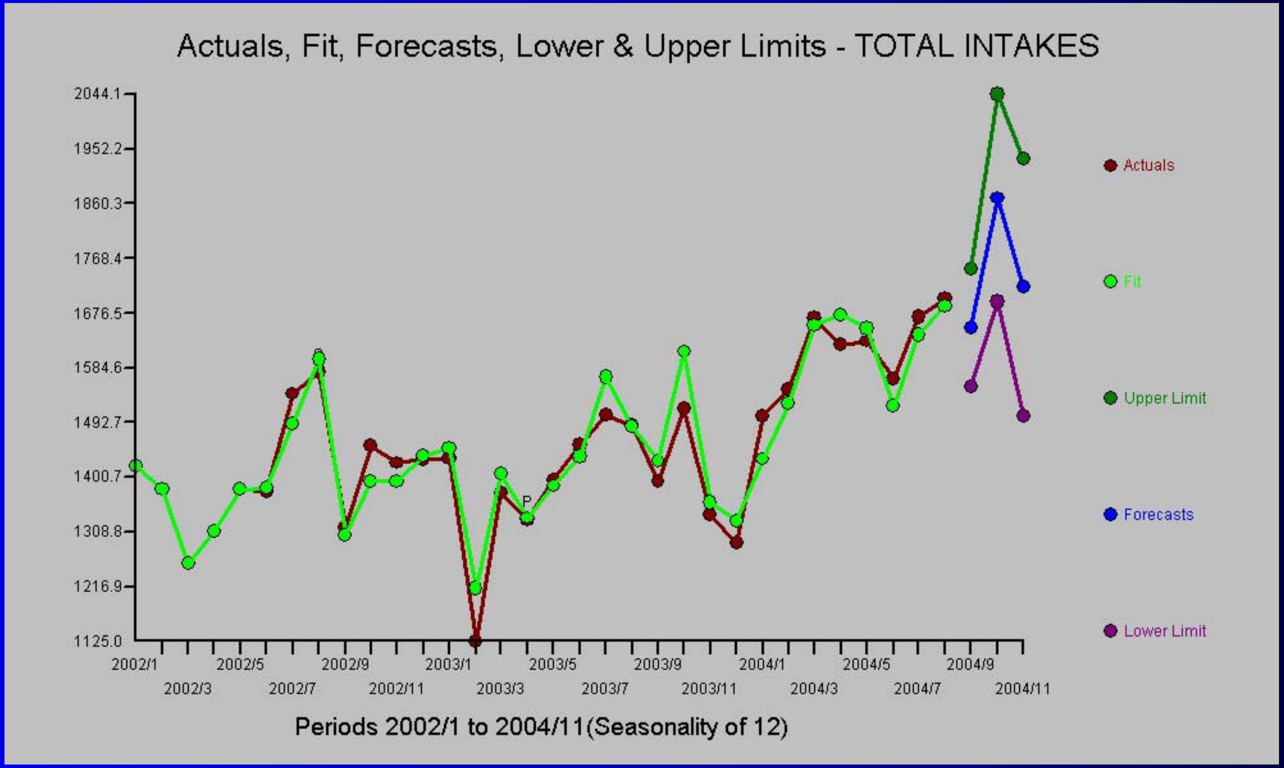
## History of Intakes and Forecasts with 95% Limits



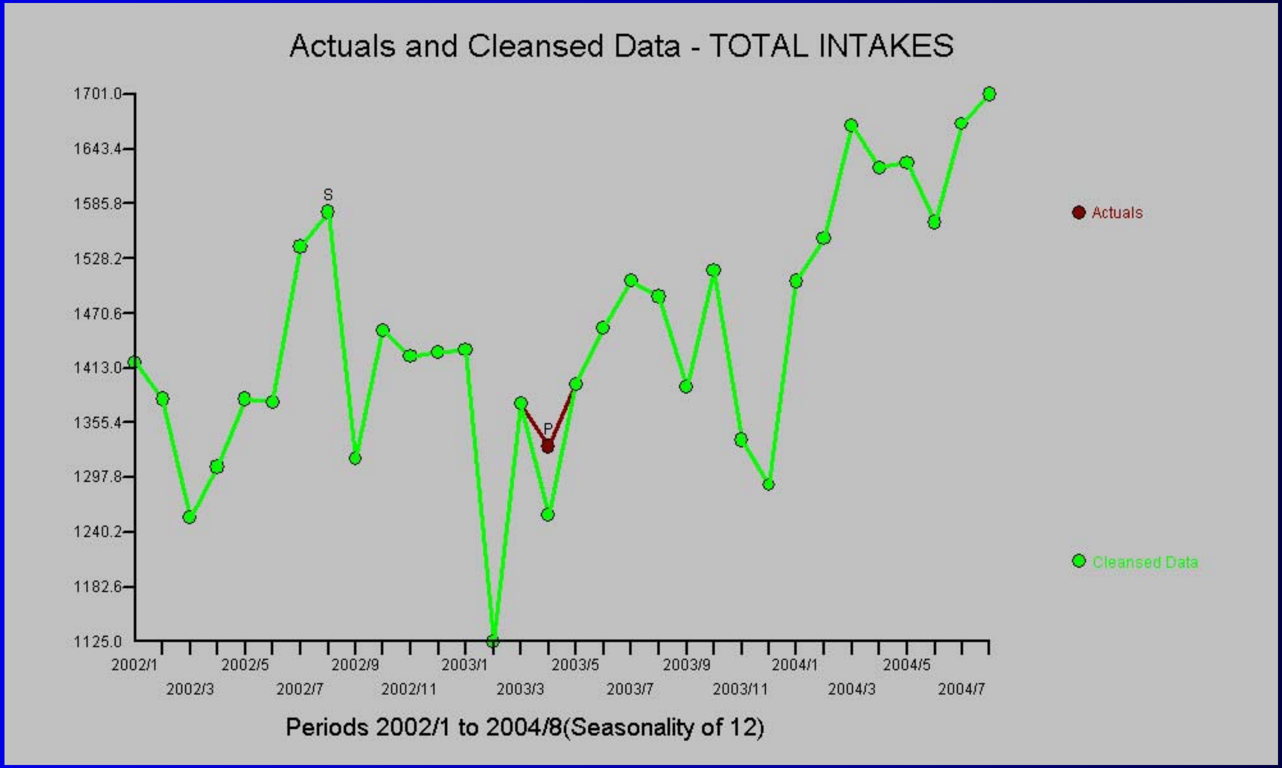
# Combining Causals, Memory and Dummy Variables



