

Forecasting

Prepared by Dr. Mohamed Sleem

Subject code : FNT 631
Data Analysis

Lecture 7



Forecasting

Lecture 7

Subject code : FNT 631
Data Analysis

Prepared by
Dr. Mohamed Sleem

www.midocean.ae

What is the purpose of regression analysis?

- Regression analysis examines relationships between dependent and independent variables.

What is the difference between simple regression and multiple regression?

- Simple regression uses one independent variable, while multiple regression uses multiple independent variables.

What are some common applications of regression analysis?

- Forecasting outcomes, solving business problems, and supporting strategic decision-making.

Forecasting Techniques

- Qualitative and judgmental
- Statistical time series models
- Explanatory/causal models

Qualitative and Judgmental Methods

- **Historical analogy** – comparative analysis with a previous situation
- **Delphi Method** – response to a sequence of questionnaires by a panel of experts

Indicators and Indexes

- **Indicators** – measures believed to influence the behavior of a variable we wish to forecast
 - Leading indicators
 - Lagging indicators
- **Index** – a weighted combination of indicators
- Indicators and indexes are often used in economic forecasting

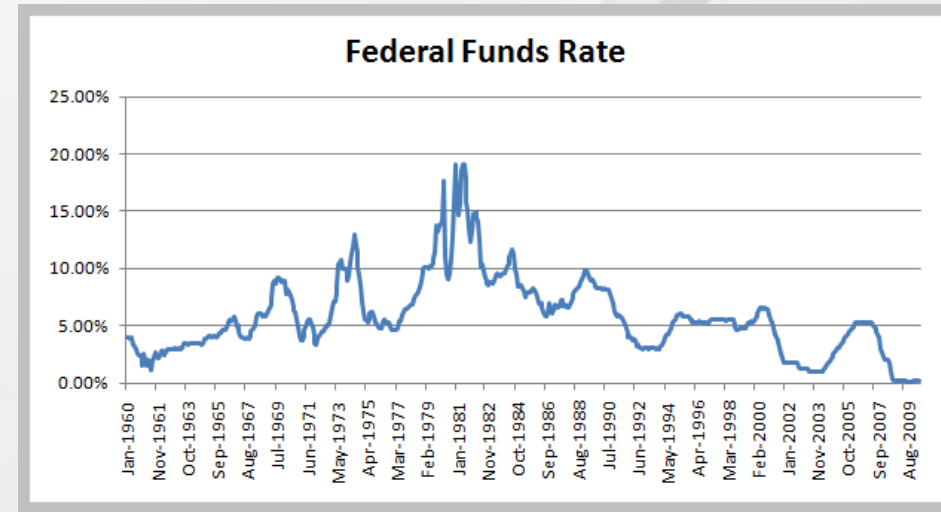
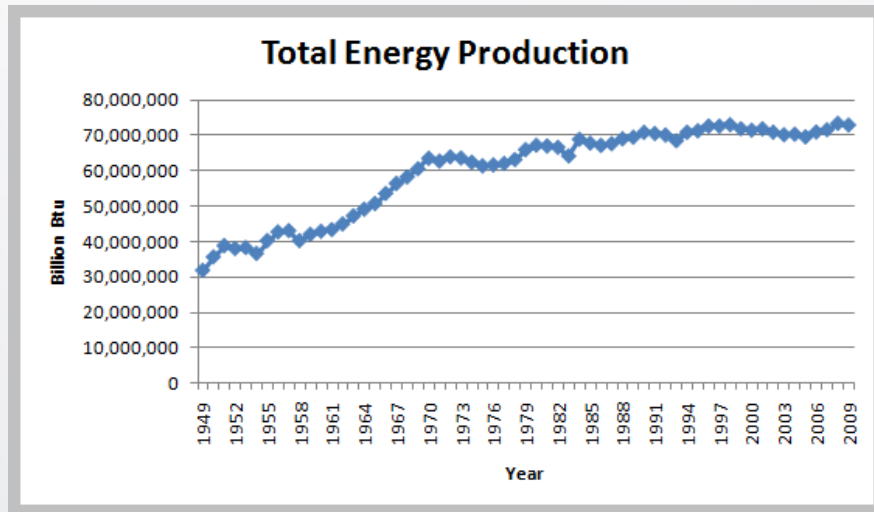
Example Dept. of Commerce Index of Leading Indicators

- Average weekly hours, manufacturing
- Average weekly initial claims, unemployment insurance
- New orders, consumer goods and materials
- Vendor performance—slower deliveries
- New orders, nondefense capital goods
- Building permits, private housing
- Stock prices, 500 common stocks (Standard & Poor)
- Money supply
- Interest rate spread
- Index of consumer expectations (University of Michigan)

Time Series

- A **time series** is a stream of historical data
- Components of time series
 - Trend
 - Short-term seasonal effects
 - Longer-term cyclical effects

Examples of Time Series



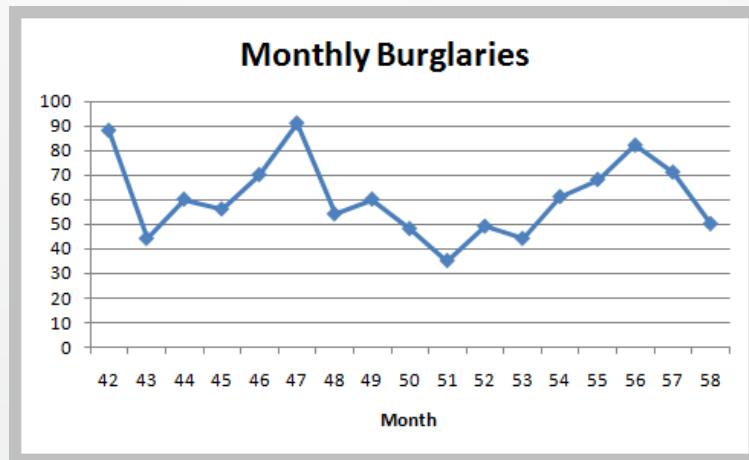
Statistical Forecasting Methods

- Moving average
- Exponential smoothing
- Regression analysis

Simple Moving Average

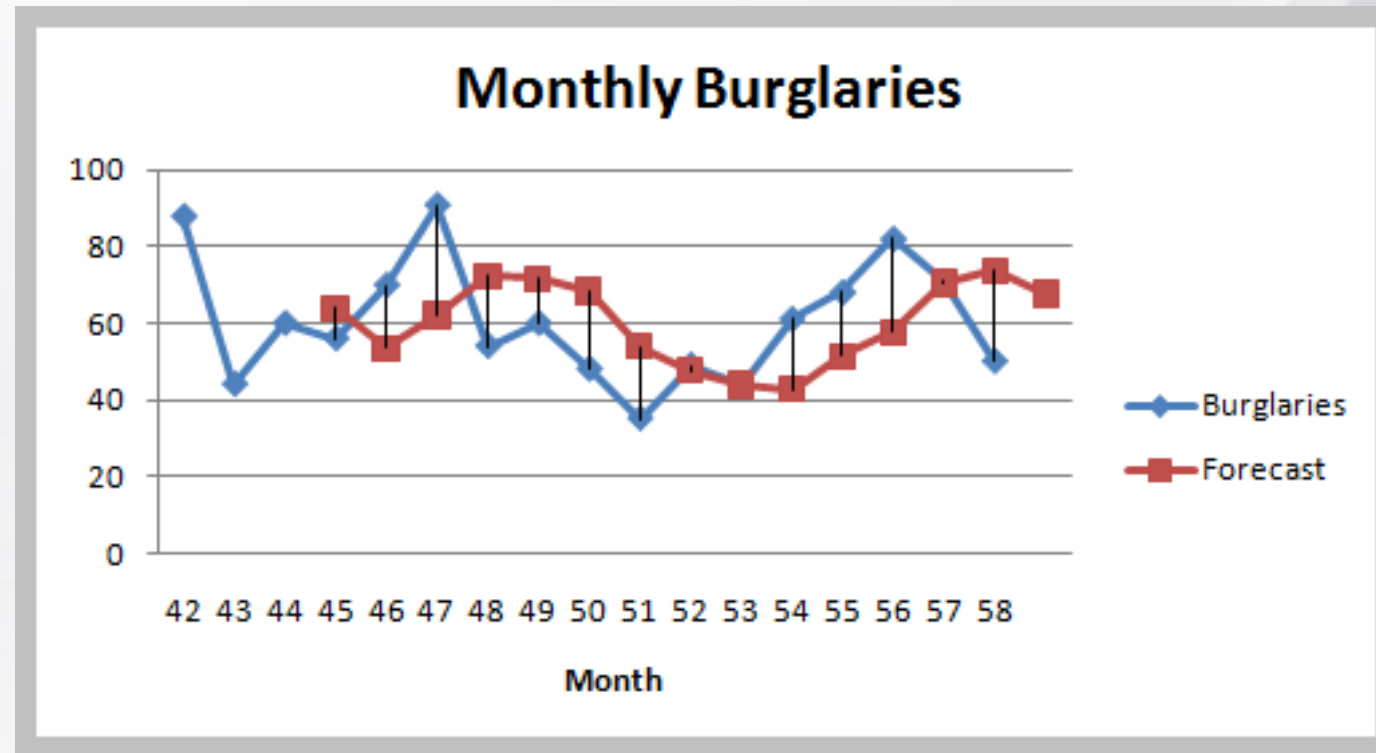
- Average random fluctuations in a time series to infer short-term changes in direction
- **Assumption:** future observations will be similar to recent past
- Moving average for next period = average of most recent k observations

