



## Monte Carlo Simulation

Presented by  
**Dr. Neil W. Polhemus**

# Monte Carlo Simulation



“Monte Carlo simulations are used to model the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables.” *Investopedia*

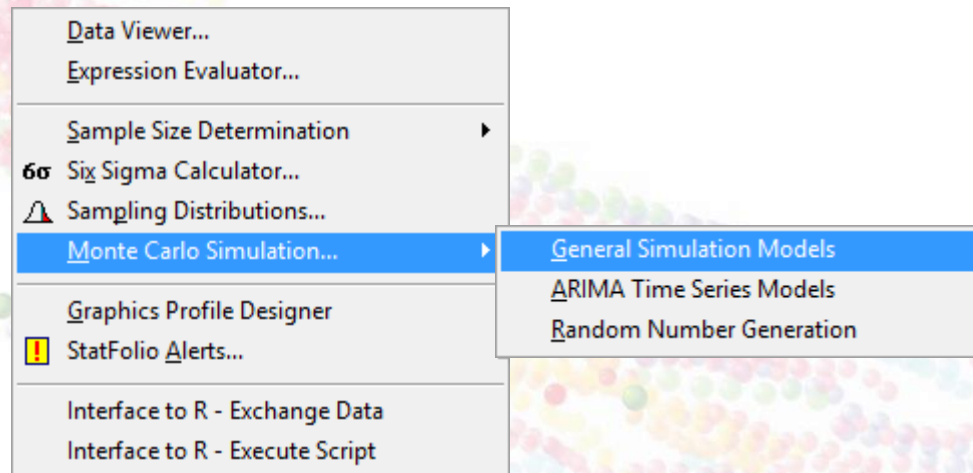
# Monte Carlo Simulation



“The world ... is full of more complicated systems .... the complex interaction of many variables — or the inherently probabilistic nature of certain phenomena — rules out a definitive prediction. So a Monte Carlo simulation uses essentially random inputs (within realistic limits) to model the system and produce probable outcomes.” *MIT News*

# Statgraphics Centurion XVII

*Tools* menu has a section for *Monte Carlo Simulation*:



# Outline

1. Random number generators
2. Example #1: Estimating the amount of oil available in an oil reservoir
3. Example #2: Simulating the consumer price index using an ARIMA time series model

# Random number generators

- Includes:
  - 7 discrete distributions
  - 39 continuous distributions
  - user-specified discrete distribution
    - user specifies values for  $x$  and  $p(x)$
  - user-specified continuous distribution
    - user specifies values for  $x$  and  $F(x)$

# Example – Gamma Distribution

The screenshot displays the 'Random Number Generator' dialog box. At the top, it shows 'Number of samples: 1', 'Sample size: 10000', 'Save to datasheet: A', and 'Random seed: 5025'. Below this is a table with columns for 'Column Name', 'Distribution', and 'Parameters'. The first row is populated with 'X1', 'Gamma', and 'GAMMA(10.0,2.0,0.0)'. A 'Gamma r.v.' sub-dialog box is open over the first row, showing 'Shape: 10.0', 'Scale: 2.0', and 'Threshold: 0.0'. Buttons for 'OK', 'Cancel', and 'Help' are visible in the sub-dialog and the main dialog.

Column Name	Distribution	Parameters
X1	Gamma	GAMMA(10.0,2.0,0.0)

# Example – Triangular Distribution

The screenshot displays the 'Random Number Generator' software interface. At the top, the title bar reads 'Random Number Generator'. Below the title bar, there are several input fields: 'Number of samples:' with a value of 2, 'Sample size:' with a value of 10000, 'Save to datasheet:' with a dropdown menu set to 'A', and 'Random seed:' with a value of 20234. A 'More' button is located to the right of these fields.

The main area of the interface is a table with three columns: 'Column Name', 'Distribution', and 'Parameters'. The table has 15 rows. The first two rows are populated:

Column Name	Distribution	Parameters
X1	Gamma	GAMMA(10.0,2.0,0.0)
X2	Triangular	TRIANGULAR(0.0,4.0,10.0)

Each row has an 'Edit' button to its right. A dialog box titled 'Triangular r.v.' is open in the foreground, showing the parameters for the triangular distribution: 'Lower Limit' is 0.0, 'Center' is 4.0, and 'Upper Limit' is 10.0. The dialog box has 'OK', 'Cancel', and 'Help' buttons.

At the bottom of the main window, there are 'OK', 'Cancel', and 'Help' buttons.



# Example – User Specified Discrete

The screenshot displays the 'Random Number Generator' software interface. At the top, it shows settings for 'Number of samples: 3', 'Sample size: 10000', 'Save to datasheet: A', and 'Random seed: 22898'. Below these settings is a table with columns for 'Column Name', 'Distribution', and 'Parameters'. The third row is highlighted, showing 'X3' with a 'User-specified discrete' distribution and parameters 'USERD(1;0.2;2;0.3;3;0.5)'. An 'Edit' button is visible next to this row. A secondary window titled 'User-Specified Discrete Distribution' is open, showing a list of values 'X:' (1, 2, 3) and their corresponding probabilities 'p(X):' (0.2, 0.3, 0.5). The 'Sort column names' checkbox is unchecked. The main window has 'OK' and 'Cancel' buttons at the bottom.

Column Name	Distribution	Parameters
1 X1	Gamma	GAMMA(10.0,2.0,0.0)
2 X2	Triangular	TRIANGULAR(0.0,4.0,10.0)
3 X3	User-specified discrete	USERD(1;0.2;2;0.3;3;0.5)
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

X:	p(X):
1	0.2
2	0.3
3	0.5

# Example – User Specified Continuous

Random Number Generator

Number of samples: 4    Sample size: 10000    Save to datasheet: A    Random seed: 22898    More

Column Name	Distribution	Parameters	
1 X1	Gamma	GAMMA(10.0,2.0,0.0)	Edit
2 X2	Triangular	TRIANGULAR(0.0,4.0,10.0)	Edit
3 X3	User-specified discrete	USERD(1;0.2;2;0.3;3;0.5)	Edit
4 X4	User-specified continuous	USERC(1;0.0;2;0.1;3;0.3;4;0.7;5;1.0)	Edit
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			

OK    Cancel

User-Specified Continuous Distribution

X: 1 2 3 4 5

F(X): 0.0 0.1 0.3 0.7 1.0

Sort column names

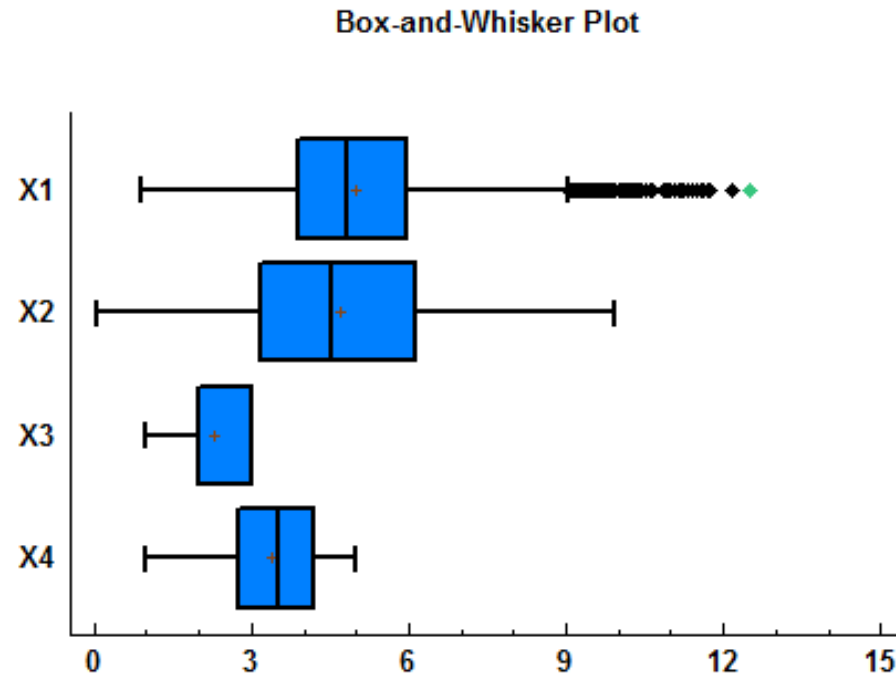
OK    Cancel    Help

# Datasheet

The screenshot shows a window titled "<untitled>" containing a data table with 18 rows and 4 columns labeled X1, X2, X3, and X4. The cell at row 1, column X1 (value 4.4263) is selected. The table data is as follows:

	X1	X2	X3	X4
1	4.4263	8.13002	3.0	4.43163
2	5.34978	3.42481	2.0	3.93414
3	5.09337	2.23698	3.0	3.38972
4	5.16672	4.29812	3.0	2.75472
5	3.42699	1.80771	2.0	3.99378
6	5.53609	0.932991	3.0	3.75126
7	2.89852	4.8701	2.0	3.24061
8	4.39979	4.49062	3.0	1.61507
9	5.33839	1.01202	3.0	2.94342
10	5.71663	7.69211	2.0	4.96178
11	6.15975	5.00934	3.0	3.13898
12	3.82894	8.36813	1.0	1.04288
13	4.61559	3.20668	3.0	4.79058
14	4.84695	4.00717	3.0	2.28208
15	4.51131	5.62881	2.0	3.14273
16	8.05416	3.10584	2.0	3.20243
17	5.78243	4.03803	3.0	4.97599
18	7.45231	2.58465	3.0	2.34332

# Box-and-Whisker Plot



# General Simulation Models

- Begin with a mathematical function:

$$Y = f(X_1, X_2, X_3, \dots, X_p)$$

- Specify probability distributions for  $X_1$  through  $X_p$ .
- Generate  $n$  sets of random numbers and plug them into the function.
- Calculate statistics from the  $n$  values of  $Y$ .