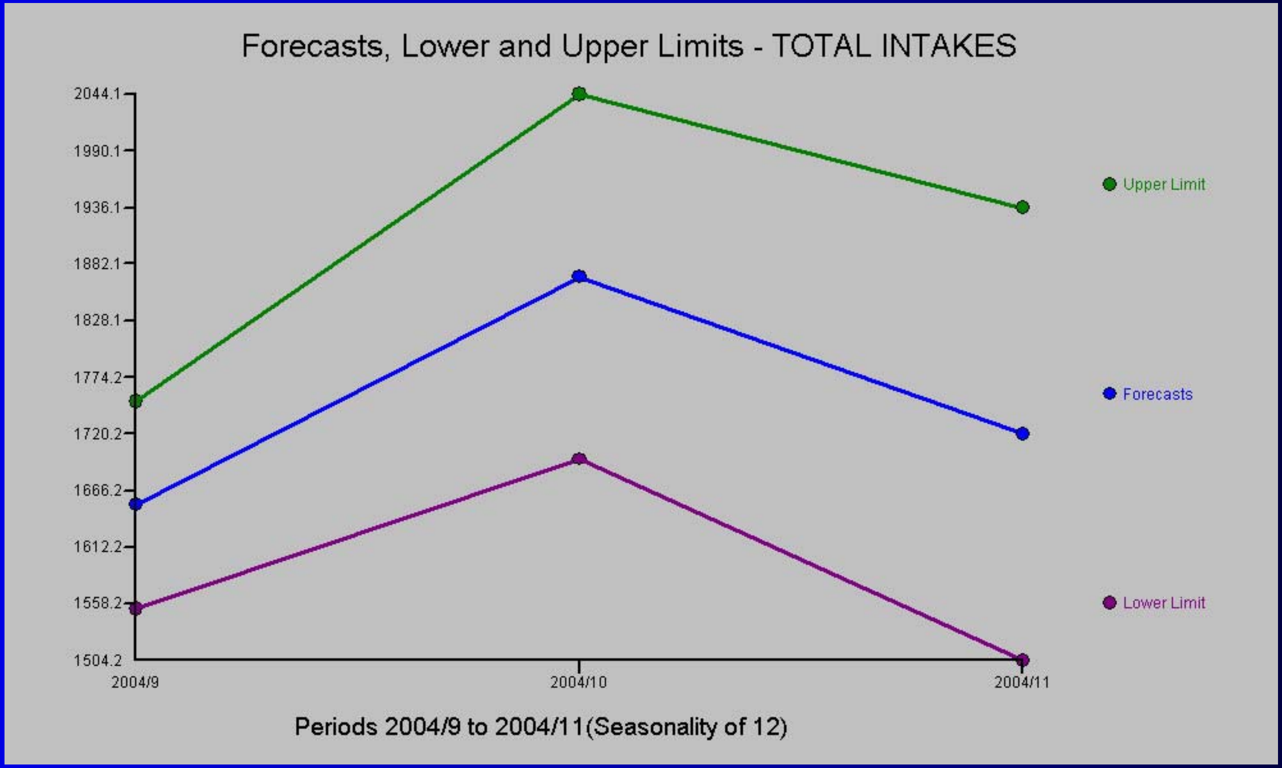
# Combining Causals, Memory and Dummy Variables



# Combining Causals, Memory and Dummy Variables



# Combining Causals, Memory and Dummy Variables



View Options Process Help

storical Data | Future Values | Forecast Data | Graph | **Reports** | WhatIf

eports
- DETAILS.HTM
- INTRVENT.HTM
- EQUATION.TXT
- VERBAL.TXT
- STAT.HTM
- RHSIDE.TXT

| VARIABLE | LAG | REGRESSION COEFFICIENT |
|---|---|---|
| TOTALREPCRIM | 0 | .109473 |
| | 1 | -.158757 |
| | 2 | .089026 |
| TOTALCASES | 0 | .385059 |
| | 1 | -.558409 |
| | 2 | .313137 |
| LABOR_FORCE | 1 | -.009094 |
| | 2 | .013188 |
| | 3 | -.007395 |
| UNEMPLOYMENT | 1 | 121.749462 |
| | 2 | -14.318887 |
| | 3 | -85.956013 |
| | 4 | 58.969877 |
| | 5 | 40.917930 |
| TEMP+DEF_F | 1 | -8.378548 |
| | 2 | 17.266206 |
| | 3 | -14.232346 |
| | 4 | 4.160186 |
| TOTAL_RELEASES | 0 | .231719 |
| | 1 | -.336037 |
| | 2 | .188439 |
| I~S00008TOTAL_INTAKES | 0 | -79.396684 |
| | 1 | 115.140462 |
| | 2 | -64.566921 |
| TOTAL_INTAKES | 1 | 1.450192 |
| | 2 | -.813219 |