Example: Moving Average Forecast With $k = 3$



	C	D	E	F
1	After Citizen-Police Program		Moving Average	
2	Month	Monthly burglaries	Forecast	
3	42	88		
4	43	44		
5	44	60		
6	45	56	64.00	Forecast for month 45 =AVERAGE(D3:D5)
7	46	70	53.33	
8	47	91	62.00	
9	48	54	72.33	
10	49	60	71.67	
11	50	48	68.33	
12	51	35	54.00	
13	52	49	47.67	
14	53	44	44.00	
15	54	61	42.67	
16	55	68	51.33	
17	56	82	57.67	
18	57	71	70.33	
19	58	50	73.67	Forecast for month 59 =AVERAGE(D17:D19)
20	59		67.67	
21				

Time Series Data and Moving Averages



Alternative Moving Average Models

	A	B	C	D	E	F	G	H
1	After Citizen-Police Program						3 Period	
2	Month	Monthly burglaries	k = 2	Error	k = 3	Error	Weighted	Error
3	42	88						
4	43	44						
5	44	60	66.00	-6.00				
6	45	56	52.00	4.00	64.00	-8.00	58.00	-2.00
7	46	70	58.00	12.00	53.33	16.67	56.00	14.00
8	47	91	63.00	28.00	62.00	29.00	64.80	26.20
9	48	54	80.50	-26.50	72.33	-18.33	81.20	-27.20
10	49	60	72.50	-12.50	71.67	-11.67	66.70	-6.70
11	50	48	57.00	-9.00	68.33	-20.33	61.30	-13.30
12	51	35	54.00	-19.00	54.00	-19.00	52.20	-17.20
13	52	49	41.50	7.50	47.67	1.33	41.40	7.60
14	53	44	42.00	2.00	44.00	0.00	44.70	-0.70
15	54	61	46.50	14.50	42.67	18.33	44.60	16.40
16	55	68	52.50	15.50	51.33	16.67	54.70	13.30
17	56	82	64.50	17.50	57.67	24.33	63.50	18.50
18	57	71	75.00	-4.00	70.33	0.67	75.70	-4.70
19	58	50	76.50	-26.50	73.67	-23.67	74.00	-24.00
20	59		60.50		67.67		59.50	

Error Metrics and Forecast Accuracy

- Mean absolute deviation (MAD)

$$\text{MAD} = \frac{\sum_{i=1}^n |A_t - F_t|}{n}$$

- Mean square error (MSE)

$$\text{MSE} = \frac{\sum_{i=1}^n (A_t - F_t)^2}{n}$$

- Mean absolute percentage error (MAPE)

$$\text{MAPE} = \frac{\sum_{i=1}^n \left| \frac{A_t - F_t}{A_t} \right|}{n}$$

Summary of Error Metrics for Burglary Data

TABLE 7.2 Error Metrics for Moving Average Models of Burglary Data

	$k = 2$	$k = 3$	Three-Period Weighted
MAD	13.63	14.86	13.70
MSE	254.38	299.84	256.31
RMSE	15.95	17.32	16.01
MAPE	23.63%	26.53%	24.46%

Exponential Smoothing

- Exponential smoothing model:

$$\begin{aligned}F_{t+1} &= (1 - \alpha)F_t + \alpha A_t \\ &= F_t + \alpha (A_t - F_t)\end{aligned}$$

- F_{t+1} is the forecast for time period $t+1$,
- F_t is the forecast for period t ,
- A_t is the observed value in period t , and
- α is a constant between 0 and 1, called the smoothing constant.

Exponential Smoothing Example

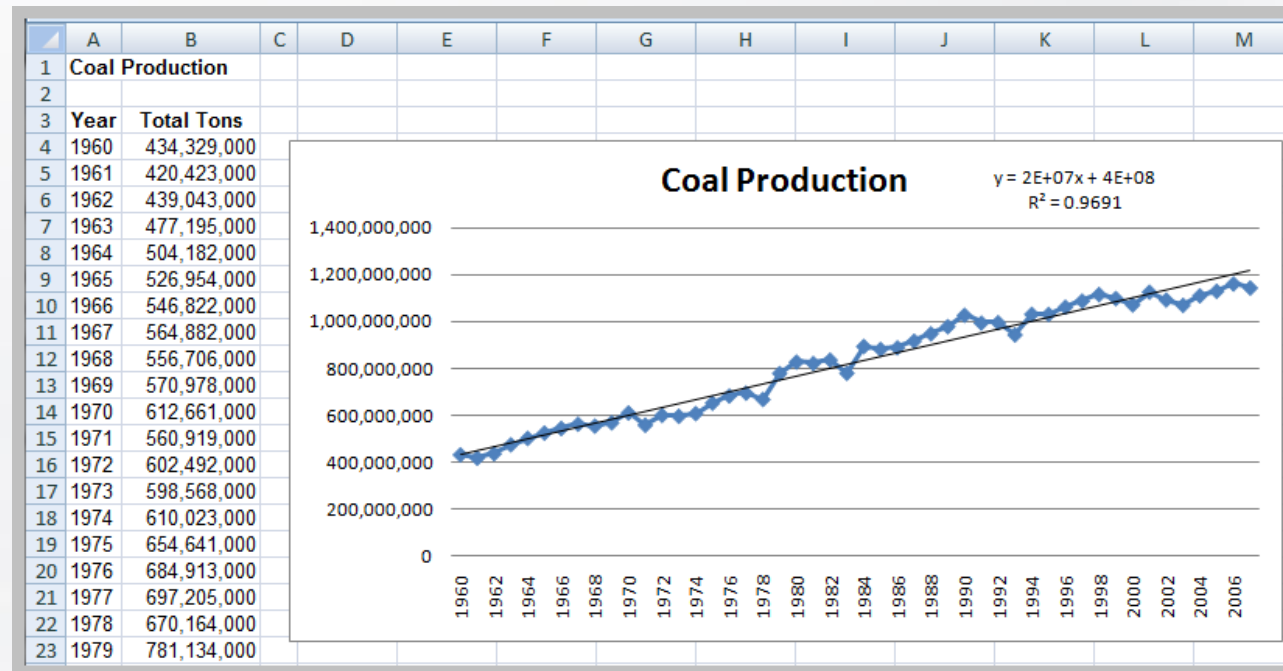
	C	D	E	F	G	H	I	J	K	L	M
1	After Citizen-Police Program		Smoothing Constant								
2	Month	Monthly burglaries	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3	42	88	88.00	88.00	88.00	88.00	88.00	88.00	88.00	88.00	88.00
4	43	44	88.00	88.00	88.00	88.00	88.00	88.00	88.00	88.00	88.00
5	44	60	83.60	79.20	74.80	70.40	66.00	61.60	57.20	52.80	48.40
6	45	56	81.24	75.36	70.36	66.24	63.00	60.64	59.16	58.56	58.84
7	46	70	78.72	71.49	66.05	62.14	59.50	57.86	56.95	56.51	56.28
8	47	91	77.84	71.19	67.24	65.29	64.75	65.14	66.08	67.30	68.63
9	48	54	79.16	75.15	74.37	75.57	77.88	80.66	83.53	86.26	88.76
10	49	60	76.64	70.92	68.26	66.94	65.94	64.66	62.86	60.45	57.48
11	50	48	74.98	68.74	65.78	64.17	62.97	61.87	60.86	60.09	59.75
12	51	35	72.28	64.59	60.45	57.70	55.48	53.55	51.86	50.42	49.17
13	52	49	68.55	58.67	52.81	48.62	45.24	42.42	40.06	38.08	36.42
14	53	44	66.60	56.74	51.67	48.77	47.12	46.37	46.32	46.82	47.74
15	54	61	64.34	54.19	49.37	46.86	45.56	44.95	44.70	44.56	44.37
16	55	68	64.00	55.55	52.86	52.52	53.28	54.58	56.11	57.71	59.34
17	56	82	64.40	58.04	57.40	58.71	60.64	62.63	64.43	65.94	67.13
18	57	71	66.16	62.83	64.78	68.03	71.32	74.25	76.73	78.79	80.51
19	58	50	66.65	64.47	66.65	69.22	71.16	72.30	72.72	72.56	71.95
20	59		64.98	61.57	61.65	61.53	60.58	58.92	56.82	54.51	52.20
21		MAD	19.33	17.16	16.15	15.36	14.93	14.71	14.72	14.88	15.36
22		MSE	496.07	390.84	359.18	346.56	340.77	338.41	339.03	343.32	352.38
23		RMSE	22.273	19.77	18.952	18.616	18.46	18.396	18.413	18.529	18.772
24		MAPE	38.28%	32.71%	30.12%	28.36%	27.54%	27.09%	27.09%	27.38%	28.23%

Forecasting Models With Linear Trends

- Double Moving Average
- Double Exponential Smoothing
- Based on the linear trend equation