# Example: Oil Reserves

Murtha presents the following volumetric model for oil in place:

$$N = 7,758Ah\phi(1 - S_w) / B_o \quad \text{stock-tank barrels}$$

where

- $A$  = area
- $h$  = net pay
- $\phi$  = porosity
- $S_w$  = water saturation
- $B_o$  = formation volume factor

Question: Given a selected site, how small or large might  $N$  reasonably be?

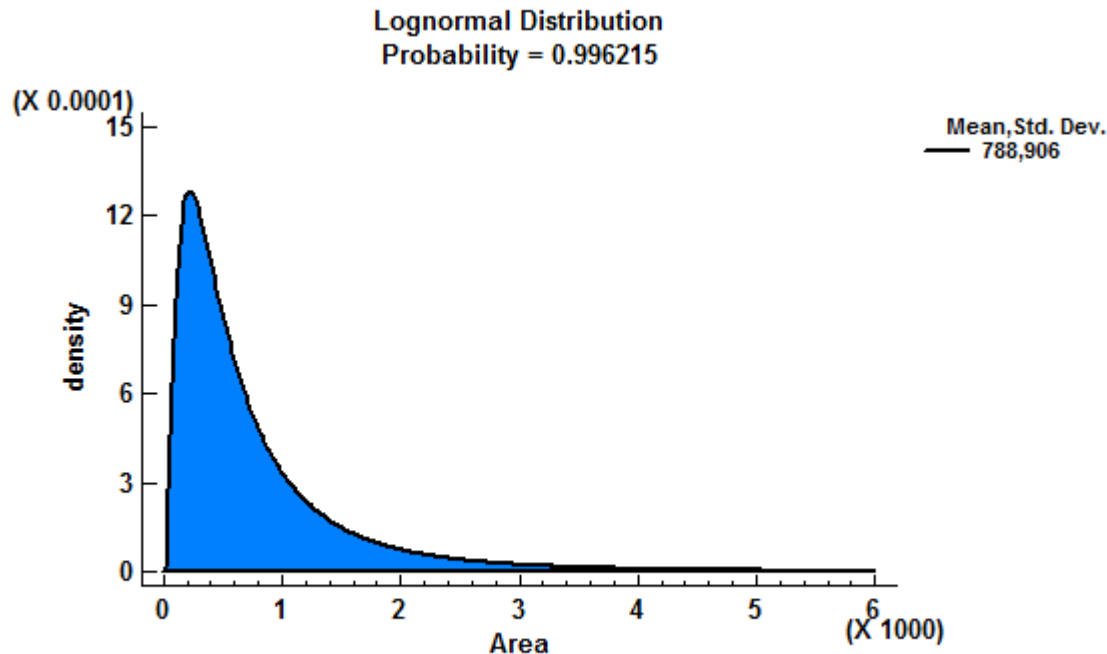
Source: Murtha, J.A. 1994. "Incorporating Historical Data Into Monte Carlo Simulation". *SPE Comp App* 6 (2): 11-17. SPE-26245-PA.

# How do get parameter distributions?

- If we have historical data, we can fit a probability distribution using the *Distribution Fitting (Uncensored Data)* procedure in Statgraphics.
  - Murtha used data from 83 similar reservoirs
- We could specify the minimum value, most likely value, and maximum value and use a triangular distribution.
- We could specify the mean and standard deviation and use a normal distribution.
- We could specify 2 or more percentiles and find a matching non-normal distribution.

# Area

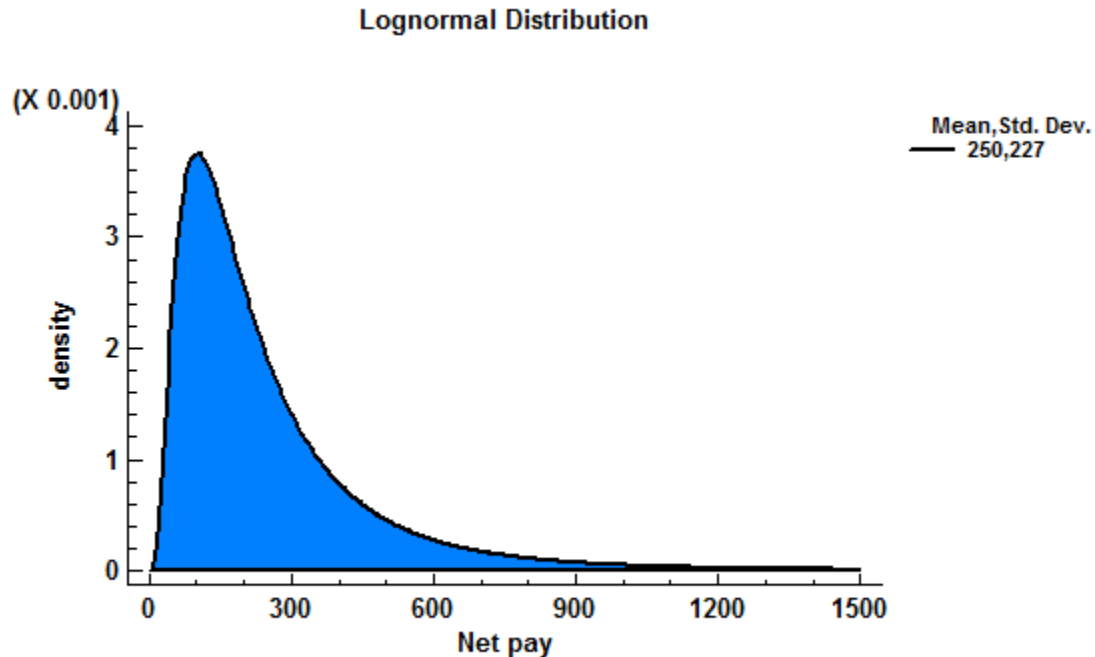
- Characterized by a lognormal distribution with mean = 788 and standard deviation = 906





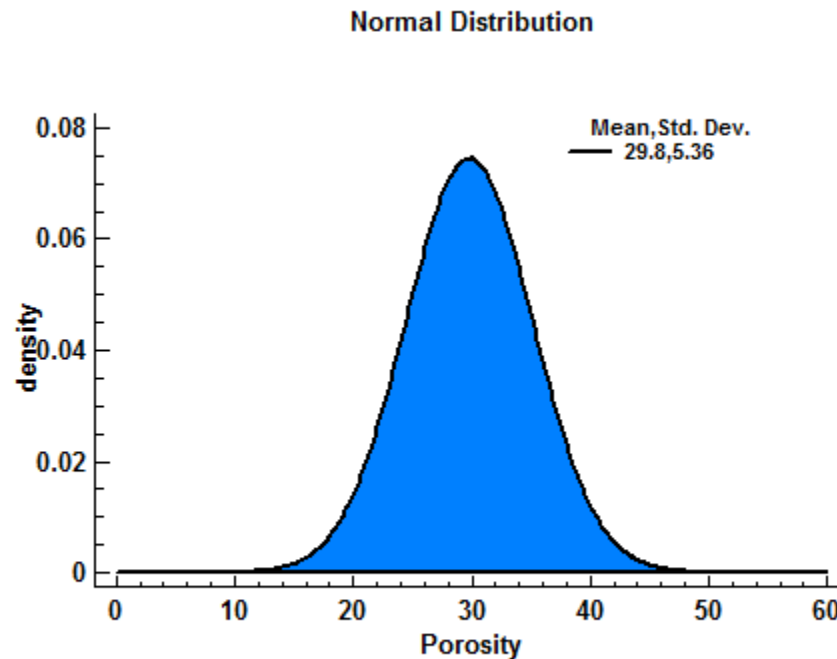
# Net pay

- Characterized by a lognormal distribution with mean = 250 and standard deviation = 227