nt Status | Engine = M | 10/4/2004 | 8:57 AM

View Options Process Help

storical Data | Future Values | Forecast Data | Graph | **Reports** | WhatIf

eports
- DETAILS.HTM
- INTRVENT.HTM
- EQUATION.TXT
- VERBAL.TXT
- STAT.HTM
- RHSIDE.TXT

```
MODEL STATISTICS AND EQUATION FOR THE CURRENT EQUATION (DETAILS FOLLOW).

Estimation/Diagnostic Checking for Variable Y  = TOTAL_INTAKES
                                            X1 = TOTALREPCRIM
                                            X2 = TOTALCASES
                                            X3 = LABOR_FORCE
                                            X4 = UNEMPLOYMENT
                                            X5 = TEMP+DEF_F
                                            X6 = TOTAL_RELEASES
          : NEWLY IDENTIFIED VARIABLE      X7 = I~S00008 2002/  8    SEASP
          : NEWLY IDENTIFIED VARIABLE      X8 = I~P00016 2003/  4    PULSE



           MODEL STATISTICS IN TERMS OF THE ORIGINAL DATA

      Number of Residuals (R)          =n                  27
      Number of Degrees of Freedom     =n-m                13
      Residual Mean                    =Sum R / n      -4.23168
      Sum of Squares                   =Sum R**2        45754.3
      Variance                      var=SOS/(n)         1694.60
      Adjusted Variance                =SOS/(n-m)       3519.56
      Standard Deviation               =SQRT(Adj Var)   59.3259
      Standard Error of the Mean       =Standard Dev/   16.4540
      Mean / its Standard Error        =Mean/SEM       -.257182
      Mean Absolute Deviation          =Sum(ABS(R))/n   32.5775
      AIC Value ( Uses var )           =nln  +2m        228.750
      SBC Value ( Uses var )           =nln  +m*lnn     246.892
      BIC Value ( Uses var )           =see Wei p153    195.943
      R Square                      =                   .901626
      Durbin-Watson Statistic       =[A-A(T-1)]**2/A**2 1.52442

    D-W STATISTIC IS INCONCLUSIVE.

 THE DURBIN-WATSON STATISTIC IS VALID ONLY FOR MODELS THAT HAVE NO ARIMA
 COMPONENT AND NO LAGS OF THE Y SERIES OTHERWISE IT IS INVALID.
 IN THIS CASE THE TEST IS INVALID.
```

nt Status | Engine = M | 10/4/2004 | 8:58 AM

View Options Process Help

storical Data | Future Values | Forecast Data | Graph | **Reports** | WhatIf

eports
- DETAILS.HTM
- INTRVENT.HTM
- EQUATION.TXT
- VERBAL.TXT
- STAT.HTM
- RHSIDE.TXT

```
  1CONSTANT                          317.        280.        .2717      1.13
  2Autoregressive-Factor #  1    1   1.45        .998E-01    .0000     14.53
  3                              2  -.813        .982E-01    .0000     -8.28

INPUT SERIES X1   TOTALREPCRIM

  4Omega (input) -Factor #  2    0   .109        .333E-01    .0041      3.29

INPUT SERIES X2   TOTALCASES

  5Omega (input) -Factor #  3    0   .385        .492E-01    .0000      7.83

INPUT SERIES X3   LABOR_FORCE

  6Omega (input) -Factor #  4    1  -.909E-02    .287E-02    .0054     -3.17

INPUT SERIES X4   UNEMPLOYMENT

  7Omega (input) -Factor #  5    1   122.        25.2        .0001      4.83
  8                              2  -162.        34.8        .0002     -4.66
  9                              3  -50.3        19.6        .0195     -2.56

INPUT SERIES X5   TEMP+DEF_F

 10Omega (input) -Factor #  6    1  -8.38        2.24        .0015     -3.74
 11                              2  -5.12        2.41        .0479     -2.12

INPUT SERIES X6   TOTAL_RELEASES

 12Omega (input) -Factor #  7    0   .232        .638E-01    .0019      3.63

INPUT SERIES X7   I~S00008 2002/  8    SEASP

 13Omega (input) -Factor #  8    0  -79.4        24.8        .0049     -3.21

INPUT SERIES X8   I~P00016 2003/  4    PULSE

 14Omega (input) -Factor #  9    0   71.9        31.2        .0335      2.30
```

nt Status | Engine = M | 10/4/2004 | 8:59 AM

View Options Process Help

storical Data | **Future Values** | Forecast Data | Graph | Reports | WhatIf

| | TOTALREPCRIM | TOTALCASES | LABOR FORCE | UNEMPLOYMEN | TEMP+DEF F | RH | TOTAL RELEASES |
|---|---|---|---|---|---|---|---|
| 04/9 | 2600.00000000 | 1300.00000000 | 291100.00000000 | 7.58600000 | 73.39000000 | 71.00000000 | 1418.00000000 |
| 04/10 | 2700.00000000 | 1400.00000000 | 291900.00000000 | 7.05500000 | 61.14000000 | 70.00000000 | 1380.00000000 |
| 04/11 | 2700.00000000 | 1350.00000000 | 302100.00000000 | 6.68400000 | 57.32000000 | 66.00000000 | 1362.00000000 |