$$F_{t+k} = a_t + b_t k$$

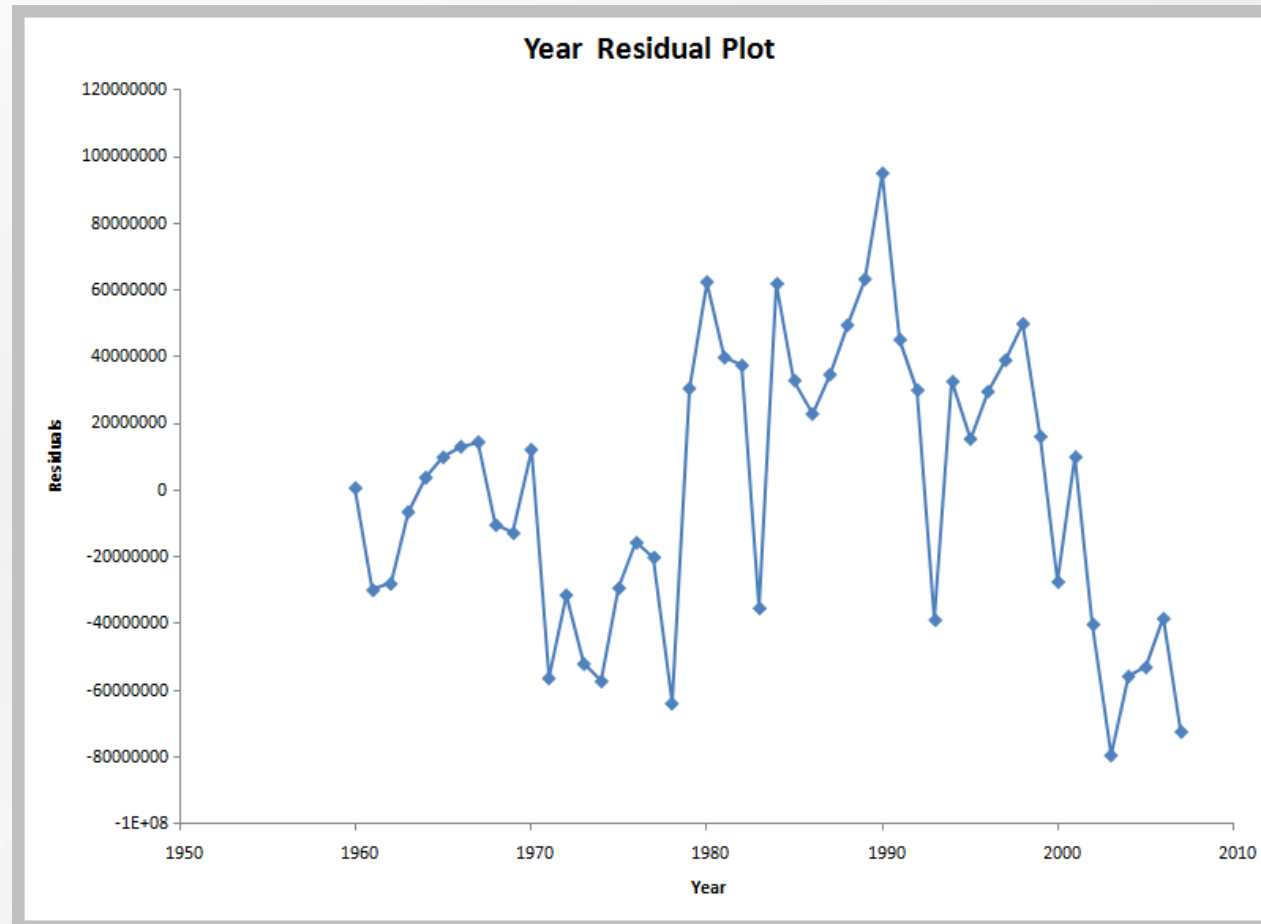
Regression-Based Forecasting



A forecast for 2008 would be:

$$\text{Tons} = 416,896,322.7 + 16,685,398.57 (49) = 1,234,480,853$$

Autocorrelated Data



Autoregressive Models

- First-order autoregressive model

- $Y_i = a_0 + a_1 Y_{i-1} + \delta_i$

- Second-order autoregressive model

- $Y_i = a_0 + a_1 Y_{i-1} + a_2 Y_{i-2} + \delta_i$

Portion of Coal Production File for Autoregressive Modeling

	A	B	C	D	E
1	Coal Production				
2					
3	Year	Total Tons	Year - 1	Year - 2	Year - 3
4	1960	434,329,000			
5	1961	420,423,000	434,329,000		
6	1962	439,043,000	420,423,000	434,329,000	
7	1963	477,195,000	439,043,000	420,423,000	434,329,000
8	1964	504,182,000	477,195,000	439,043,000	420,423,000
9	1965	526,954,000	504,182,000	477,195,000	439,043,000
10	1966	546,822,000	526,954,000	504,182,000	477,195,000

Third-Order Autoregressive Model

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	<i>Regression Statistics</i>						
4	Multiple R	0.988073439					
5	R Square	0.97628912					
6	Adjusted R Square	0.974554178					
7	Standard Error	35247696.59					
8	Observations	45					
9							
10	ANOVA						
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
12	Regression	3	2.09738E+18	6.99125E+17	562.721337	2.49811E-33	
13	Residual	41	5.09384E+16	1.2424E+15			
14	Total	44	2.14831E+18				
15							
16		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
17	Intercept	50332718.33	20917609.66	2.406236618	0.020709694	8088749.075	92576687.59
18	Year - 1	0.565292456	0.154196228	3.666058917	0.000700825	0.25388686	0.876698052
19	Year - 2	0.247260179	0.174243467	1.41904993	0.163442473	-0.104631637	0.599151996
20	Year - 3	0.156914847	0.151580327	1.035192694	0.306646344	-0.14920783	0.463037525

Second-Order Autogressive Model

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	<i>Regression Statistics</i>						
4	Multiple R	0.988567258					
5	R Square	0.977265224					
6	Adjusted R Square	0.976207792					
7	Standard Error	34986337.08					
8	Observations	46					
9							
10	ANOVA						
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
12	Regression	2	2.26249E+18	1.13125E+18	924.1877675	4.66498E-36	
13	Residual	43	5.26339E+16	1.22404E+15			
14	Total	45	2.31513E+18				
15							
16		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
17	Intercept	43087157.38	19471726.51	2.212806213	0.032269724	3818678.935	82355635.83
18	Year - 1	0.63224491	0.142019472	4.451818489	5.95528E-05	0.345835353	0.918654467
19	Year - 2	0.341258327	0.141128508	2.418067984	0.019912562	0.056645569	0.625871085

Forecasting Models With Seasonality

- Additive model

$$F_{t+k} = a_t + S_{t-s+k}$$

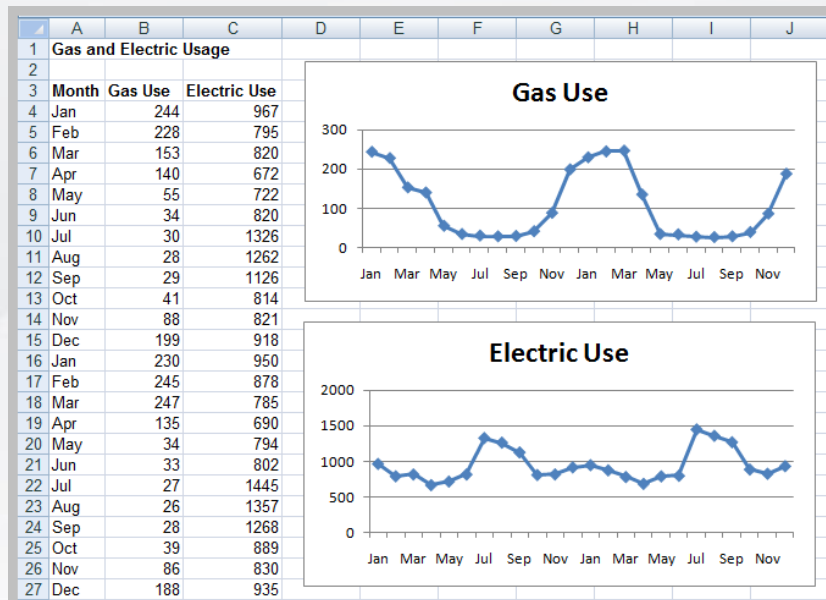
- Multiplicative model

$$F_{t+k} = a_t S_{t-s+k}$$

Incorporating Seasonality into Regression Models

- Use dummy variables. Example:

$$\begin{aligned} \text{Gas usage} = & \beta_0 + \beta_1 \text{Time} + \beta_2 \text{February} + \beta_3 \text{March} + \beta_4 \text{April} + \beta_5 \text{May} \\ & + \beta_6 \text{June} + \beta_7 \text{July} + \beta_8 \text{August} + \beta_9 \text{September} + \beta_{10} \text{October} \\ & + \beta_{11} \text{November} + \beta_{12} \text{December} \end{aligned}$$



Data Matrix

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
3	Month	Gas Use	Time	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
4	Jan	244	1	0	0	0	0	0	0	0	0	0	0	0
5	Feb	228	2	1	0	0	0	0	0	0	0	0	0	0
6	Mar	153	3	0	1	0	0	0	0	0	0	0	0	0
7	Apr	140	4	0	0	1	0	0	0	0	0	0	0	0
8	May	55	5	0	0	0	1	0	0	0	0	0	0	0
9	Jun	34	6	0	0	0	0	1	0	0	0	0	0	0
10	Jul	30	7	0	0	0	0	0	1	0	0	0	0	0
11	Aug	28	8	0	0	0	0	0	0	1	0	0	0	0
12	Sep	29	9	0	0	0	0	0	0	0	1	0	0	0
13	Oct	41	10	0	0	0	0	0	0	0	0	1	0	0
14	Nov	88	11	0	0	0	0	0	0	0	0	0	1	0
15	Dec	199	12	0	0	0	0	0	0	0	0	0	0	1
16	Jan	230	13	0	0	0	0	0	0	0	0	0	0	0
17	Feb	245	14	1	0	0	0	0	0	0	0	0	0	0
18	Mar	247	15	0	1	0	0	0	0	0	0	0	0	0
19	Apr	135	16	0	0	1	0	0	0	0	0	0	0	0
20	May	34	17	0	0	0	1	0	0	0	0	0	0	0
21	Jun	33	18	0	0	0	0	1	0	0	0	0	0	0
22	Jul	27	19	0	0	0	0	0	1	0	0	0	0	0
23	Aug	26	20	0	0	0	0	0	0	1	0	0	0	0
24	Sep	28	21	0	0	0	0	0	0	0	1	0	0	0
25	Oct	39	22	0	0	0	0	0	0	0	0	1	0	0
26	Nov	86	23	0	0	0	0	0	0	0	0	0	1	0
27	Dec	188	24	0	0	0	0	0	0	0	0	0	0	1