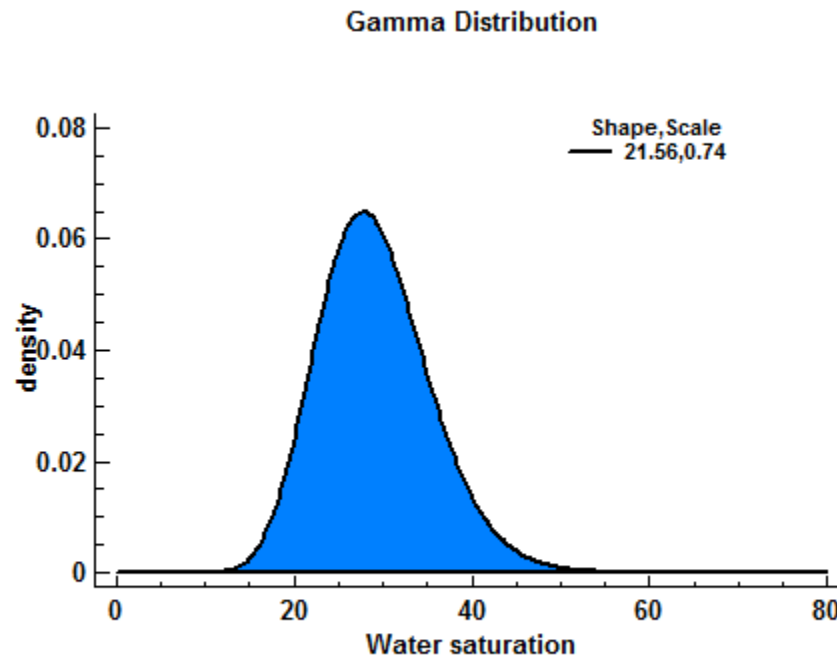
# Porosity as %

- Characterized by a normal distribution with mean = 29.8 and standard deviation = 5.36



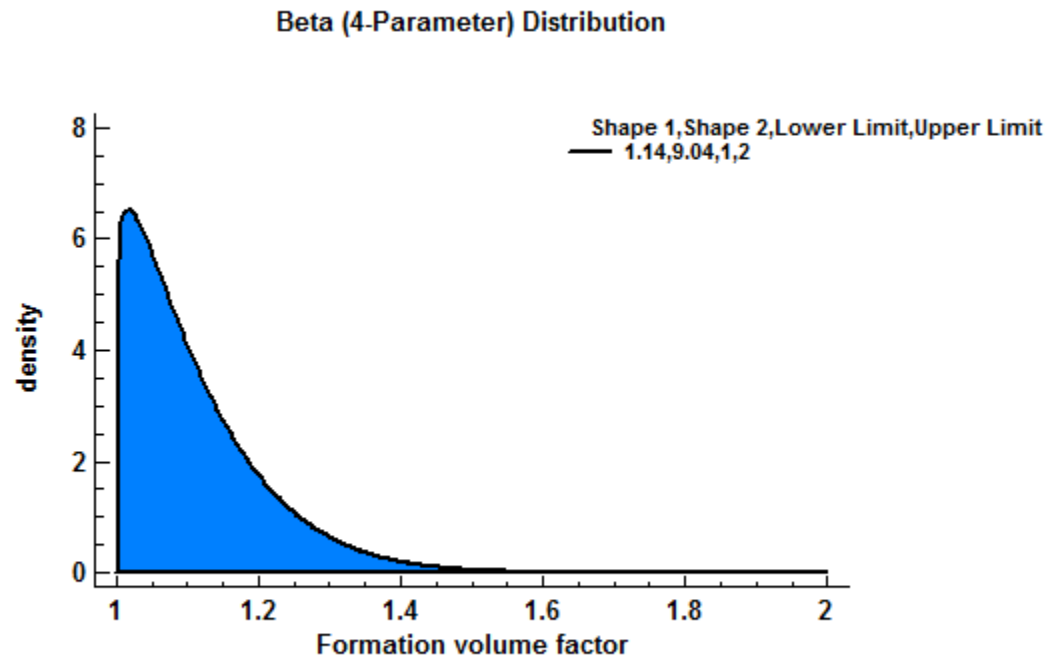
# Water saturation as %

- Characterized by a gamma distribution with shape = 21.56 and scale = 0.74



# Formation volume factor

- Characterized by a beta distribution with shape1 = 1.14 and shape2 = 9.04, in the interval [1,2]



# Matching Percentiles

- *Water saturation*: suppose we are 90% certain that it is between 20 and 40 and want to use a gamma distribution.
- I have written an R script that I've saved in *matchpercentiles.sgp* that will find the parameters of various distributions for which the cdf satisfies:
  - $F(20) = 0.05$
  - $F(40) = 0.95$

# R Script

```
Interface to R - Execute Script

#LOAD REQUIRED LIBRARY
library(nleqslv)

## Warning: package 'nleqslv' was built under R version 3.2.5

#SET VALUES FOR X WITH PROBABILITIES
X<-c(20,40)
p<-c(0.05,0.95)

#DEFINE FUNCTION TO RETURN DIFFERENCES FROM TARGET VALUES
fn <- function(parms,X,p,dist) {
  result<-c(999,999)
  if(parms[1]>0.0&&parms[2]>0.0) {
    if(dist==1) result <- p - pgamma(X,parms[1],parms[2])
    if(dist==2) result <- p - plnorm(X,parms[1],parms[2])
    if(dist==3) result <- p - pnorm(X,parms[1],parms[2]) }
  return (result) }

#SOLVE FOR GAMMA PARAMETER VALUES
sol<-nleqslv(c(1.0,0.01),fn,p=p,X=X,dist=1,control=list(allowSingular=TRUE))
sol$x

## [1] 22.9498931  0.7839015

#SOLVE FOR LOGNORMAL PARAMETER VALUES
sol<-nleqslv(c(10,1),fn,p=p,X=X,dist=2,control=list(allowSingular=TRUE))
c(exp(sol$x[1]+0.5*sol$x[2]^2),sqrt(exp(2*sol$x[1]+sol$x[2]^2)*(exp(sol$x[2]^2)-1)))

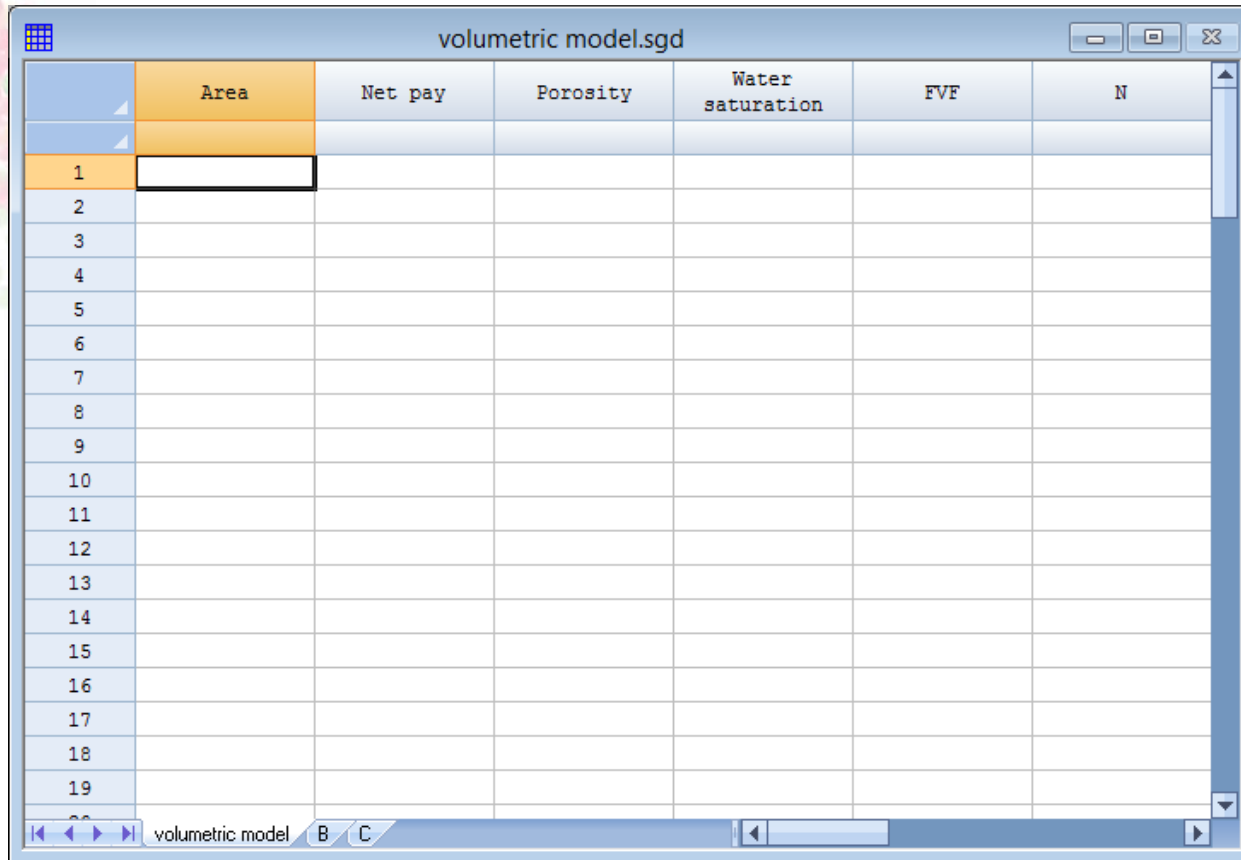
## [1] 28.919135  6.161571

#SOLVE FOR NORMAL PARAMETER VALUES
sol<-nleqslv(c(2000,500),fn,p=p,X=X,dist=3,control=list(allowSingular=TRUE))
sol$x

## [1] 30.000000  6.079568
```

# General Simulation Models

- Step 1: create datasheet with variables



	Area	Net pay	Porosity	Water saturation	FVF	N
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						

