Applications of Dynamic Regression Models in Business and Industry

By

Robert M. Lucas, PhD.

Introduction

Dynamic regression models extend multiple regression models by allowing for independent variables to be incorporated as leading indicators of the dependent variable and to account for autocorrelation of the dependent variable. Dynamic regression models can be built in the JMP® Time Series Platform.

Manufacturing processes data is often a time series with process input characteristics and the final output quality varying over time. A dynamic regression model may be used to ascertain how to adjust process parameters to control the variability of process output quality.

In business, market mix models are used to evaluate the return on investment of different advertising strategies. A dynamic regression model can be used to evaluate the return on investment of advertising spend by different media, a market mix model. One can use the model to more efficiently allocate advertising budgets.

Background

A simple example of a multiple linear regression model is given by the following equation.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

The model assumes that the mean of the response given the independent or input variables can be represented by a linear combination of the inputs. It also assumes that the deviations of the observed values from the mean represented by the error term are independent and normally distributed with a mean zero and constant variance.

A comparable time series regression model is represented by the following equation.

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t$$

The only difference in notation is the inclusion of the subscript t. However, this implies that the order of the data is important. The data must be sorted from oldest to the most recent observation and must have equally spaced time intervals between the observations, a time series.

Dynamic regression models are an extension of multiple linear regression. One extension is to accommodate the situation when the deviations from the mean are not independent but correlated. The other extension is to have more complicated relations between an input and the response. The equation below represents both of these extensions.

$$(1 - \theta_1 B)(Y_t - \beta_0 - \beta_1 X_{1t} - \beta_2 X_{1(t-1)}) = \varepsilon_t$$

The Backward Shift Operator, B, is defined as

$$B(Z_t) = Z_{t-1}$$

JMP® uses the equivalent expression below.

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{1(t-1)} = (1/(1 - \theta_1 B))\varepsilon_t$$

The response depends on both the current value of the input and on the previous value of the input. The response also depends on the previous value of itself. Consequently the term autocorrelation, correlation with itself.

Models like these can be fit in the JMP® **Time Series Platform** using the **Transfer Function Specification** window. The term transfer function is natural. The function describes how variation in the input is transferred to the response.

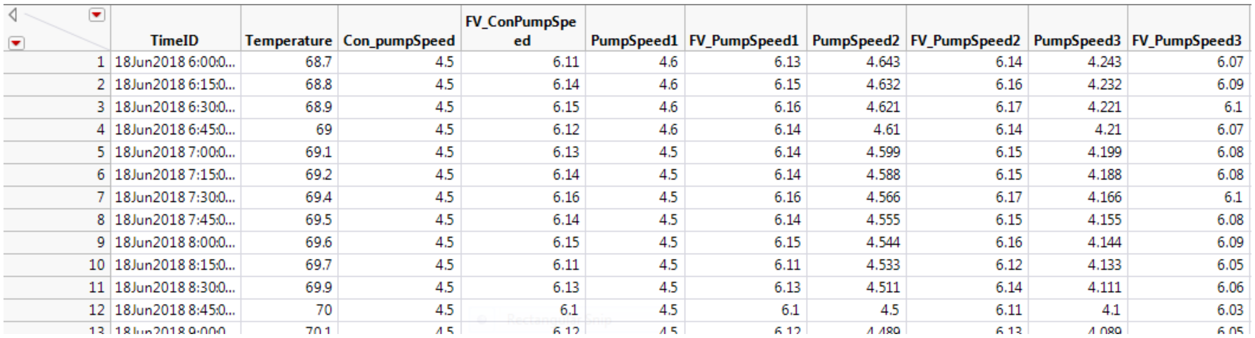
Manufacturing Process Description

A pharmaceutical manufacture of over-the-counter medications sells a product in 6 fluid ounce bottles. Because the viscosity of the product depends on ambient temperature, bottles tend to contain more product when temperatures are higher, and less when cooler. The space is conditioned but the temperature still varies by time of day. Fill volume can be adjusted by controlling the speed of the pump. The filling line fills 12 bottles about every five seconds. The target is set about 6.15 ounces to essentially eliminate the chance of under fill.