Regression Model Results

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	<i>Regression Statistics</i>						
4	Multiple R	0.985480895					
5	R Square	0.971172595					
6	Adjusted R Square	0.948997667					
7	Standard Error	19.54432831					
8	Observations	24					
9							
10	<i>ANOVA</i>						
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
12	Regression	10	167292.2083	16729.22083	43.79597661	2.33344E-08	
13	Residual	13	4965.75	381.9807692			
14	Total	23	172257.9583				
15							
16		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
17	Intercept	236.75	9.772164157	24.22697738	3.33921E-12	215.6385229	257.8614771
18	Mar	-36.75	16.92588482	-2.171230656	0.04901621	-73.31615098	-0.183849024
19	Apr	-99.25	16.92588482	-5.863799799	5.55744E-05	-135.816151	-62.68384902
20	May	-192.25	16.92588482	-11.35834268	4.02824E-08	-228.816151	-155.683849
21	Jun	-203.25	16.92588482	-12.00823485	2.07264E-08	-239.816151	-166.683849
22	Jul	-208.25	16.92588482	-12.30364038	1.54767E-08	-244.816151	-171.683849
23	Aug	-209.75	16.92588482	-12.39226204	1.41949E-08	-246.316151	-173.183849
24	Sep	-208.25	16.92588482	-12.30364038	1.54767E-08	-244.816151	-171.683849
25	Oct	-196.75	16.92588482	-11.62420766	3.05791E-08	-233.316151	-160.183849
26	Nov	-149.75	16.92588482	-8.847395666	7.30451E-07	-186.316151	-113.183849
27	Dec	-43.25	16.92588482	-2.555257847	0.023953114	-79.81615098	-6.683849024

Model

Gas Usage = 236.75 – 36.75 March – 99.25 April – 192.25 May – 203.25 June
– 208.25 July – 209.75 August – 208.25 September – 196.75 October
– 149.75 November – 43.25 December

Models for Trend and Seasonality

- Holt-Winters Additive Model

$$F_{t+1} = a_t + b_t + S_{t-s+1}$$

- Holt-Winters Multiplicative Model

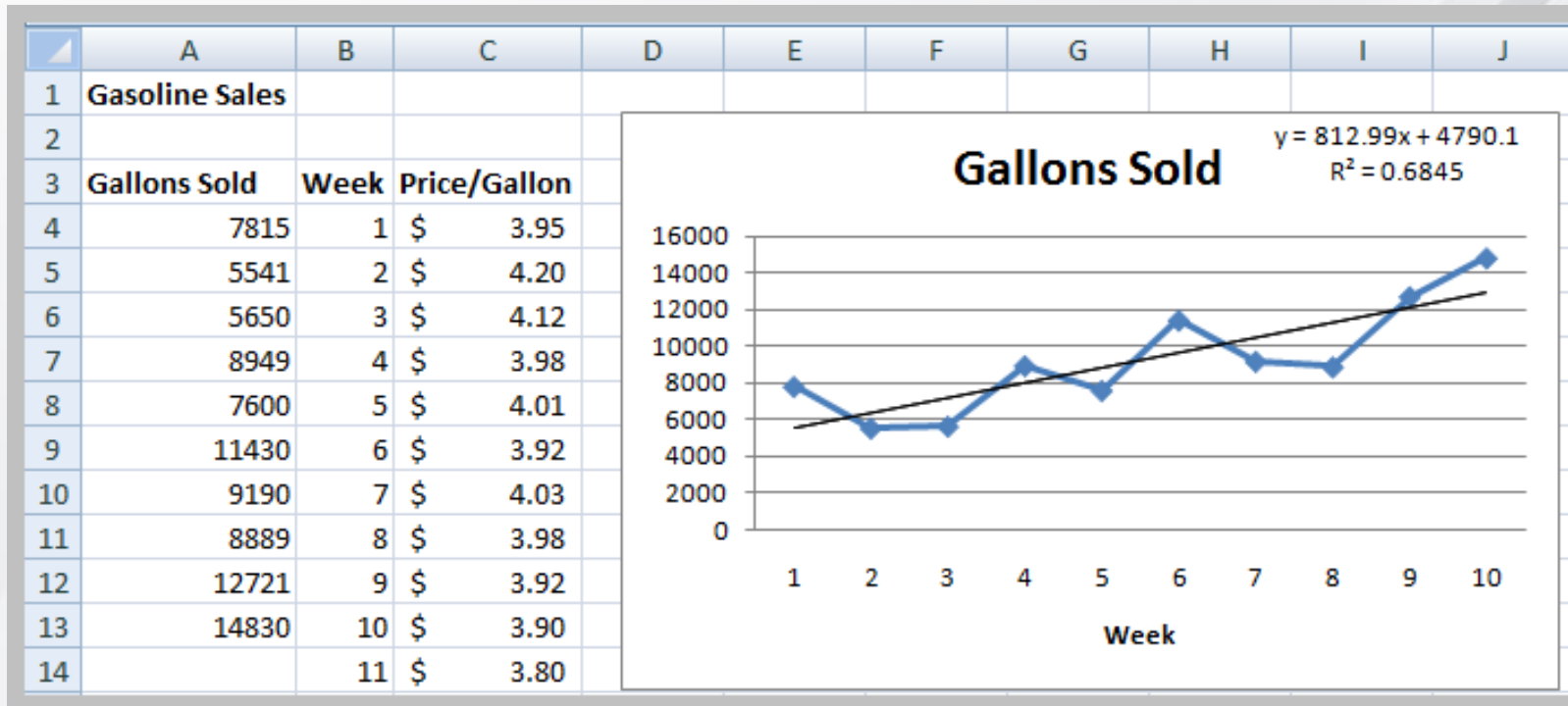
$$F_{t+1} = (a_t + b_t) S_{t-s+1}$$

Model Choice

TABLE 7.3 Forecasting Model Choice

	No Seasonality	Seasonality
No Trend	Single moving average or single exponential smoothing	Seasonal additive or seasonal multiplicative model
Trend	Double moving average or double exponential smoothing	Holt–Winters additive or Holt–Winters multiplicative model

Regression Forecasting With Causal Variables



Sales (week 11) = $4790.1 + 812.00(11) = 13,733$ gallons

Causal Model

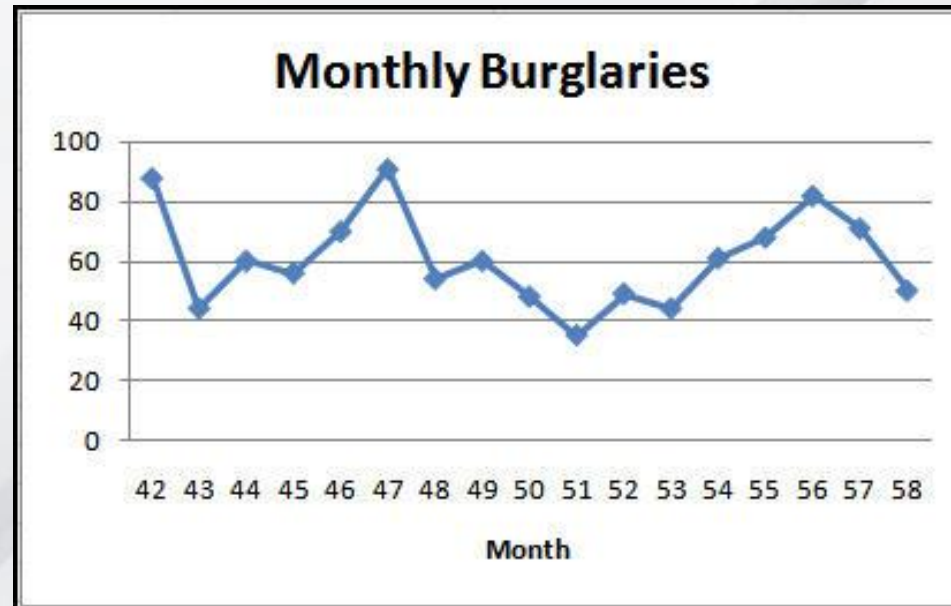
- Sales = $\beta_0 + \beta_1 \text{Week} + \beta_2 \text{Price/Gallon}$
- Sales = 72333.08 + 508.67 Week – 16463.20 Price/Gallon

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	Regression Statistics						
4	Multiple R	0.930528528					
5	R Square	0.865883342					
6	Adjusted R Square	0.827564297					
7	Standard Error	1235.400329					
8	Observations	10					
9							
10	ANOVA						
11		df	SS	MS	F	Significance F	
12	Regression	2	68974748.7	34487374.35	22.59668368	0.000883465	
13	Residual	7	10683497.8	1526213.972			
14	Total	9	79658246.5				
15							
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
17	Intercept	72333.08447	21969.92267	3.292368642	0.013259225	20382.47253	124283.6964
18	Week	508.6681395	168.1770861	3.024598364	0.019260863	110.9925233	906.3437558
19	Price/Gallon	-16463.19901	5351.082403	-3.076611005	0.017900405	-29116.49823	-3809.899789