nt Status | Engine = M | 10/4/2004 | 9:00 AM

View Options Process Help

storical Data | Future Values | **Forecast Data** | Graph | Reports | WhatIf

| | TOTAL_INTAKES |
|---|---|
| 2004/5 | 1652.27770977 |
| 2004/6 | 1870.05793692 |
| 2004/7 | 1719.92521075 |

nt Status | Engine = M | 10/4/2004 | 9:00 AM

# A Forecasting Model is a Planning Tool Not Just An End In Itself !

# Impact Assessment <> What if Unemployment Slowly Rises?

CAUSAL3.ASC FreeFore Professional Build: 0.1.30

File View Options Process Help

Run
RunWhatIf

Historical Data | F | Forecast Data | Graph | Reports | WhatIf

| | TOTALREPCRIM | TOTALCASES | LABOR FORCE | UNEMPLOYMEN | TEMP+DEF F | RH | TOTAL RELEASES |
|---|---|---|---|---|---|---|---|
| 2004/9 | 2600.00000000 | 1300.00000000 | 291100.00000000 | 7.80000000 | 73.39000000 | 71.00000000 | 1418.00000000 |
| 2004/10 | 2700.00000000 | 1400.00000000 | 291900.00000000 | 7.90000000 | 61.14000000 | 70.00000000 | 1380.00000000 |
| 2004/11 | 2700.00000000 | 1350.00000000 | 302100.00000000 | 8.00000000 | 57.32000000 | 66.00000000 | 1362.00000000 |

View Options Process Help

storical Data | Future Values | **Forecast Data** | Graph | Reports | WhatIf

| | TOTAL INTAKES |
|---|---|
| 2004/9 | 1652.27770977 |
| 2004/10 | 1896.11232177 |
| 2004/11 | 1857.52313329 |

nt Status | Engine = M | 10/4/2004 | 9:57 AM

SCENARIO

| | 1 | | 2 | |
|---|---|---|---|---|
| | UNEMPLOYMENT | INTAKES | UNEMPLOYMENT | INTAKES |
| 2004/9 | 7.586 | 1652 | 7.8 | 1652 |
| 2004/10 | 7.055 | 1870 | 7.9 | 1896 |
| 2004/11 | 6.684 | 1719 | 8 | 1857 |

# A Memory Component in a Causal Model is a Proxy for an Omitted Variable

If Y(T) = [ X(T) ] + g[ Z(T) ] + A(T) and you omit Z(T) then
Y(T) = [ X(T) ] + V(T) where V(T)= g[ Z(T) ] + A(T) .

If Z(T) is an auto-projective sequence then V(T) will be auto-projective and thus auto-correlated yielding V(T)= [T(B)/P(B)] A(T )

Y(T) = X(T ) + [T(B)/P(B)] A(T )

- where {T(B)/P(B)}=ARMA model for unobserved series Z(T)

# Intervention Analysis/AIA References


Box, G.E.P., and Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control, 2nd ed. San Francisco: Holden Day.

Box, G.E.P., and Tiao, G. (1975). "Intervention Analysis with Applications to Economic and Environmental Problems," Journal of the American Statistical Association, Vol 70, pp. 70-79.

Chang, I., and Tiao, G.C. (1983). "Estimation of Time Series Parameters in the Presence of Outliers," Technical Report #8, Statistics Research Center, Graduate School of Business, University of Chicago, Chicago.

McCleary, R., and Hay, R. (1980). Applied Time Series Analysis for the Social Sciences. Los Angeles: Sage.

# Intervention Analysis/AIA References

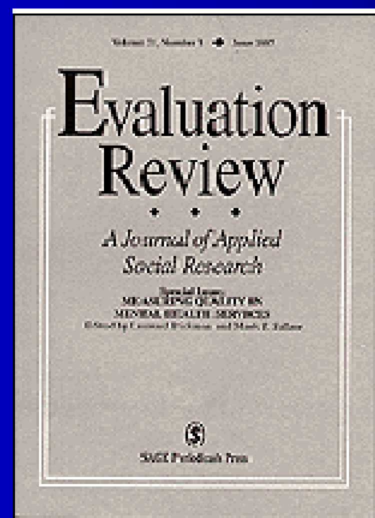

Reilly, D.P. (1980). "Experiences with an Automatic Box-Jenkins Modeling Algorithm," in Time Series Analysis, ed. O.D. Anderson. (Amsterdam: North-Holland), pp. 493-508.

Reilly, D.P. (1987). "Experiences with an Automatic Transfer Function Algorithm," in Computer Science and Statistics Proceedings of the 19th Symposium on the Interface, ed. R.M. Heiberger, (Alexandria, VI: American Statistical Association), pp. 128-135.

Tsay, R.S. (1986). "Time Series Model Specification in the Presence of Outliers," Journal of the American Statistical Society, Vol. 81, pp. 132-141.

Wei, W. (1989). Time Series Analysis Univariate and Multivariate Methods. Redwood City: Addison Wesley.

# Boston Armed Robbery References



# Boston Armed Robbery References


Campbell, D.T. and J.C. Stanley (1966), Experimental and Quasi-Experimental Designs For Research. Chicago: Rand-McNally

Chang, I., and Tiao, G.C. (1983). "Estimation of Time Series Parameters in the Presence of Outliers," Technical Report #8, Statistics Research Center, Graduate School of Business, University of Chicago, Chicago.

Deutsch, S.J. and F.B. Alt (1977) ," The Effect of Massachusetts gun control law on gun-related crimes in the city of Boston" Evaluation Quarterly, 1, 543-568.

McCleary, R., and Hay, R. (1980). Applied Time Series Analysis for the Social Sciences. Los Angeles: Sage.

Reilly, D.P. (1980). "Experiences with an Automatic Box-Jenkins Modeling Algorithm," in Time Series Analysis, ed. O.D. Anderson. (Amsterdam: North-Holland), pp. 493-508.

Reilly, D.P. (1987). "Experiences with an Automatic Transfer Function Algorithm," in Computer Science and Statistics Proceedings of the 19th Symposium on the Interface, ed. R.M. Heiberger, (Alexandria, VI: American Statistical Association), pp. 128-135.