This demo illustrates how one can quantify the relationship among temperature, pump speed, and fill volume. Once quantified, the variability of fill volume can be reduced by adjusting the pump speed to compensate for changes in temperature that affect the viscosity of the product. With less variability, the target fill volume can be reduced with very small chance of under fill. Hence, substantial saving can be attained by reducing the average fill volume.

Manufacturing Process Data Demonstration.

A subset of the manufacturing process data table is show below.

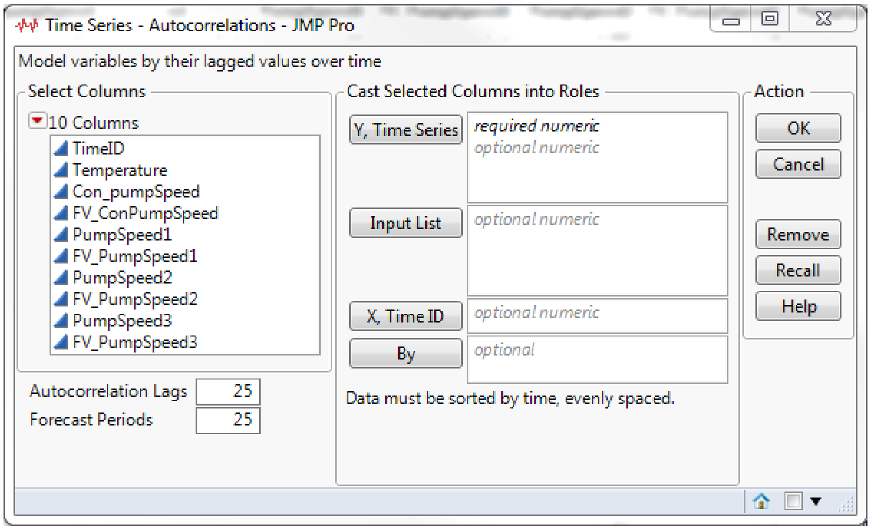


	TimeID	Temperature	Con_pumpSpeed	FV_ConPumpSpeed	PumpSpeed1	FV_PumpSpeed1	PumpSpeed2	FV_PumpSpeed2	PumpSpeed3	FV_PumpSpeed3
1	18Jun2018 6:00:0...	68.7	4.5	6.11	4.6	6.13	4.643	6.14	4.243	6.07
2	18Jun2018 6:15:0...	68.8	4.5	6.14	4.6	6.15	4.632	6.16	4.232	6.09
3	18Jun2018 6:30:0...	68.9	4.5	6.15	4.6	6.16	4.621	6.17	4.221	6.1
4	18Jun2018 6:45:0...	69	4.5	6.12	4.6	6.14	4.61	6.14	4.21	6.07
5	18Jun2018 7:00:0...	69.1	4.5	6.13	4.5	6.14	4.599	6.15	4.199	6.08
6	18Jun2018 7:15:0...	69.2	4.5	6.14	4.5	6.14	4.588	6.15	4.188	6.08
7	18Jun2018 7:30:0...	69.4	4.5	6.16	4.5	6.16	4.566	6.17	4.166	6.1
8	18Jun2018 7:45:0...	69.5	4.5	6.14	4.5	6.14	4.555	6.15	4.155	6.08
9	18Jun2018 8:00:0...	69.6	4.5	6.15	4.5	6.15	4.544	6.16	4.144	6.09
10	18Jun2018 8:15:0...	69.7	4.5	6.11	4.5	6.11	4.533	6.12	4.133	6.05
11	18Jun2018 8:30:0...	69.9	4.5	6.13	4.5	6.13	4.511	6.14	4.111	6.06
12	18Jun2018 8:45:0...	70	4.5	6.1	4.5	6.1	4.5	6.11	4.1	6.03
13	18Jun2018 9:00:0	70.1	4.5	6.12	4.5	6.12	4.489	6.13	4.089	6.05