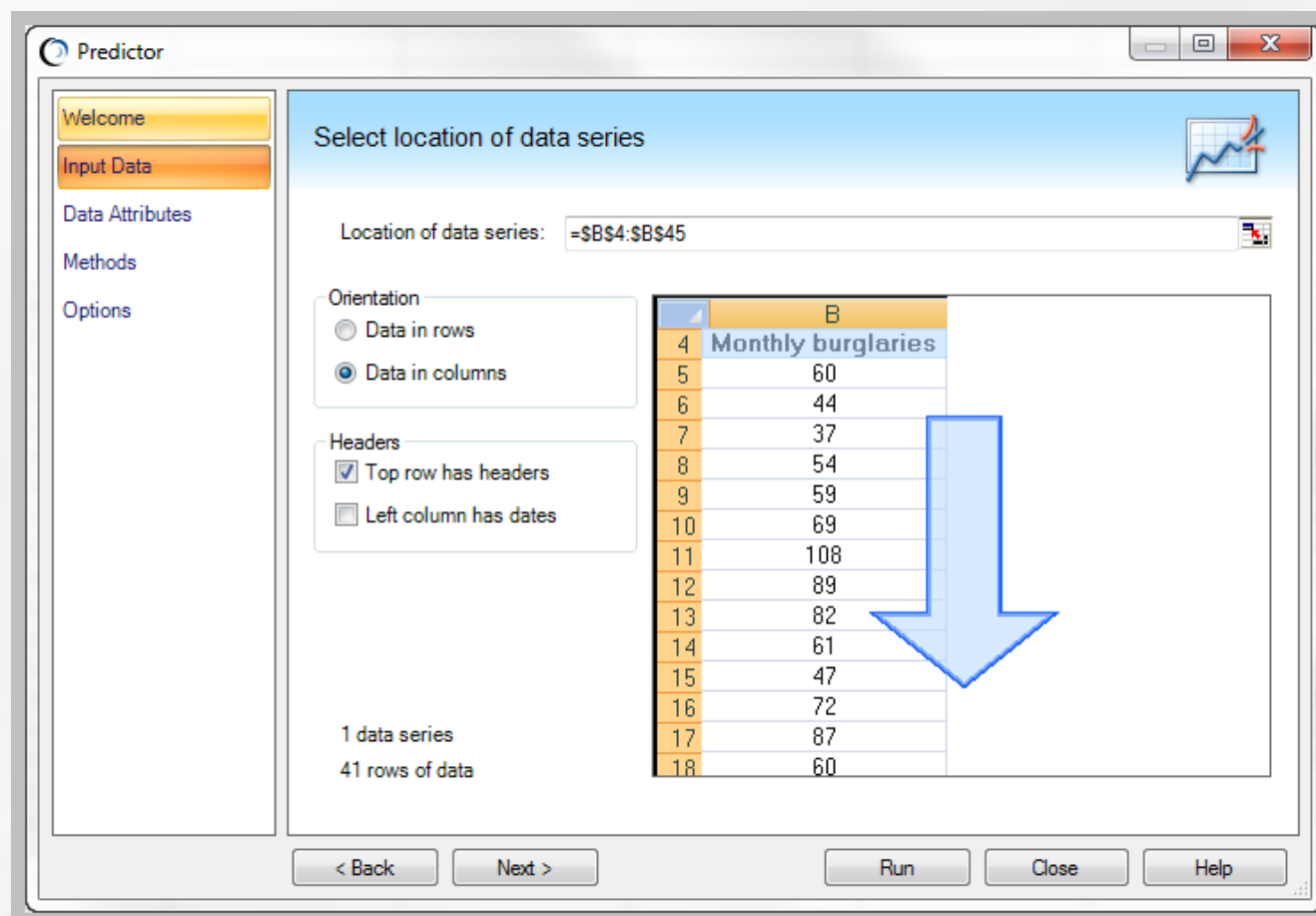
Sales (week 11) = 72333.08 + 508.67(11) - 16463.2(3.80) = 13,733 gallons

CB Predictor

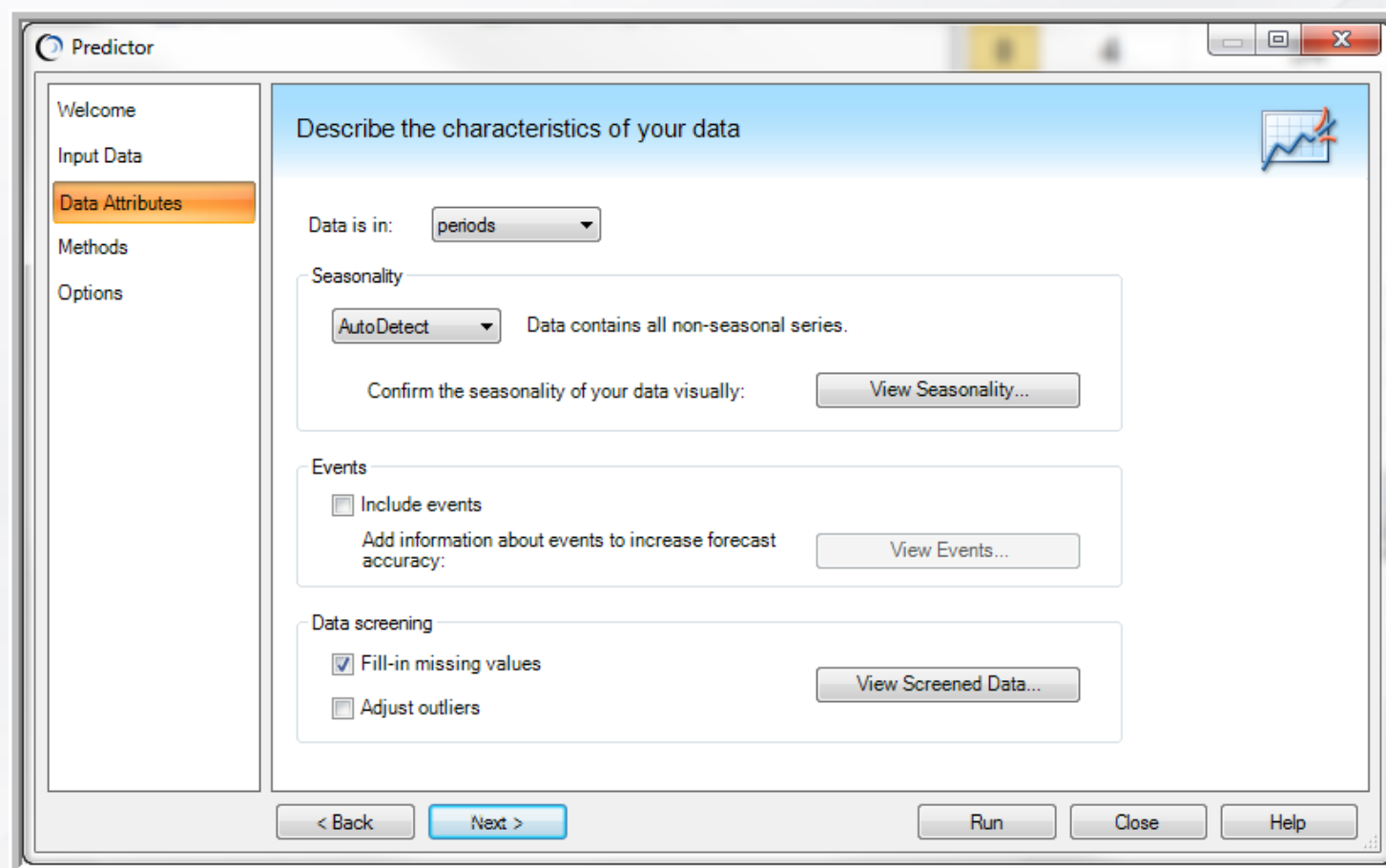
- Excel add-in for forecasting
- Integrated with *Crystal Ball* software
- Example: Burglary Data