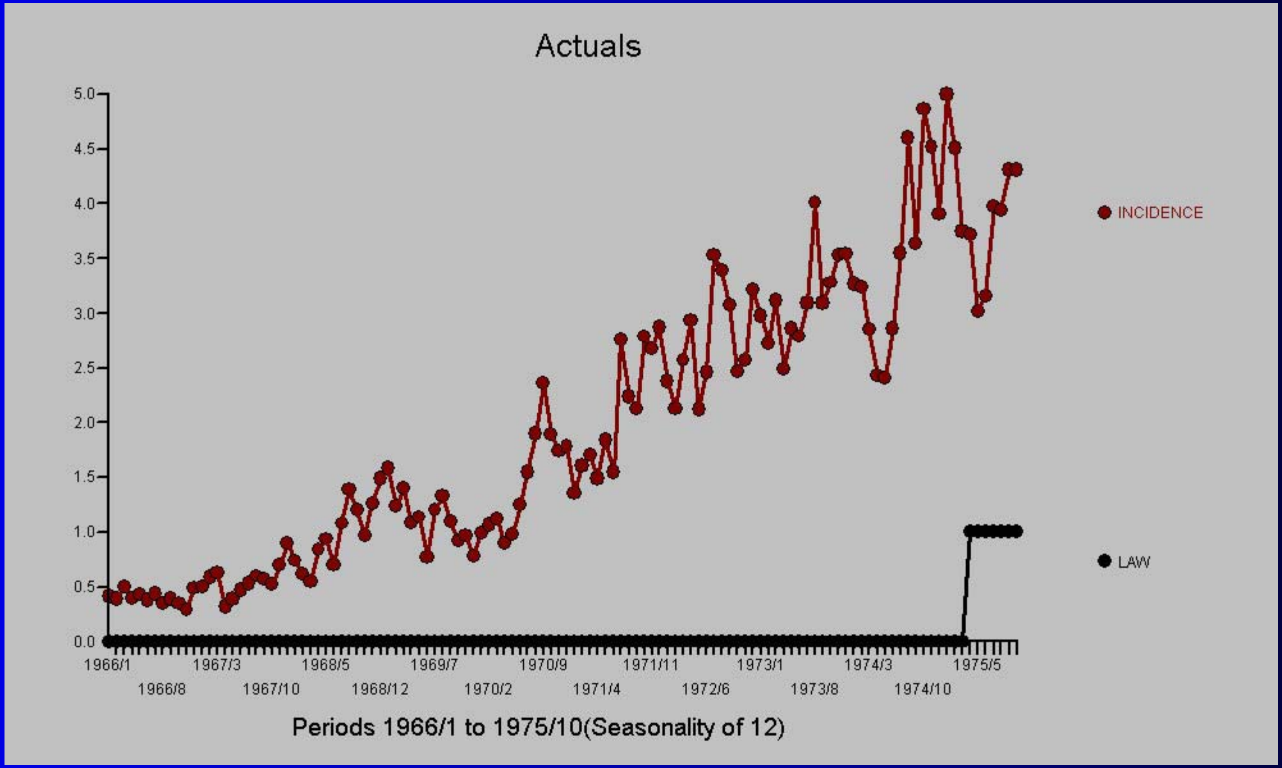
Tsay, R.S. (1986). "Time Series Model Specification in the Presence of Outliers," Journal of the American Statistical Society, Vol. 81, pp. 132-141.

Wei, W. (1989). Time Series Analysis Univariate and Multivariate Methods. Redwood City: Addison Wesley.

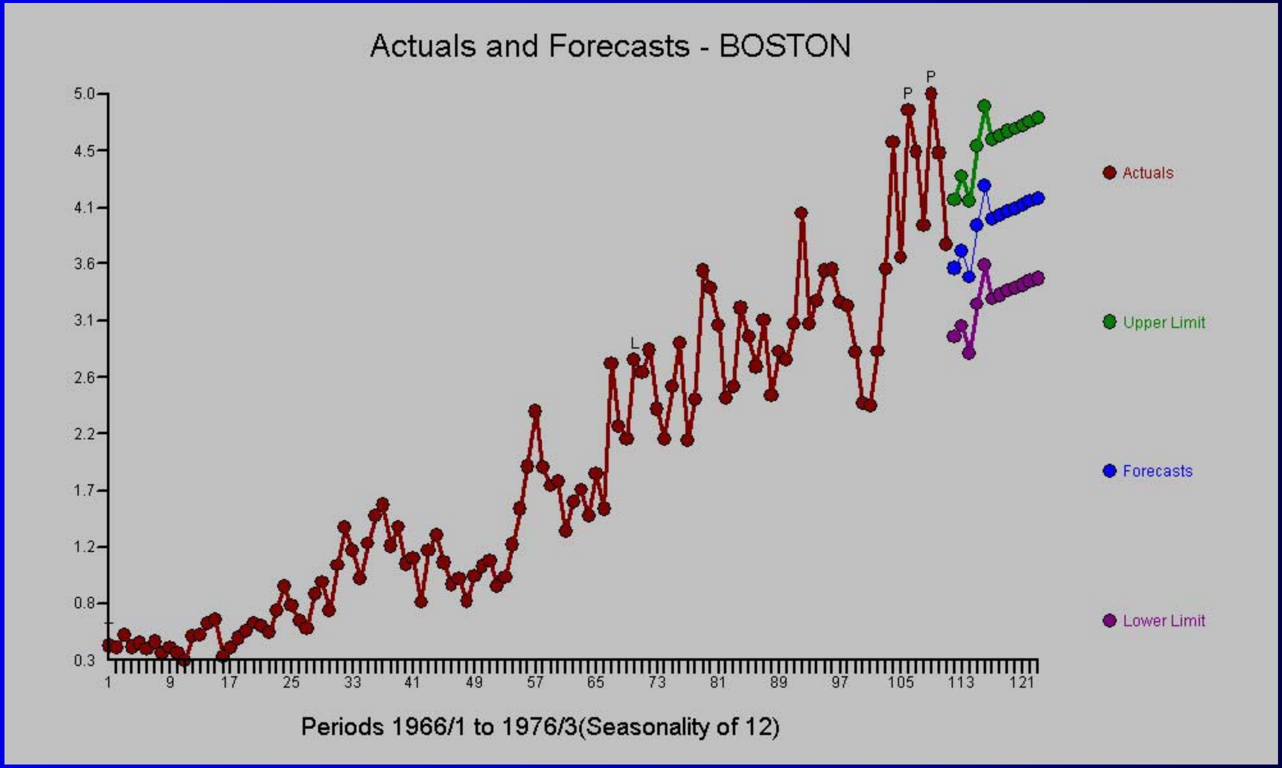
## Did Getting Tough On Boston Criminals Pay ? *