The data table contains nine columns. The first column is the TimeID, a JMP® date-time value in 15 minute increments. Temperature is degrees Fahrenheit of the plant. The column FV_ConPumpSpeed is the fill volume in fluid ounces of the product when the pump speed does not vary. The pump speed is gallons per minute. To ascertain the effect of pump speed and temperature, pump speed was varied as a function of temperature as shown in the PumpSpeed1 column. The measured fill volumes as given in the FV_PumpSpeed1 column. The PumpSpeed2 column gives the adjusted pump speed based on the results of the experiment. FV_PumpSpeed2 gives the fill volumes for PumpSpeed2. Because the variability of the process was reduced, the average pumps speed can be reduced as shown in the PumpSpeed3 columns. The FV_pumpSpped3 column contains the measured pump speed for PumpSpeed3.

The above data are simulated but realistic based on the actual process. The realism was confirmed by my anonymous source of the process details.

To open the **Time Series Platform**, select **Analyz**e, then **Specialized Modeling**, and then **Time Series**. The **Time Series** dialog opens.



Time Series - Autocorrelations - JMP Pro

Model variables by their lagged values over time

Y, Time Series: required numeric, optional numeric

Input List: optional numeric

X, Time ID: optional numeric

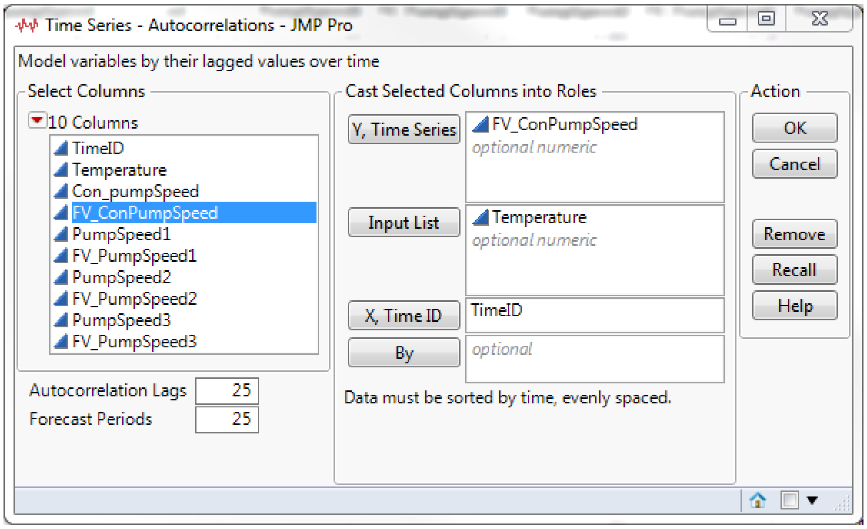
By: optional

Autocorrelation Lags: 25

Forecast Periods: 25

Data must be sorted by time, evenly spaced.

Assign the columns to their roles as show below.