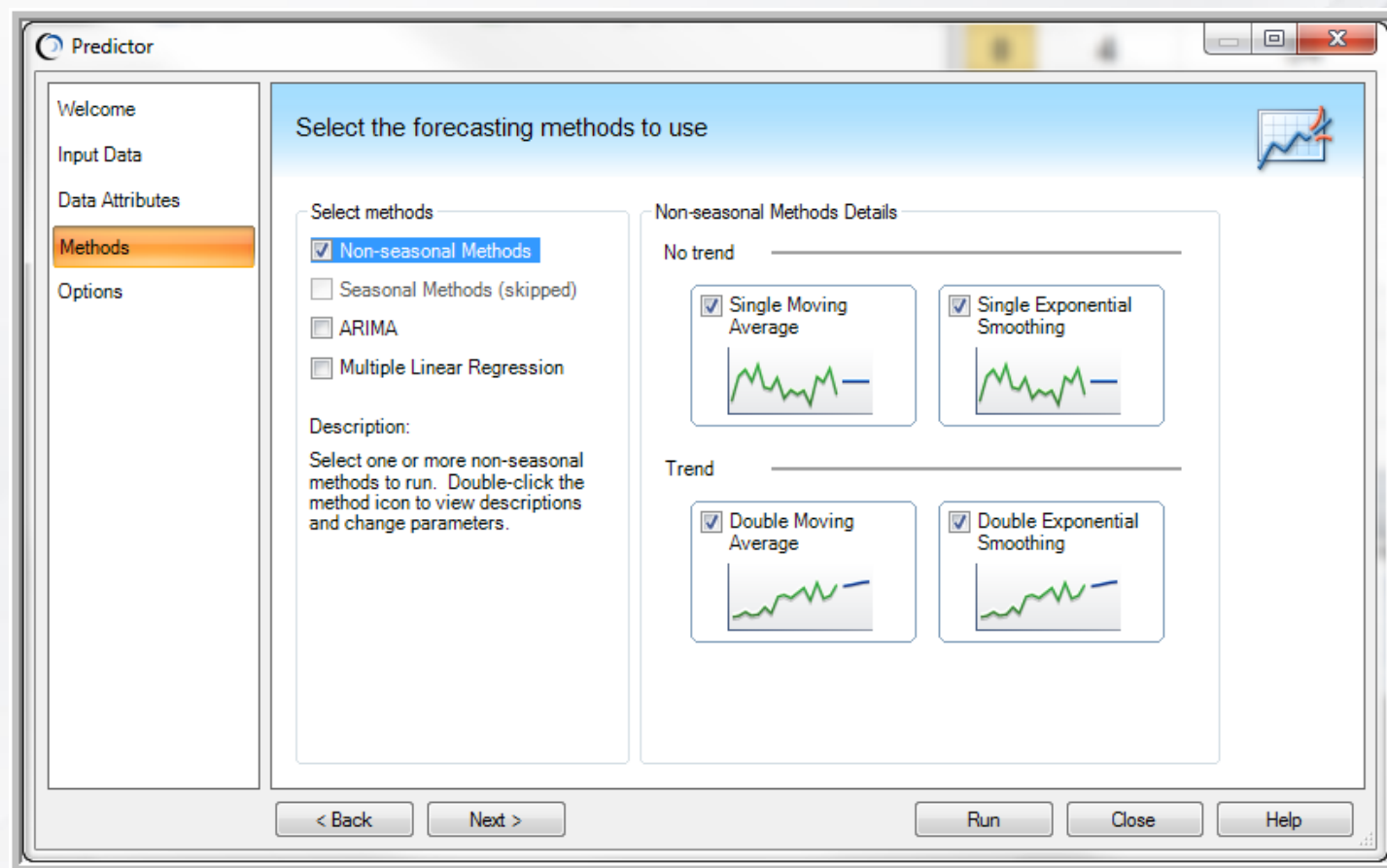
Input Data Dialog



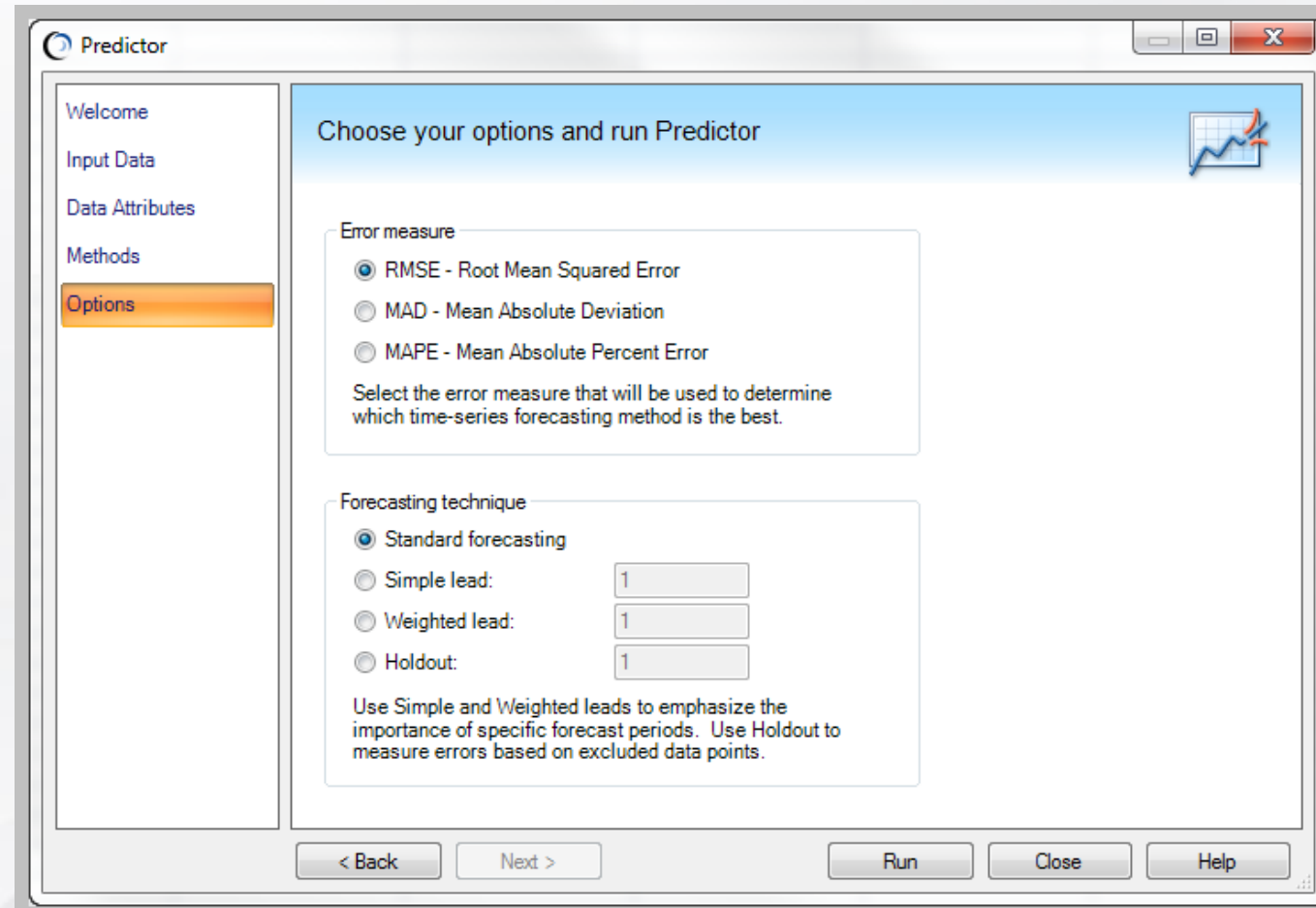
Data Attributes Dialog



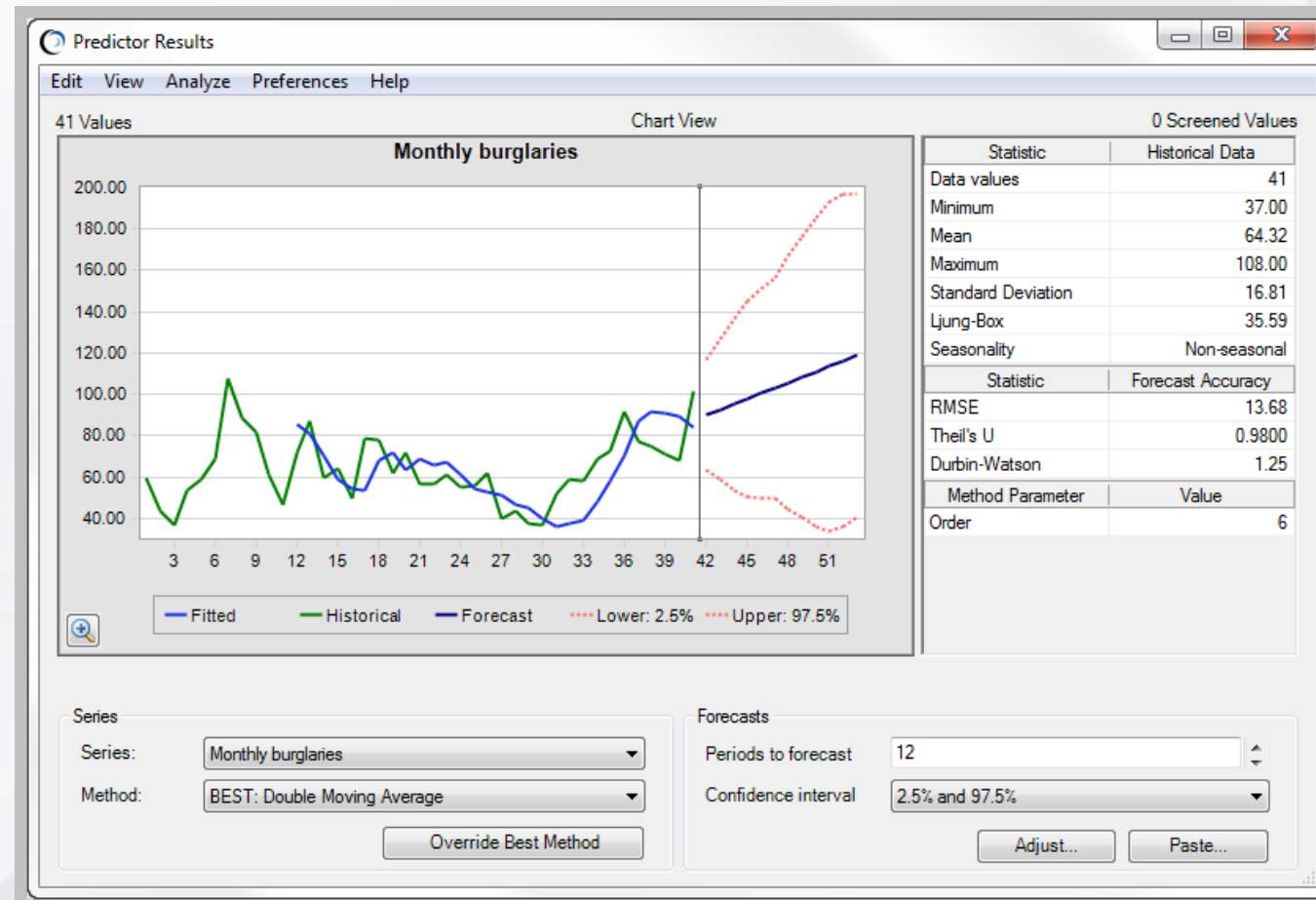
Methods Dialog



Options Dialog



CB Predictor Results for Forecasting Burglaries



CB Predictor Table View

Predictor Results

Edit View Analyze Preferences Help

41 Values Table View 0 Screened Values

Period	Historical Data	Lower: 2.5%	Fit & Forecast	Upper: 97.5%	Residuals
40	68.00		89.74		-21.74
41	102.00		83.78		18.22
42		63.23	90.05	116.87	
43		58.67	92.68	126.70	
44		53.71	95.32	136.92	
45		50.60	97.95	145.30	
46		50.11	100.58	151.06	
47		50.05	103.22	156.38	
48		44.85	105.85	166.85	
49		40.59	108.48	176.38	
50		36.47	111.12	185.76	
51		34.10	113.75	193.40	
52		35.93	116.38	196.84	
53		40.74	119.02	197.29	

Statistic	Historical Data
Data values	41
Minimum	37.00
Mean	64.32
Maximum	108.00
Standard Deviation	16.81
Ljung-Box	35.59
Seasonality	Non-seasonal