# General Simulation Models

- Step 2: define the simulation

Monte Carlo Simulation

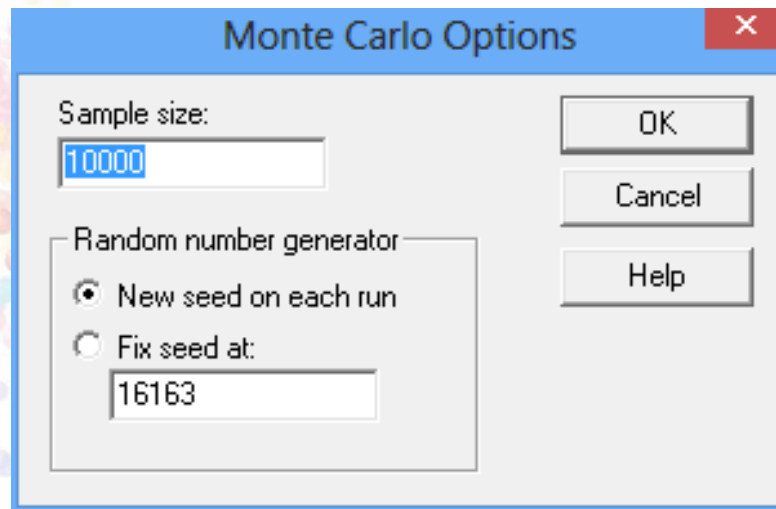
Number of variables:

	Variable	Type	Definition	
1	Area	Lognormal r.v.	LOGNORMAL(788.0,906.0,0.0)	Edit
2	Net pay	Lognormal r.v.	LOGNORMAL(250.0,227.0,0.0)	Edit
3	Porosity	Normal r.v.	NORMAL(29.8,5.36)	Edit
4	Water saturation	Gamma r.v.	GAMMA(21.56,0.74,0.0)	Edit
5	FVF	Beta r.v.	BETA(1.14,9.04,1.0,2.0)	Edit
6	N	Function	7758*Area*Net pay*(Porosity/100)*(1-Water saturation/100)/FVF	Edit
7				Edit
8				Edit
9				Edit
10				Edit
11				Edit
12				Edit
13				Edit
14				Edit
15				Edit



# General Simulation Models

- Step 3: run the simulation



A screenshot of a 'Monte Carlo Options' dialog box. The dialog has a blue title bar with a close button (X) in the top right corner. The main area is light gray and contains the following elements:

- Sample size:** A text input field containing the value '10000'.
- Random number generator:** A section containing two radio buttons:
  - New seed on each run
  - Fix seed at:
    - A text input field containing the value '16163'.
- Buttons:** Three buttons are stacked vertically on the right side: 'OK', 'Cancel', and 'Help'.

# General Simulation Models

- Step 4: calculate summary statistics

Summary Statistics

	<i>Area</i>	<i>Net pay</i>	<i>Porosity</i>	<i>Water saturation</i>	<i>FVF</i>	<i>N</i>
Count	10000	10000	10000	10000	10000	10000
Average	798.257	254.146	29.7738	29.1382	1.11129	3.02279E8
Median	520.151	187.103	29.7858	28.6549	1.08525	1.41263E8
Standard deviation	921.451	235.396	5.34707	6.2976	0.094791	5.81547E8
Coeff. of variation	115.433%	92.6223%	17.959%	21.6128%	8.5298%	192.388%
Minimum	13.2084	11.7636	3.71162	10.7716	1.00003	2.31833E6
Maximum	18030.1	3364.35	50.0127	58.1647	1.62905	1.8165E10
Range	18016.9	3352.59	46.3011	47.3931	0.62902	1.81627E10
Lower quartile	284.913	110.926	26.1484	24.6363	1.03933	6.15614E7
Upper quartile	963.019	318.933	33.3238	33.0429	1.15893	3.22183E8
Std. skewness	189.92	139.082	-0.101744	19.5701	55.7774	372.656
Std. kurtosis	882.858	425.331	0.724102	7.0223	41.8117	3189.81

# General Simulation Models

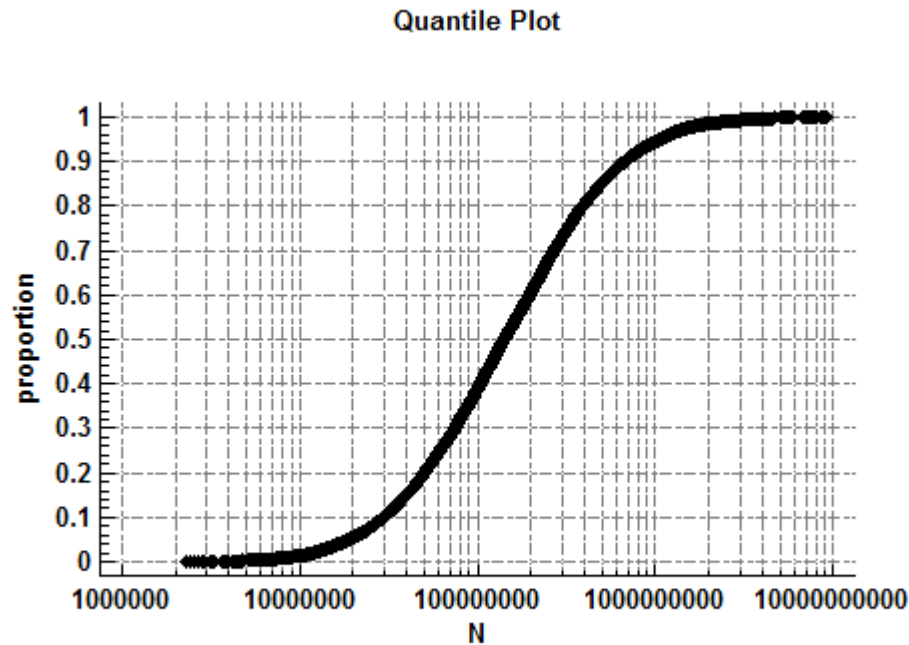
- Step 5: calculate percentiles

## Percentiles

<i>Percentage</i>	<i>Area</i>	<i>Net pay</i>	<i>Porosity</i>	<i>Water saturation</i>	<i>FVF</i>	<i>N</i>
1.0%	63.8958	30.9285	17.3382	16.6558	1.00185	8.72609E6
5.0%	117.535	50.8366	21.0517	19.7407	1.00836	1.90795E7
10.0%	162.918	67.8868	22.8418	21.4664	1.01613	2.95456E7
25.0%	284.913	110.926	26.1484	24.6363	1.03933	6.15614E7
50.0%	520.151	187.103	29.7858	28.6549	1.08525	1.41263E8
75.0%	963.019	318.933	33.3238	33.0429	1.15893	3.22183E8
90.0%	1733.26	512.351	36.5647	37.4472	1.24328	6.72699E8
95.0%	2352.44	676.942	38.6333	40.3498	1.30265	1.07355E9
99.0%	4462.55	1157.71	42.4143	46.0589	1.41934	2.5423E9