Time Series - Autocorrelations - JMP Pro

Model variables by their lagged values over time

Y, Time Series: FV_ConPumpSpeed, optional numeric

Input List: Temperature, optional numeric

X, Time ID: TimeID

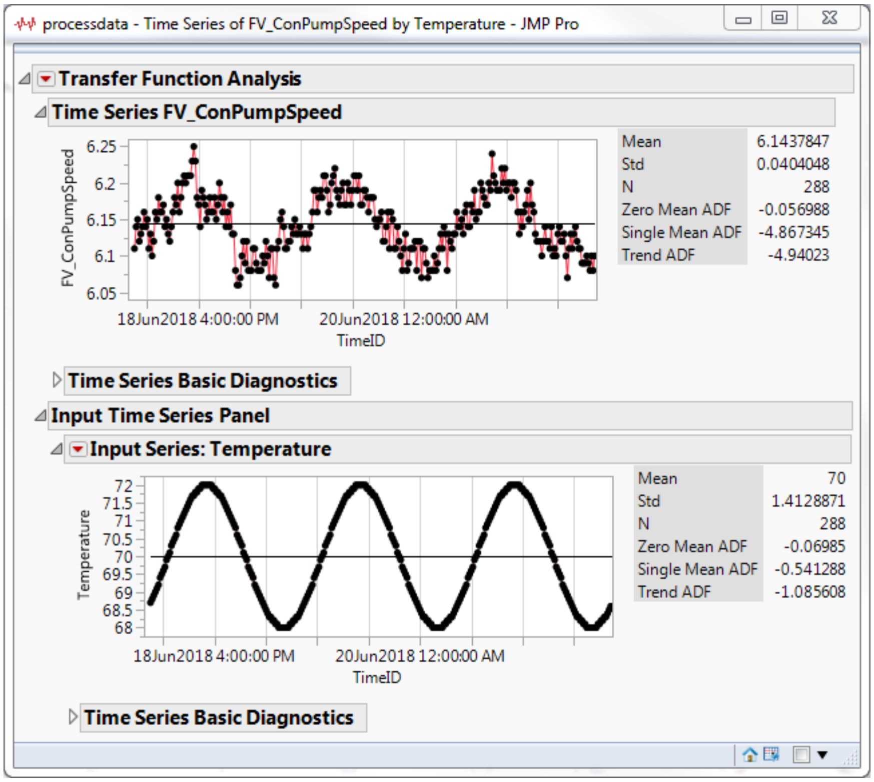
By: optional

Autocorrelation Lags: 25

Forecast Periods: 25

Data must be sorted by time, evenly spaced.

After selecting **OK**, the following graphs appear.

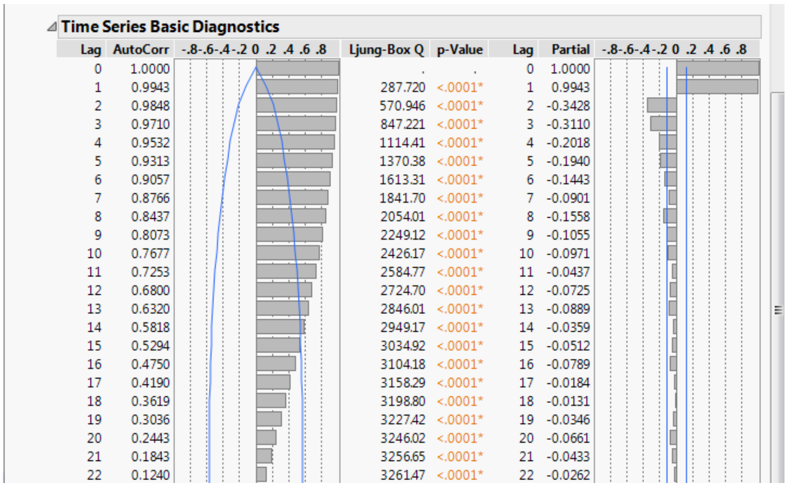

processdata - Time Series of FV_ConPumpSpeed by Temperature - JMP Pro
Transfer Function Analysis
Time Series FV_ConPumpSpeed
FV_ConPumpSpeed
6.25
6.2
6.15
6.1
6.05
18Jun2018 4:00:00 PM
20Jun2018 12:00:00 AM
TimeID
Time Series Basic Diagnostics
Input Time Series Panel
Input Series: Temperature
Temperature
72
71.5
71
70.5
70
69.5
69
68.5
68
18Jun2018 4:00:00 PM
20Jun2018 12:00:00 AM
TimeID
Time Series Basic Diagnostics

Mean	6.1437847
Std	0.0404048
N	288
Zero Mean ADF	-0.056988
Single Mean ADF	-4.867345
Trend ADF	-4.94023

Mean	70
Std	1.4128871
N	288
Zero Mean ADF	-0.06985
Single Mean ADF	-0.541288
Trend ADF	-1.085608

Obviously, fill volume and temperature are highly correlated. The fill volume is highest when the temperature is warmest.