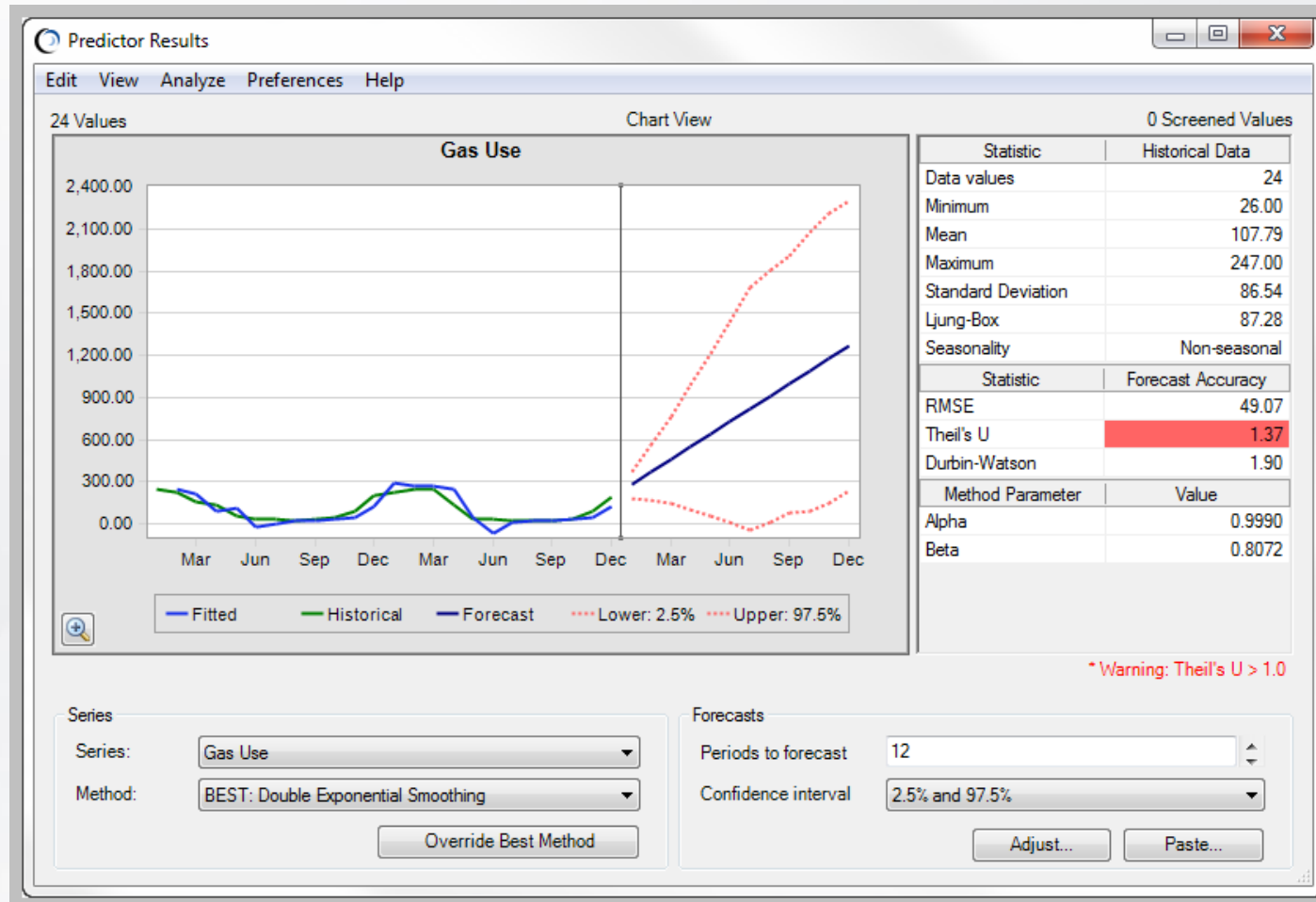
Statistic	Forecast Accuracy
RMSE	13.68
Theil's U	0.9800
Durbin-Watson	1.25

Method Parameter	Value
Order	6

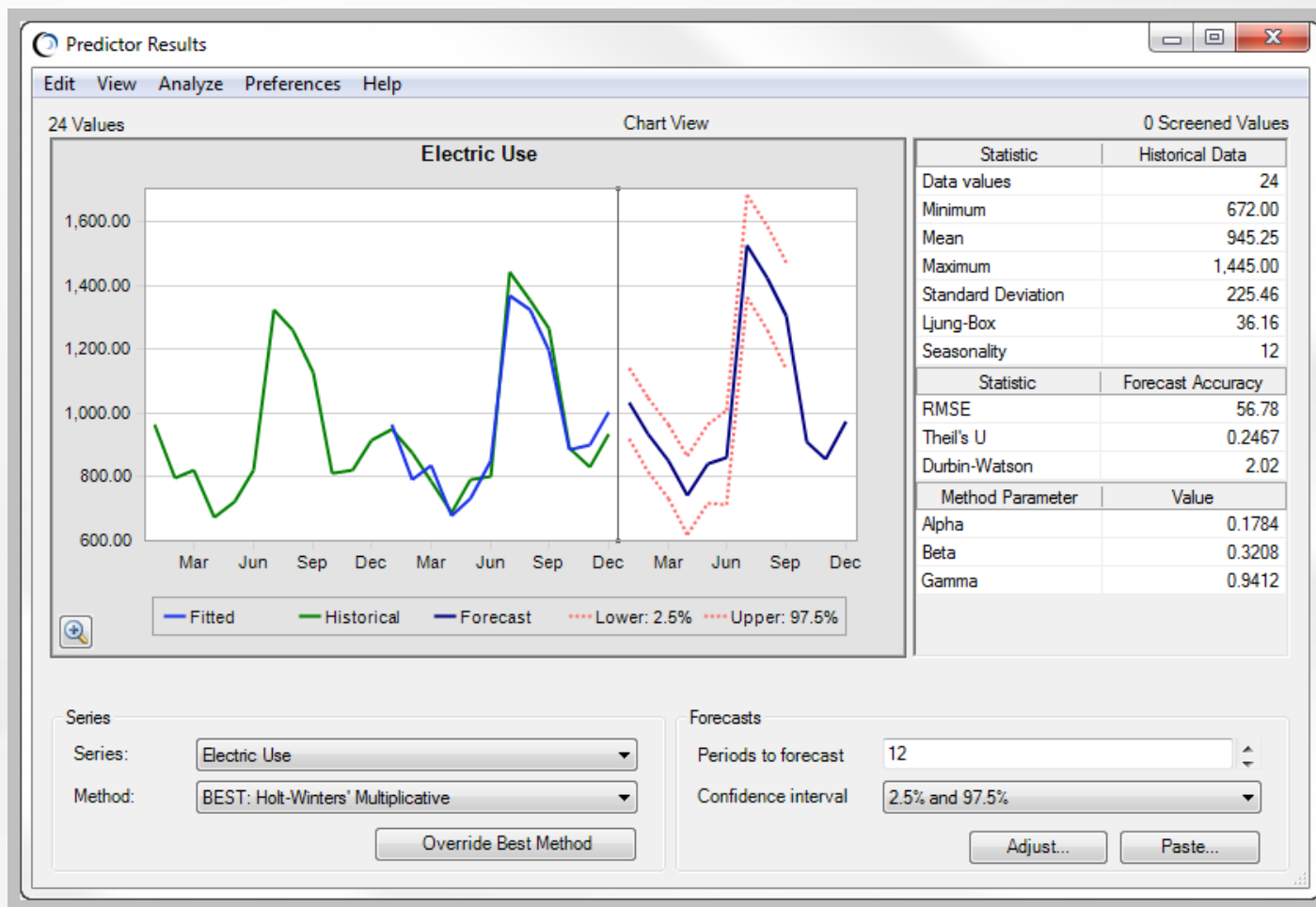
Series: Monthly burglaries
Method: BEST: Double Moving Average
Override Best Method

Forecasts: Periods to forecast: 12
Confidence interval: 2.5% and 97.5%
Adjust... Paste...

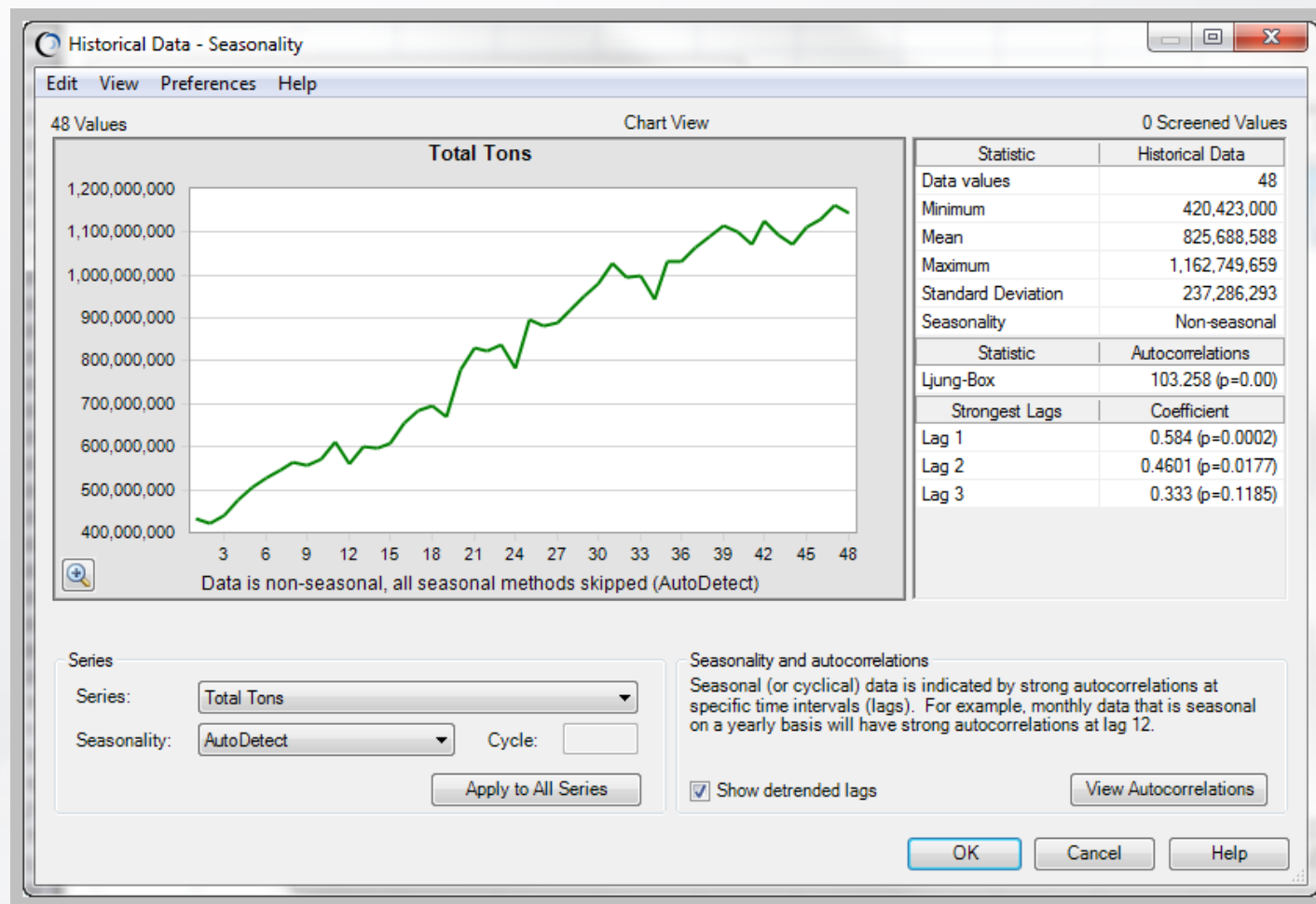
CB Predictor Results for Gas Use Example