- In April of 1975, Massachusetts enacted a gun control law relating to armed robbery. It is natural to want to asses the impact of the law on the incidence of armed robberies in different geographical areas.
- One approach is to use historical data on Boston Armed Robberies from March, 1972 to March, 1975 (111 values) to develop a forecast and then to compare the actual for the next set of periods (7) up to and including October 1975 (period 118).
- We present the history , forecasts and a comparison.


- * with acknowledgement to Dr.William Sabol



Actuals
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1966/1
1967/3
1968/5
1969/7
1970/9
1971/11
1973/1
1974/3
1975/5
1966/8
1967/10
1968/12
1970/2
1971/4
1972/6
1973/8
1974/10
Periods 1966/1 to 1975/10(Seasonality of 12)
INCIDENCE
LAW


Actuals and Forecasts - BOSTON
5.0
4.5
4.1
3.6
3.1
2.6
2.2
1.7
1.2
0.8
0.3
1
9
17
25
33
41
49
57
65
73
81
89
97
105
113
121
P
P
L
Actuals
Upper Limit
Forecasts
Lower Limit
Periods 1966/1 to 1976/3(Seasonality of 12)

Edit View Insert Format Tools Data Window Help Acrobat

1975/OCT

Book1

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| | | | | ACTUALS | | FORECASTS | |
| | 112 | 1975/APRIL | | 3.72 | | 3.552649 | |
| | 113 | 1975/MAY | | 3.02 | | 3.695066 | |
| | 114 | 1975/JUNE | | 3.16 | | 3.477021 | |
| | 115 | 1975/JULY | | 3.98 | | 3.911189 | |
| | 116 | 1975/AUG | | 3.94 | | 4.237743 | |
| | 117 | 1975/SEP | | 4.31 | | 3.958033 | |
| | 118 | 1975/OCT | | 4.31 | | 3.992282 | |
| | | | | 26.44 | | 26.8 | |

Sheet1 Sheet2 Sheet3

CAPS

## Did Getting Tough On Boston Criminals Pay ? *

- Another approach , leading directly to a formal statistical test is to explicitely introduce as a supporting variable an Event Variable reflecting the known introduction of the law.he impact of the law on the incidence of armed robberies.

$$Z_t = 0,0,0,0,,,,,1,1,1,1,1,,,,,,T$$

$$\text{or } Z_t = 0 \quad t < i$$

$$Z_t = 1 \quad t > i\text{-}1$$

N.B. that this is not Intervention Detection as the variable is known and not detected

View Options Process Help

torical Data | Future Values | Forecast Data | Graph | Reports

| | INCIDENCE | LAW |
|---|---|---|
| 72/10 | 2.47000000 | 0.00000000 |
| 72/11 | 2.57000000 | 0.00000000 |
| 72/12 | 3.22000000 | 0.00000000 |
| 73/1 | 2.98000000 | 0.00000000 |
| 73/2 | 2.73000000 | 0.00000000 |
| 73/3 | 3.12000000 | 0.00000000 |
| 73/4 | 2.49000000 | 0.00000000 |
| 73/5 | 2.86000000 | 0.00000000 |
| 73/6 | 2.79000000 | 0.00000000 |
| 73/7 | 3.09000000 | 0.00000000 |
| 73/8 | 4.01000000 | 0.00000000 |
| 73/9 | 3.09000000 | 0.00000000 |
| 73/10 | 3.28000000 | 0.00000000 |
| 73/11 | 3.53000000 | 0.00000000 |
| 73/12 | 3.54000000 | 0.00000000 |
| 74/1 | 3.27000000 | 0.00000000 |
| 74/2 | 3.24000000 | 0.00000000 |
| 74/3 | 2.85000000 | 0.00000000 |
| 74/4 | 2.43000000 | 0.00000000 |
| 74/5 | 2.41000000 | 0.00000000 |
| 74/6 | 2.86000000 | 0.00000000 |
| 74/7 | 3.55000000 | 0.00000000 |
| 74/8 | 4.60000000 | 0.00000000 |
| 74/9 | 3.64000000 | 0.00000000 |
| 74/10 | 4.87000000 | 0.00000000 |
| 74/11 | 4.52000000 | 0.00000000 |
| 74/12 | 3.91000000 | 0.00000000 |
| 75/1 | 5.00000000 | 0.00000000 |
| 75/2 | 4.51000000 | 0.00000000 |
| 75/3 | 3.75000000 | 0.00000000 |
| 75/4 | 3.72000000 | 1.00000000 |
| 75/5 | 3.02000000 | 1.00000000 |
| 75/6 | 3.16000000 | 1.00000000 |
| 75/7 | 3.98000000 | 1.00000000 |
| 75/8 | 3.94000000 | 1.00000000 |
| 75/9 | 4.31000000 | 1.00000000 |
| 75/10 | 4.31000000 | 1.00000000 |