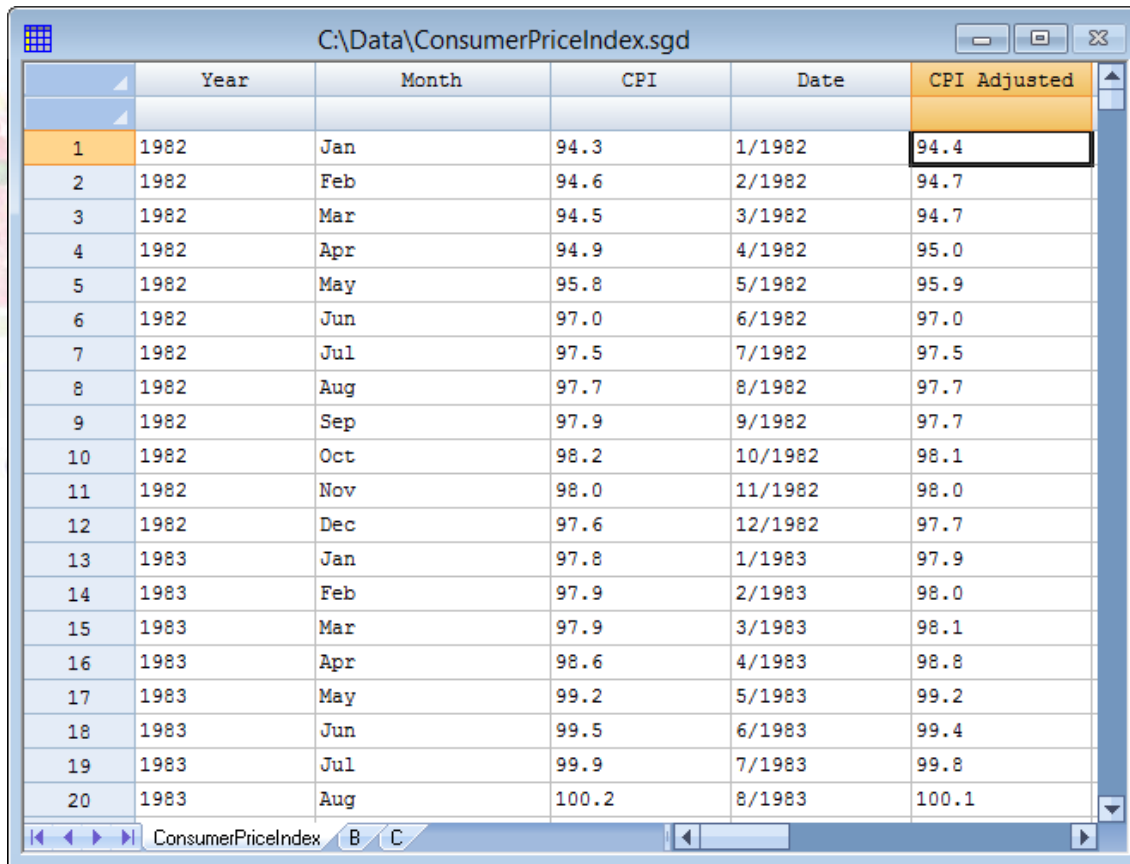
# General Simulation Models

- Step 6: create a quantile plot for the response



# Example #2: Consumer Price Index

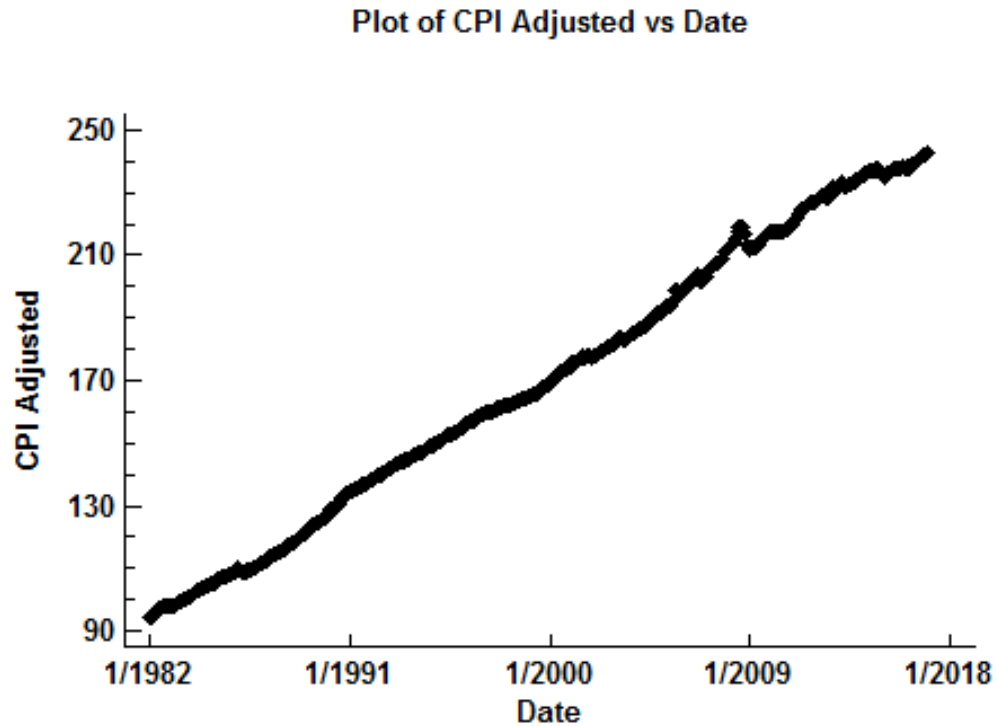
All urban consumers, all items, 1982-2016



	Year	Month	CPI	Date	CPI Adjusted
1	1982	Jan	94.3	1/1982	94.4
2	1982	Feb	94.6	2/1982	94.7
3	1982	Mar	94.5	3/1982	94.7
4	1982	Apr	94.9	4/1982	95.0
5	1982	May	95.8	5/1982	95.9
6	1982	Jun	97.0	6/1982	97.0
7	1982	Jul	97.5	7/1982	97.5
8	1982	Aug	97.7	8/1982	97.7
9	1982	Sep	97.9	9/1982	97.7
10	1982	Oct	98.2	10/1982	98.1
11	1982	Nov	98.0	11/1982	98.0
12	1982	Dec	97.6	12/1982	97.7
13	1983	Jan	97.8	1/1983	97.9
14	1983	Feb	97.9	2/1983	98.0
15	1983	Mar	97.9	3/1983	98.1
16	1983	Apr	98.6	4/1983	98.8
17	1983	May	99.2	5/1983	99.2
18	1983	Jun	99.5	6/1983	99.4
19	1983	Jul	99.9	7/1983	99.8
20	1983	Aug	100.2	8/1983	100.1

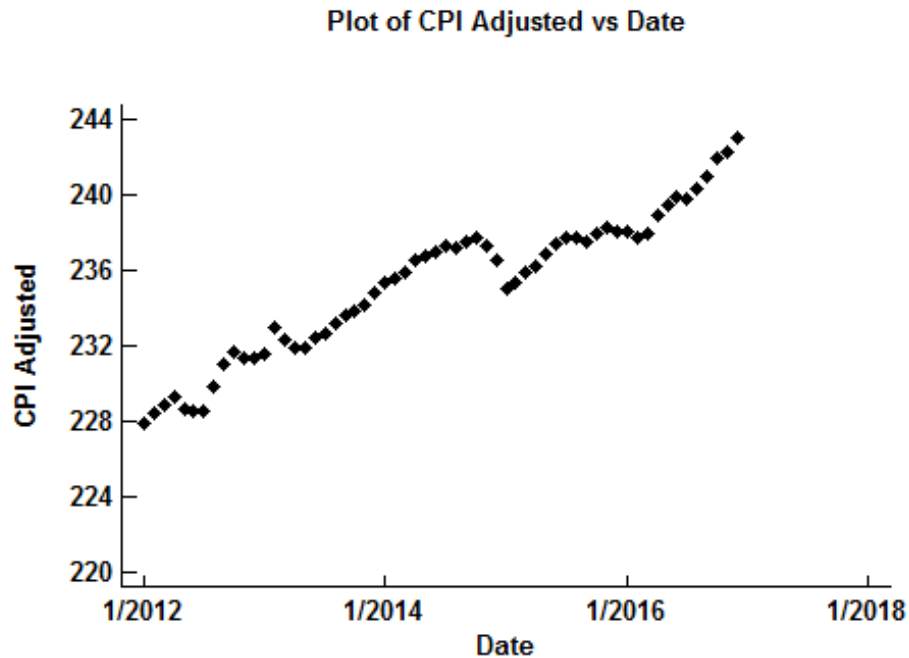
Source: <https://www.bls.gov/cpi/data.htm>

# CPI during last 35 years



Average yearly change: 4.3

# CPI during last 5 years



Average yearly change: 3.2

# Simulating ARIMA Time Series Models

- Steps
  - Fit an ARIMA model to the CPI data
  - Generate several possible realizations
  - Plot the realizations on a single graph
- Useful as input to financial models



# Automatic Forecasting

Automatic Forecasting

Year  
Month  
CPI  
Date  
CPI Adjusted

Sort column names

Data:  
CPI Adjusted

(Time Indices:)  
Date

or Sampling Interval

Once Every: 1

Starting At: 1

Year(s) (4-digit)  Hour(s)  
 Quarter(s)  Minute(s)  
 Month(s)  Second(s)  
 Day(s)  Other

(Seasonality:)

(Trading Days Adjustment:)

(Select:)

Number of Forecasts: 60 Withhold for Validation: 0

OK Cancel Delete Transform... Help

# Analysis Options

Automatic Forecasting Options

Models to Include

<input type="checkbox"/> Random Walk	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Random Walk with Drift	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Mean	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Linear Trend	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Quadratic Trend	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Exponential Trend	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> S-Curve	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Moving Average	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Simple Exp. Smoothing	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Brown's Linear Exp. Smoothing	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Holt's Linear Exp. Smoothing	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Quadratic Exp. Smoothing	<input checked="" type="checkbox"/> Optimize Parameters
<input type="checkbox"/> Winters' Exp. Smoothing	<input checked="" type="checkbox"/> Optimize Parameters
<input checked="" type="checkbox"/> ARIMA: <input checked="" type="checkbox"/> Optimize Model Order	<input checked="" type="checkbox"/> Optimize Parameters

AR Terms (p)

Nonseasonal:

Seasonal:

MA Terms (q)

Nonseasonal:

Seasonal:

Fix q at p-1

Differencing (d)

Nonseasonal:

Seasonal:

Include constant

Method Selection Criterion

Akaike Information Criterion (AIC)

Hannan-Quinn Criterion (HQC)

Schwarz Bayesian Inf. Criterion (SBIC)

Mean Squared Error (MSE)

Mean Absolute Error (MAE)

Mean Abs. Percentage Error (MAPE)

Adjustments...

Parameters...

Estimation...

Input series...