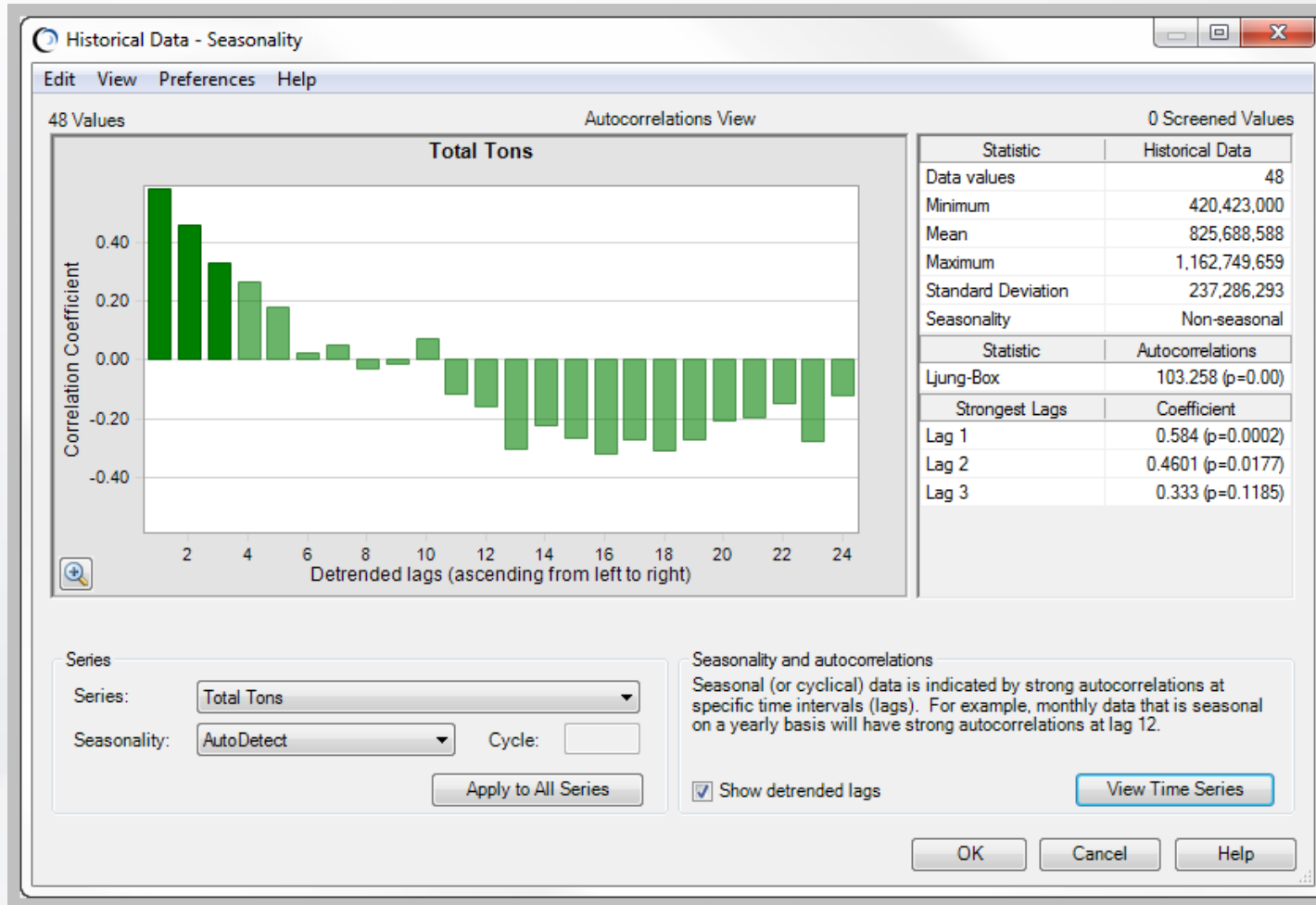
CB Predictor Results for Electric Use Example



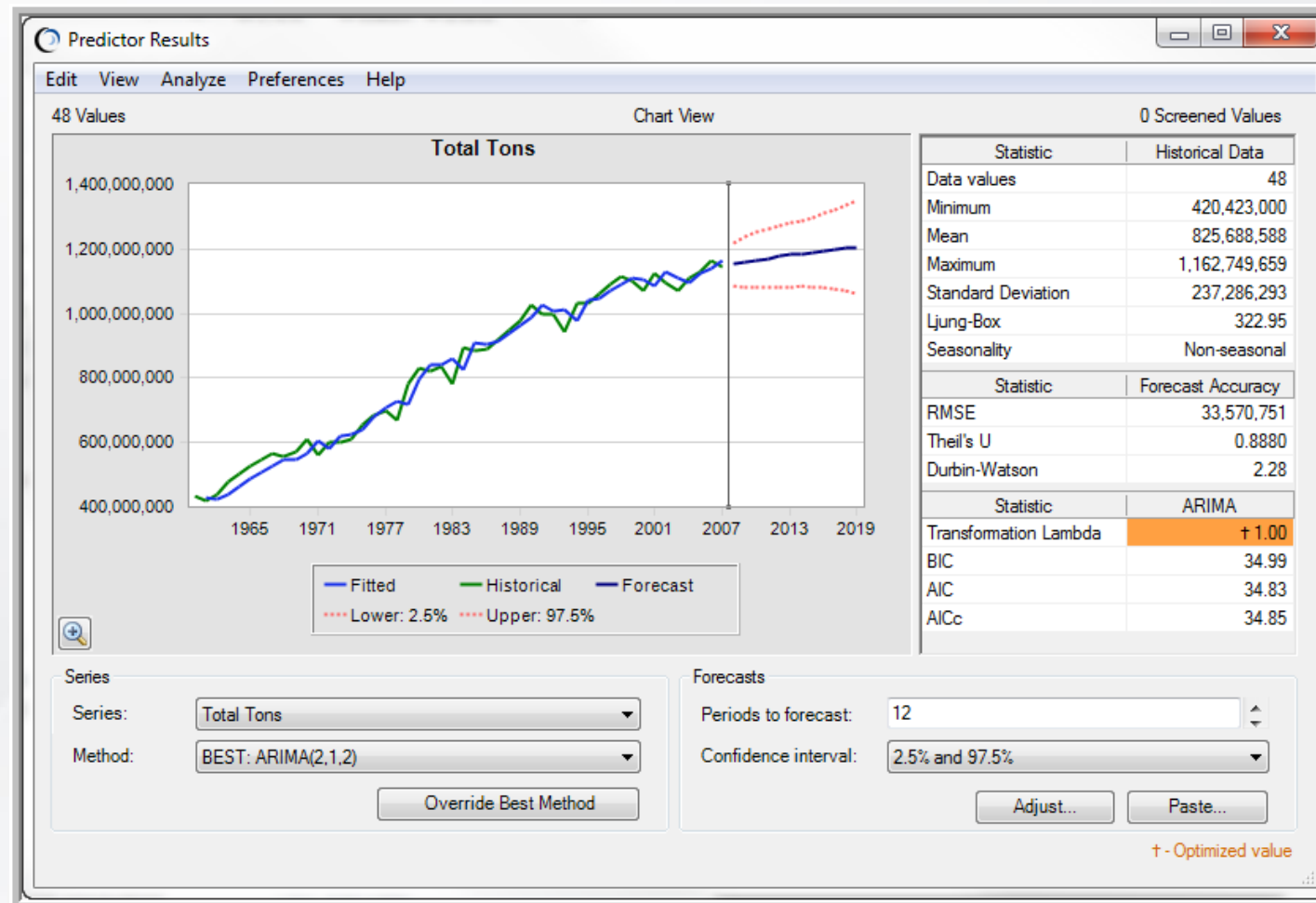
CB Predictor Seasonality Dialog



Autocorrelations in *CB Predictor*



CB Predictor ARIMA Model for Coal Production



Summary

- Forecasting Techniques
- Qualitative and Judgmental Methods
- Indicators and Indexes
- Example Dept. of Commerce Index of Leading Indicators
- Time Series
- Statistical Forecasting Methods
- Error Metrics and Forecast Accuracy

Review Questions

1. What are the components of a time series?
2. What does the Delphi Method involve?
3. Name three statistical forecasting methods mentioned in the document.



Thank You