**Series Properties**

Observations: 118

Forecasts: 12

Series: 2

Major Period: 1966

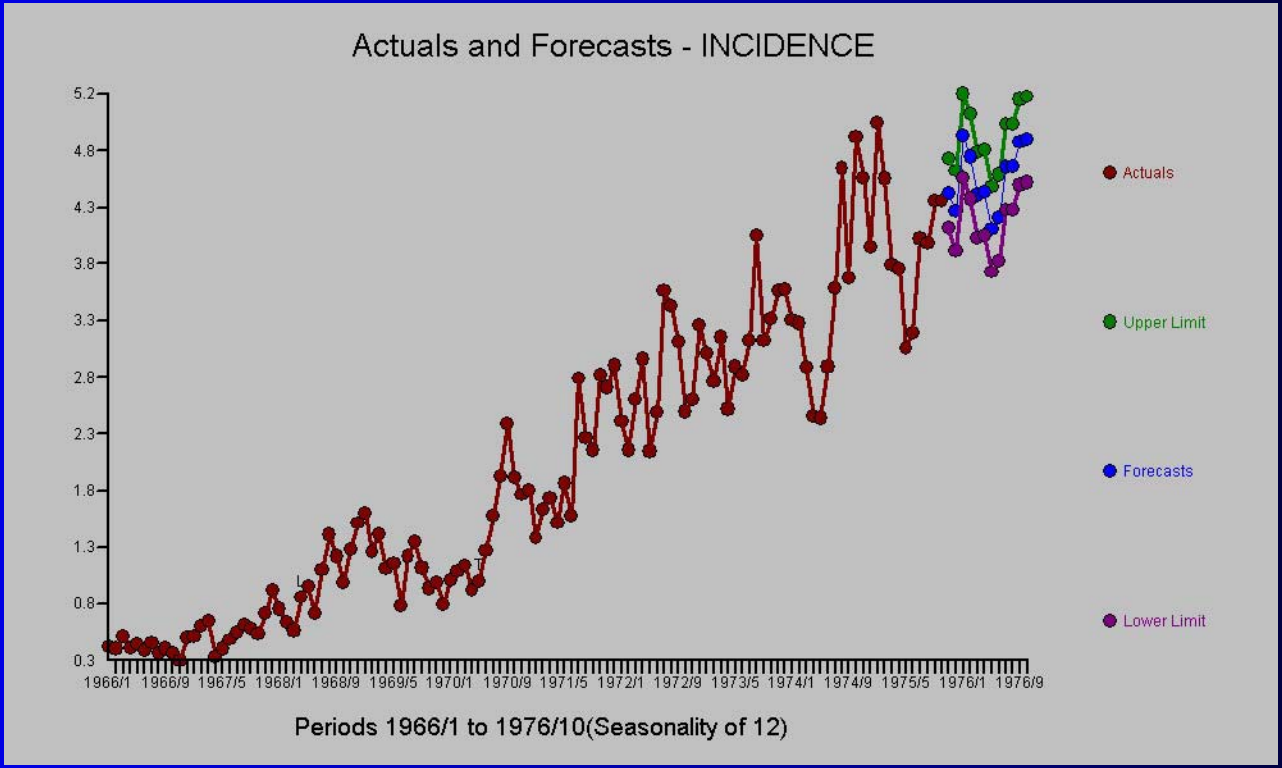
Minor Period: 1

Frequency: 12

Apply Cancel

nt Status | Engine = M | 9/30/2004 | 8:00 AM

# USING ALL 118 VALUES



View Options Process Help

storical Data | Future Values | Forecast Data | Graph | **Reports**

eports
- DETAILS.HTM
- INTRVENT.HTM
- EQUATION.TXT
- VERBAL.TXT
- STAT.HTM

THE ESTIMATED MODEL PARAMETERS

| # | MODEL COMPONENT | | LAG (BOP) | COEFF | STANDARD ERROR | P VALUE | T VALUE |
|---|---|---|---|---|---|---|---|
| 1 | CONSTANT | | | .435E-01 | .262E-01 | .0999 | 1.66 |
| 2 | Autoregressive-Factor # | 1 | 1 | .972 | .231E-01 | .0000 | 42.09 |
| 3 | Autoregressive-Factor # | 2 | 2 | -.257 | .976E-01 | .0096 | -2.64 |
| 4 | Moving Average-Factor # | 3 | 1 | .363 | .103 | .0006 | 3.54 |
| | INPUT SERIES X1 LAW | | | | | | |
| 5 | Omega (input) -Factor # | 4 | 0 | -.112 | .907 | .9021 | -.12 |

```
Y(T) =  .77504
       +[X1(T)][(-   .112)]
       +    [(1-   .972B** 1)(1+  .257B** 2)]**-1 [(1-   .363B** 1)] [A(T)]
```

A NON-CONSTANT ERROR VARIANCE HAS BEEN REMEDIED VIA WEIGHTED ESTIMATION CULMINATING AS A GENERALIZED LEAST SQUARES MODEL WITH A HOMOSCEDASTIC ERROR PROCESS.

| DIRECTION | TIME (T) | DATE | F VALUE | P VALUE |
|---|---|---|---|---|
| INCREASING | 54 | 1970/ 6 | 7.73213 | .0000 |

Since the automatic model fixup option for the variance stability test is enabled, the program will now estimate the parameters of the model with a set of weights that adjusts the residuals to account for the variance change(s).