OK

Cancel

Help

# Estimated Model

## Forecast Summary

Forecast model selected: ARIMA(1,1,2) with constant

Number of forecasts generated: 60

Number of periods withheld for validation: 0

	<i>Estimation</i>	<i>Validation</i>
<i>Statistic</i>	<i>Period</i>	<i>Period</i>
RMSE	0.42377	
MAE	0.268858	
MAPE	0.155123	
ME	-0.0000700932	
MPE	-0.000434589	

## ARIMA Model Summary

<i>Parameter</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t</i>	<i>P-value</i>
AR(1)	-0.595915	0.219866	-2.71036	0.007000
MA(1)	-1.12592	0.209141	-5.38357	0.000000
MA(2)	-0.407999	0.0889357	-4.58757	0.000006
Mean	0.355178	0.0328146	10.8238	0.000000
Constant	0.566833			

Backforecasting: yes

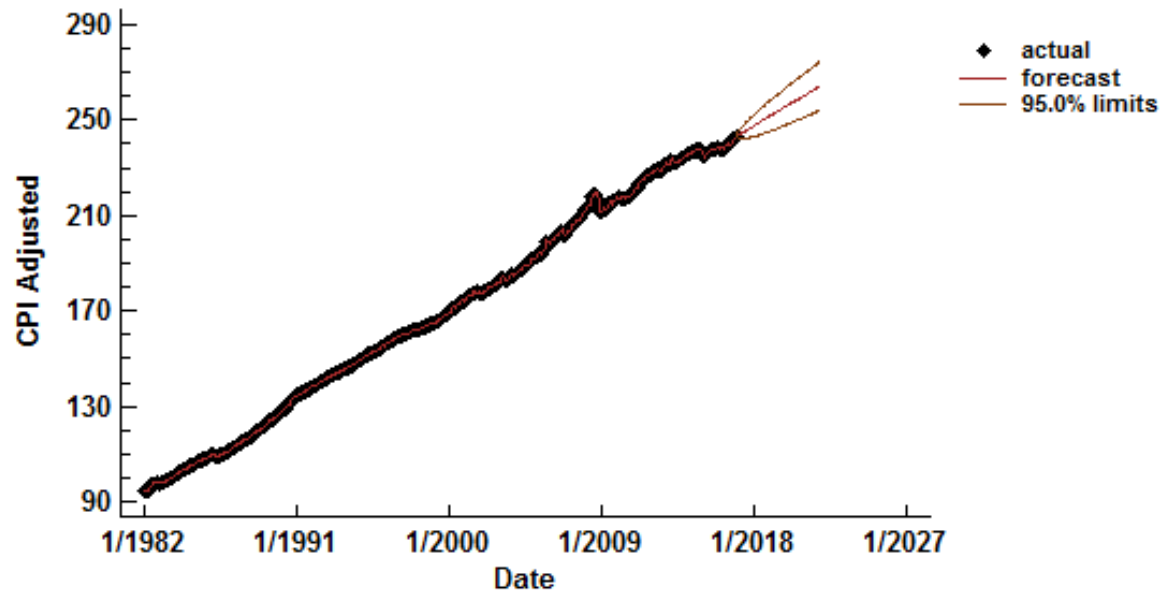
Estimated white noise variance = 0.179583 with 415 degrees of freedom

Estimated white noise standard deviation = 0.423772

Number of iterations: 8

# Forecast Plot

Time Sequence Plot for CPI Adjusted  
ARIMA(1,1,2) with constant



# ARIMA Model Simulation

ARIMA Model Simulator

Mean:  White noise sigma:  Sample size:  Random seed:

Nonseasonal Factors

p:  d:  q:

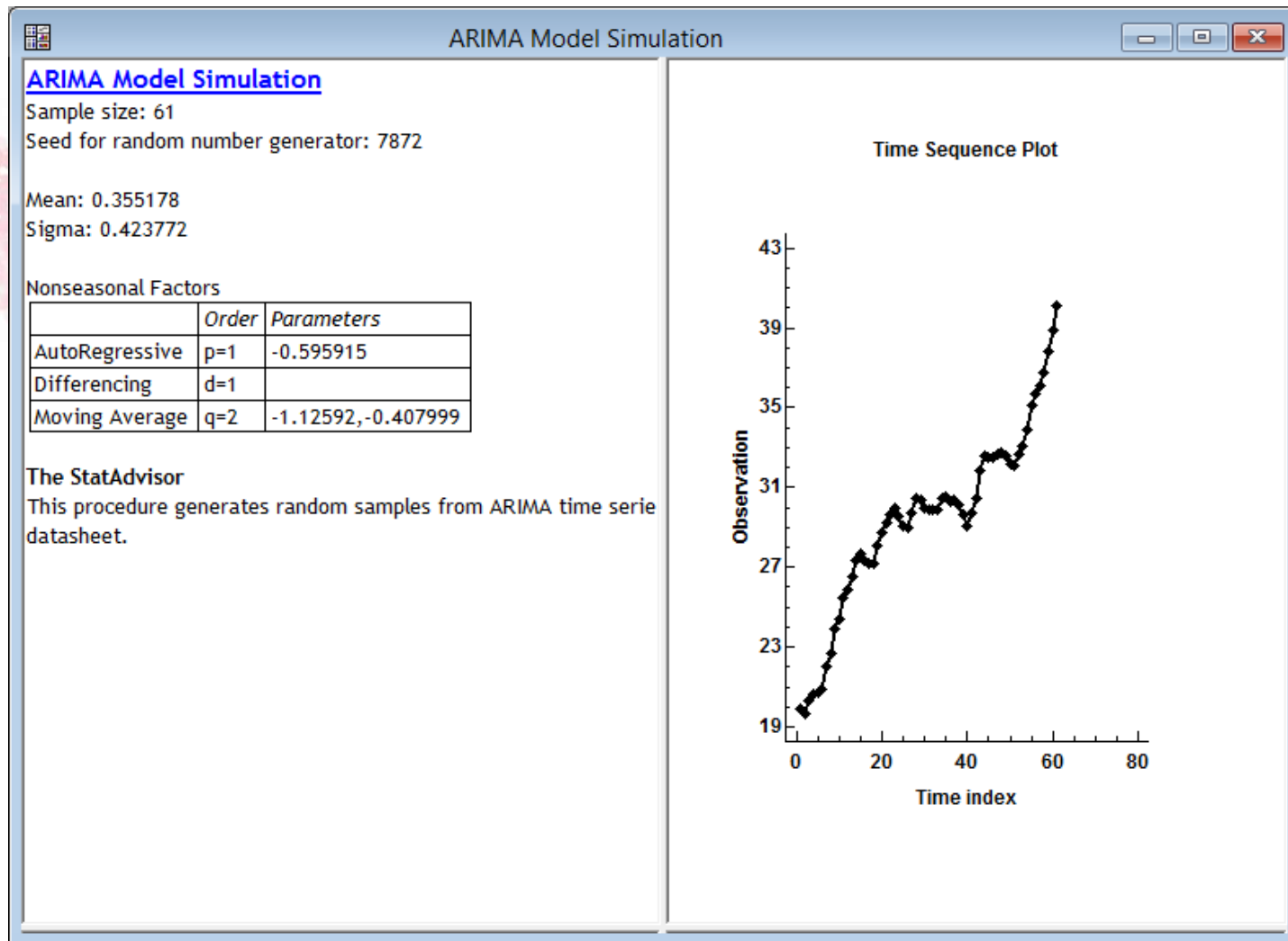
<input type="text" value="-0.595915"/>	<input type="text" value="-1.12592"/>
<input type="text" value="0.0"/>	<input type="text" value="-0.407999"/>
<input type="text" value="0.0"/>	<input type="text" value="0.0"/>
<input type="text" value="0.0"/>	<input type="text" value="0.0"/>
<input type="text" value="0.0"/>	<input type="text" value="0.0"/>
<input type="text" value="0.0"/>	<input type="text" value="0.0"/>

Seasonal Factors

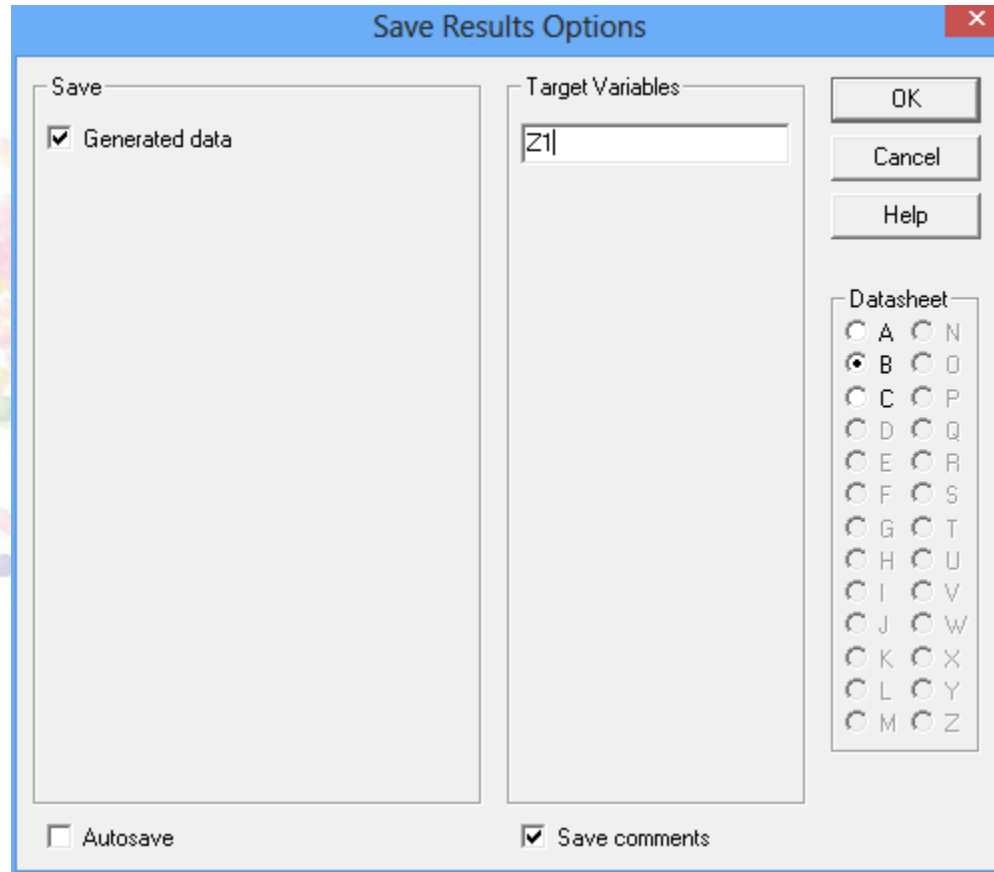
P:  D:  Q:  S:

<input type="text" value="0.0"/>	<input type="text" value="0.0"/>
<input type="text" value="0.0"/>	<input type="text" value="0.0"/>
<input type="text" value="0.0"/>	<input type="text" value="0.0"/>
<input type="text" value="0.0"/>	<input type="text" value="0.0"/>
<input type="text" value="0.0"/>	<input type="text" value="0.0"/>
<input type="text" value="0.0"/>	<input type="text" value="0.0"/>

# Simulated Time Series



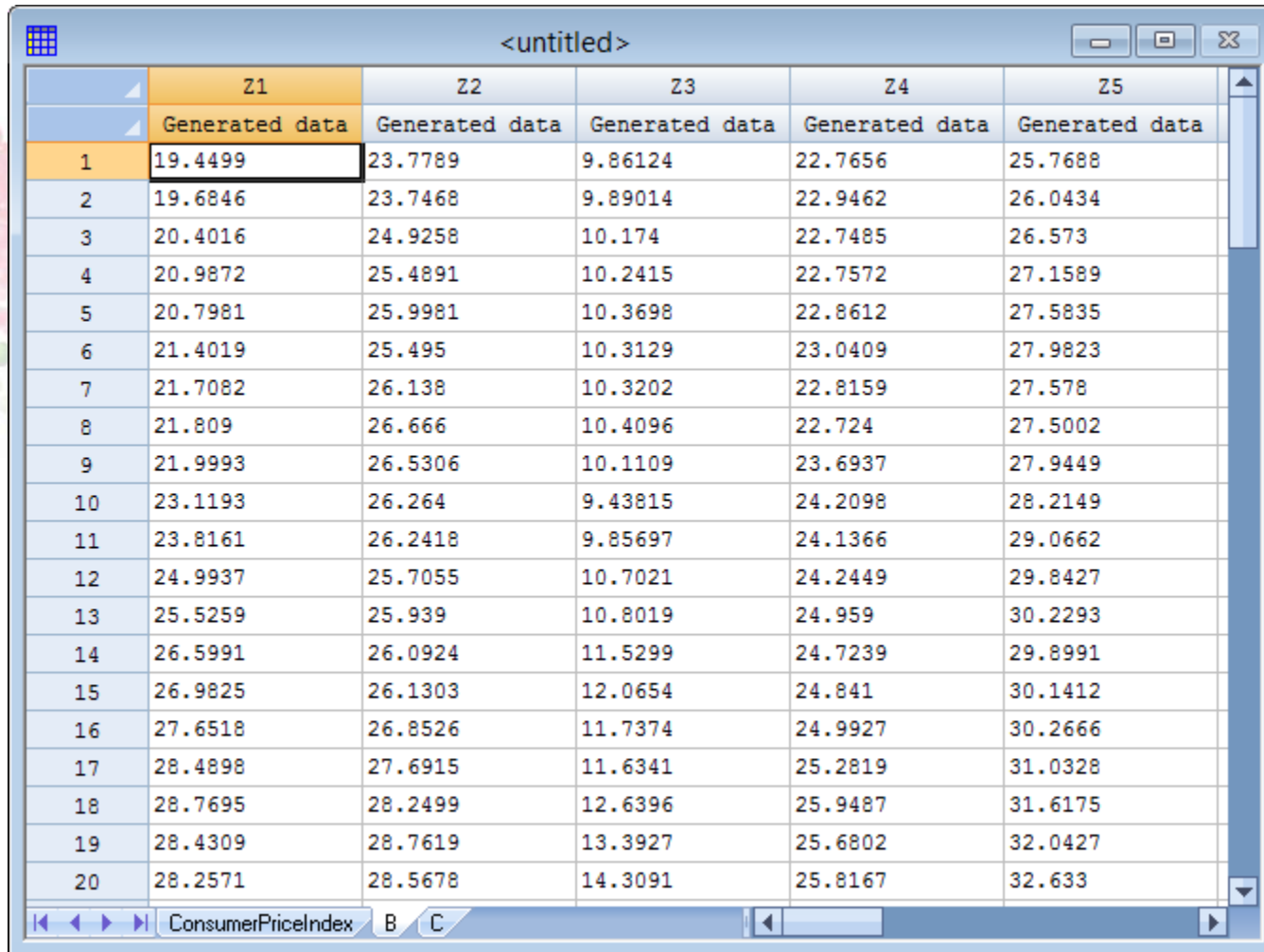
# Save Data



The image shows a 'Save Results Options' dialog box with a blue title bar and a red close button. The dialog is divided into several sections:

- Save:** A section with a checked checkbox for 'Generated data'.
- Target Variables:** A text input field containing the text 'Z1'.
- Buttons:** Three buttons on the right side: 'OK', 'Cancel', and 'Help'.
- Datasheet:** A grid of radio buttons labeled with letters A through Z. The radio button for 'B' is selected.
- Bottom Section:** Two checkboxes: 'Autosave' (unchecked) and 'Save comments' (checked).

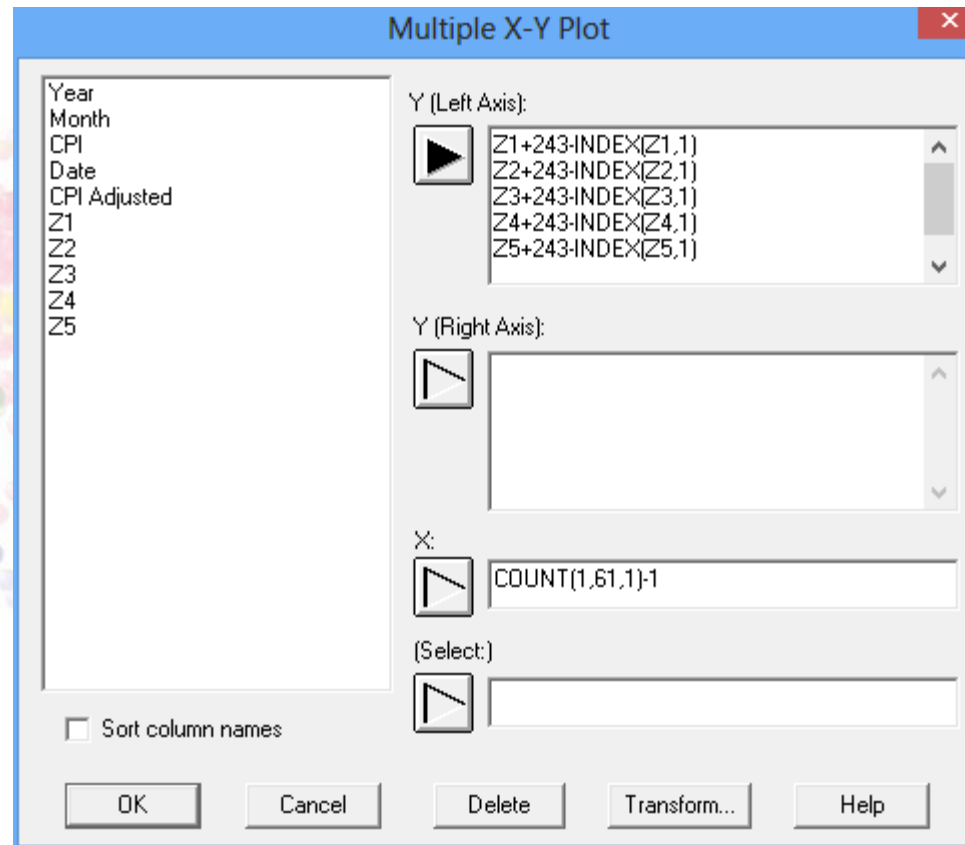
# Multiple Realizations



	Z1	Z2	Z3	Z4	Z5
	Generated data	Generated data	Generated data	Generated data	Generated data
1	19.4499	23.7789	9.86124	22.7656	25.7688
2	19.6846	23.7468	9.89014	22.9462	26.0434
3	20.4016	24.9258	10.174	22.7485	26.573
4	20.9872	25.4891	10.2415	22.7572	27.1589
5	20.7981	25.9981	10.3698	22.8612	27.5835
6	21.4019	25.495	10.3129	23.0409	27.9823
7	21.7082	26.138	10.3202	22.8159	27.578
8	21.809	26.666	10.4096	22.724	27.5002
9	21.9993	26.5306	10.1109	23.6937	27.9449
10	23.1193	26.264	9.43815	24.2098	28.2149
11	23.8161	26.2418	9.85697	24.1366	29.0662
12	24.9937	25.7055	10.7021	24.2449	29.8427
13	25.5259	25.939	10.8019	24.959	30.2293
14	26.5991	26.0924	11.5299	24.7239	29.8991
15	26.9825	26.1303	12.0654	24.841	30.1412
16	27.6518	26.8526	11.7374	24.9927	30.2666
17	28.4898	27.6915	11.6341	25.2819	31.0328
18	28.7695	28.2499	12.6396	25.9487	31.6175
19	28.4309	28.7619	13.3927	25.6802	32.0427
20	28.2571	28.5678	14.3091	25.8167	32.633