nt Status | Engine = M | 9/30/2004 | 8:02 AM

View Options Process Help

storical Data | Future Values | Forecast Data | Graph | **Reports**

eports
- DETAILS.HTM
- INTRVENT.HTM
- EQUATION.TXT
- VERBAL.TXT
- STAT.HTM

THE ESTIMATED MODEL PARAMETERS

| # | MODEL COMPONENT | | LAG (BOP) | COEFF | STANDARD ERROR | P VALUE | T VALUE |
|---|---|---|---|---|---|---|---|
| 1 | CONSTANT | | | .118 | .451E-01 | .0100 | 2.62 |
| 2 | Autoregressive-Factor # | 1 | 1 | .590 | .826E-01 | .0000 | 7.14 |
| 3 | Autoregressive-Factor # | 2 | 12 | .510 | .102 | .0000 | 5.00 |

INPUT SERIES X1 I~T00053 1970/ 5 TIME

| # | MODEL COMPONENT | | LAG (BOP) | COEFF | STANDARD ERROR | P VALUE | T VALUE |
|---|---|---|---|---|---|---|---|
| 4 | Omega (input) -Factor # | 3 | 0 | .476E-01 | .100E-01 | .0000 | 4.76 |

INPUT SERIES X2 I~L00028 1968/ 4 LEVEL

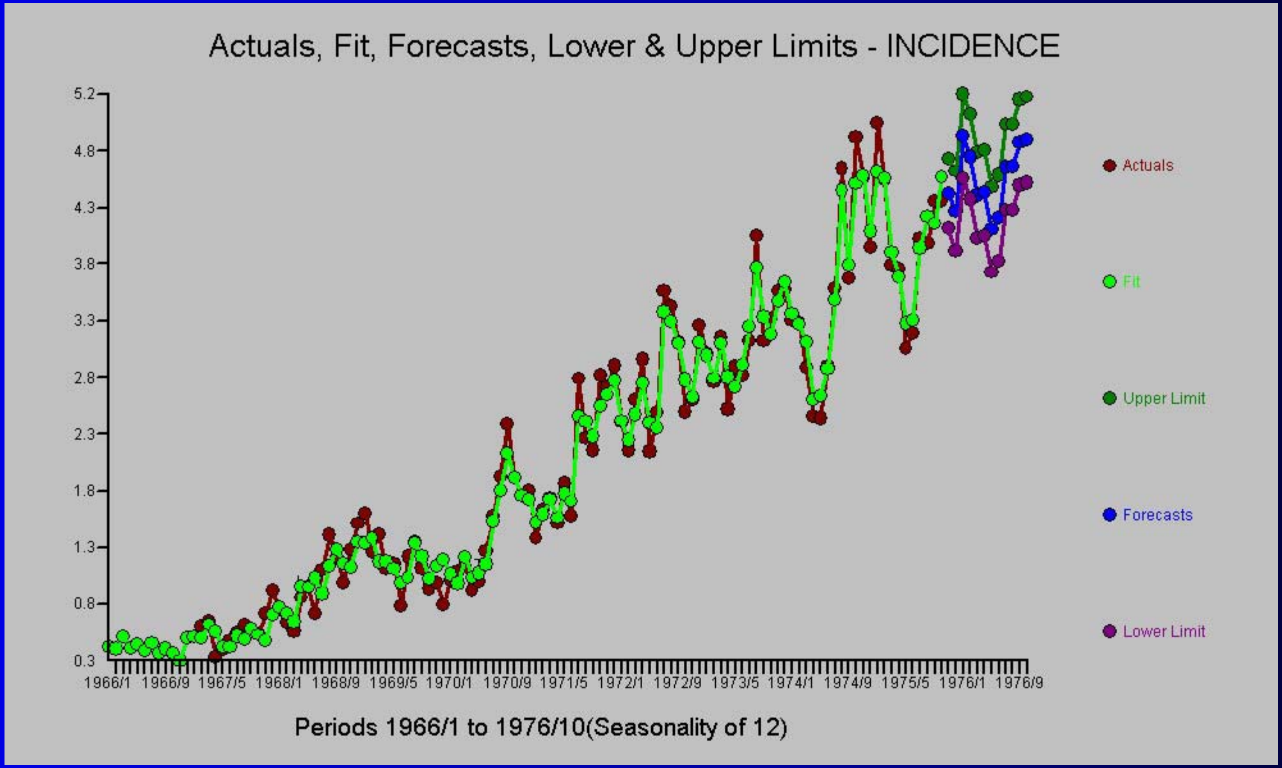
| # | MODEL COMPONENT | | LAG (BOP) | COEFF | STANDARD ERROR | P VALUE | T VALUE |
|---|---|---|---|---|---|---|---|
| 5 | Omega (input) -Factor # | 4 | 0 | .522 | .802E-01 | .0000 | 6.50 |

```
Y(T) =  .58902
       +[X1(T)][(+  .048)]
       +[X2(T)][(+  .522)]
       +    [(1-  .590B** 1)(1-  .510B** 12)]**-1  [A(T)]
```

A NON-CONSTANT ERROR VARIANCE HAS BEEN REMEDIED VIA WEIGHTED ESTIMATION CULMINATING AS A GENERALIZED LEAST SQUARES MODEL WITH A HOMOSCEDASTIC ERROR PROCESS.

| DIRECTION | TIME (T) | DATE | F VALUE | P VALUE |
|---|---|---|---|---|
| INCREASING | 54 | 1970/ 6 | 7.73213 | .0000 |

nt Status | Engine = M | 9/30/2004 | 8:20 AM



Actuals, Fit, Forecasts, Lower & Upper Limits - INCIDENCE
5.2
4.8
4.3
3.8
3.3
2.8
2.3
1.8
1.3
0.8
0.3
1966/1 1966/9 1967/5 1968/1 1968/9 1969/5 1970/1 1970/9 1971/5 1972/1 1972/9 1973/5 1974/1 1974/9 1975/5 1976/1 1976/9
Periods 1966/1 to 1976/10(Seasonality of 12)
Actuals
Fit
Upper Limit
Forecasts
Lower Limit