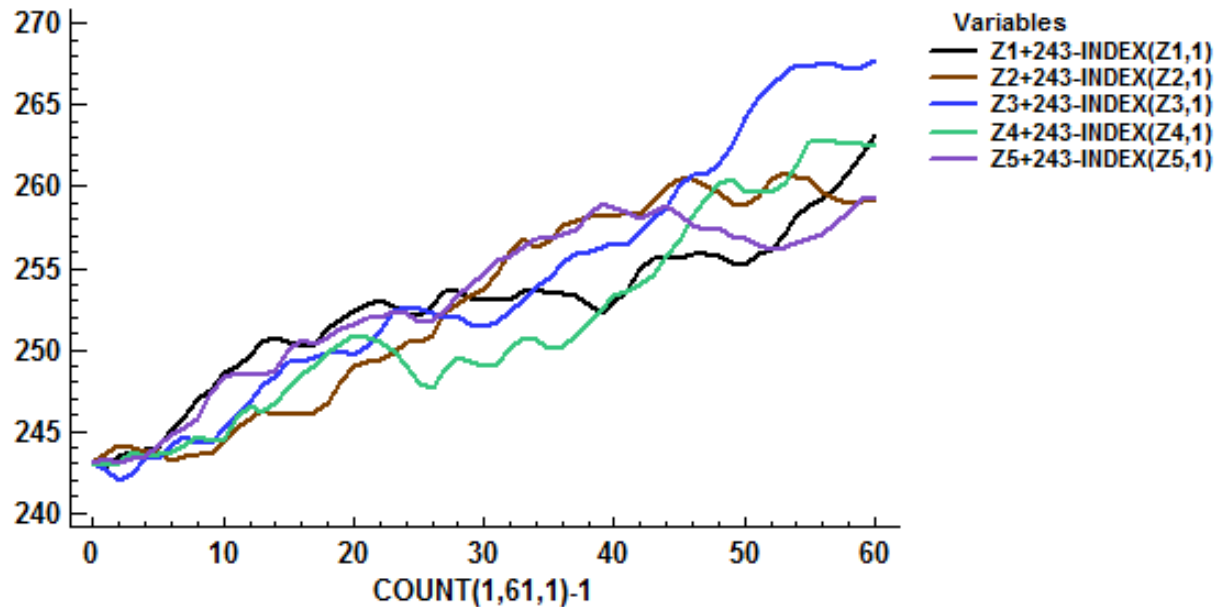


# Plot Multiple Realizations

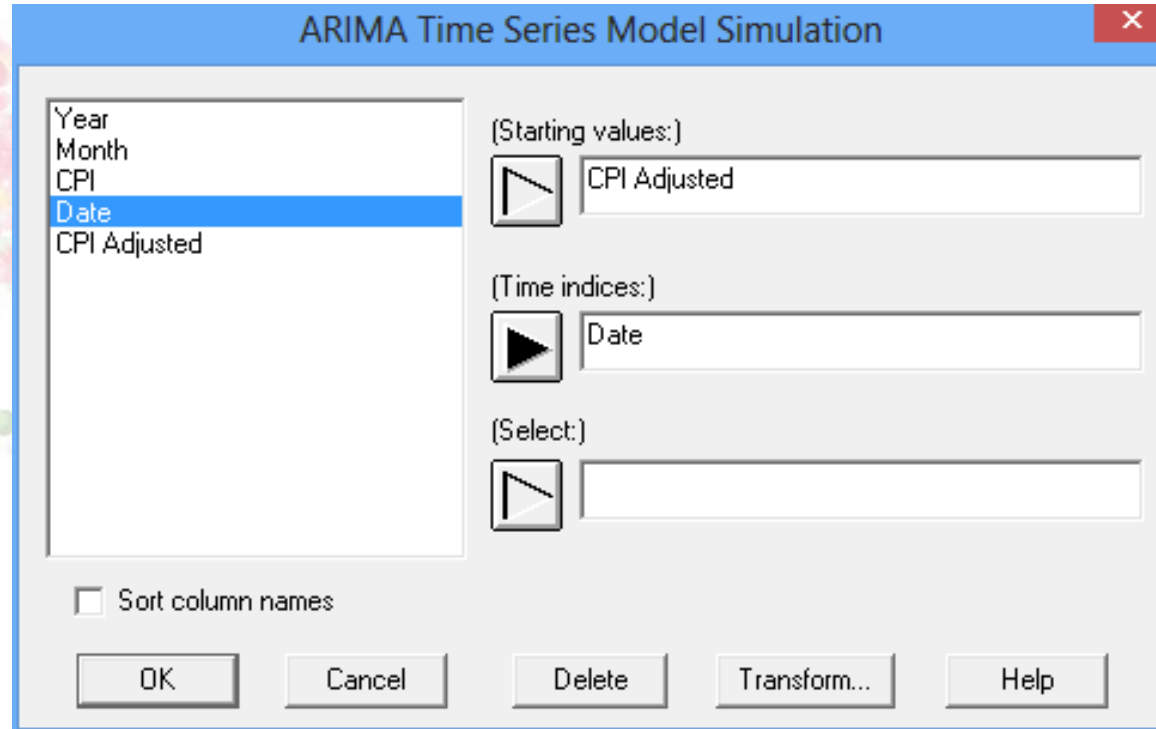


# Plot Multiple Realizations

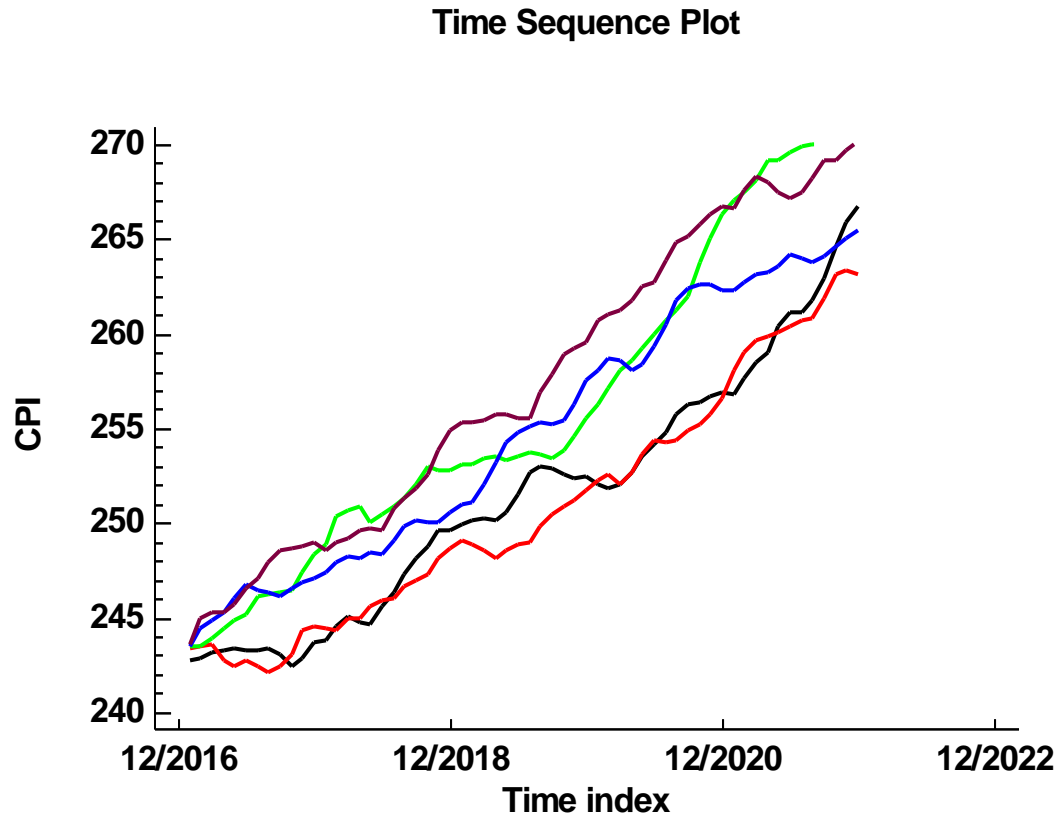
Multiple X-Y Plot



# Upcoming Enhancement



# Using Historical Data as Starting Values



# Resources

- StatFolios and data files are at: [www.statgraphics.com/webinars](http://www.statgraphics.com/webinars)
- Recorded webinar will be posted soon.