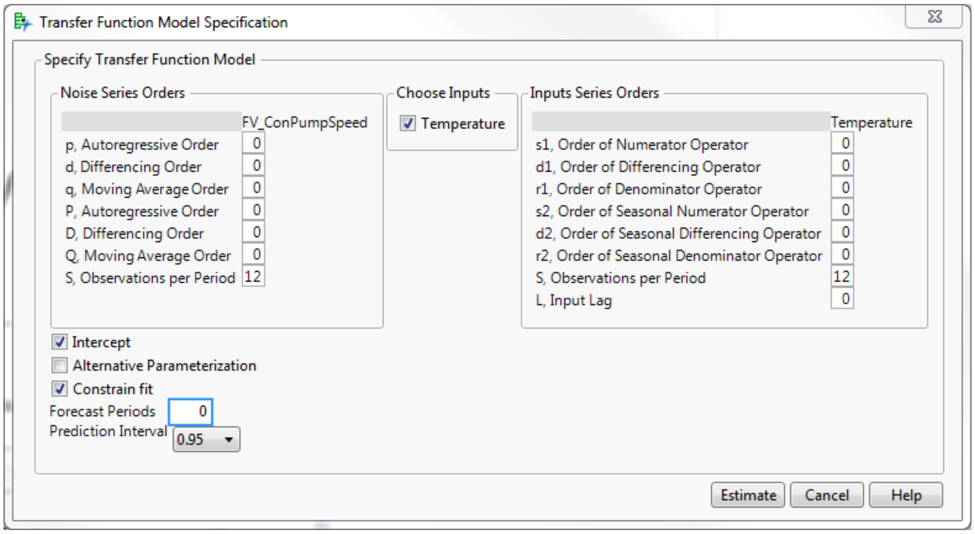
Expand the **Time Series Basic Diagnostics**.


Time Series Basic Diagnostics

Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.9943		287.720	<.0001*	1	0.9943	
2	0.9848		570.946	<.0001*	2	-0.3428	
3	0.9710		847.221	<.0001*	3	-0.3110	
4	0.9532		1114.41	<.0001*	4	-0.2018	
5	0.9313		1370.38	<.0001*	5	-0.1940	
6	0.9057		1613.31	<.0001*	6	-0.1443	
7	0.8766		1841.70	<.0001*	7	-0.0901	
8	0.8437		2054.01	<.0001*	8	-0.1558	
9	0.8073		2249.12	<.0001*	9	-0.1055	
10	0.7677		2426.17	<.0001*	10	-0.0971	
11	0.7253		2584.77	<.0001*	11	-0.0437	
12	0.6800		2724.70	<.0001*	12	-0.0725	
13	0.6320		2846.01	<.0001*	13	-0.0889	
14	0.5818		2949.17	<.0001*	14	-0.0359	
15	0.5294		3034.92	<.0001*	15	-0.0512	
16	0.4750		3104.18	<.0001*	16	-0.0789	
17	0.4190		3158.29	<.0001*	17	-0.0184	
18	0.3619		3198.80	<.0001*	18	-0.0131	
19	0.3036		3227.42	<.0001*	19	-0.0346	
20	0.2443		3246.02	<.0001*	20	-0.0661	
21	0.1843		3256.65	<.0001*	21	-0.0433	
22	0.1240		3261.47	<.0001*	22	-0.0262	

The autocorrelation and partial autocorrelation plots are displayed. These plots are useful to identify candidate models that account for the autocorrelation that is often present in time series data. However, because of the high correlation with temperature, it is premature at this time to construct models.

To fit a linear regression model with temperature as the independent variable, from the red triangle beside **Transfer Function Analysis** drop down menu select **Transfer Function**. The following window appears.



Specify Transfer Function Model

Noise Series Orders

	FV_ConPumpSpeed
p, Autoregressive Order	0
d, Differencing Order	0
q, Moving Average Order	0
P, Autoregressive Order	0
D, Differencing Order	0
Q, Moving Average Order	0
S, Observations per Period	12

Choose Inputs

☑ Temperature

Inputs Series Orders

	Temperature
s1, Order of Numerator Operator	0
d1, Order of Differencing Operator	0
r1, Order of Denominator Operator	0
s2, Order of Seasonal Numerator Operator	0
d2, Order of Seasonal Differencing Operator	0
r2, Order of Seasonal Denominator Operator	0
S, Observations per Period	12
L, Input Lag	0

☑ Intercept

□ Alternative Parameterization

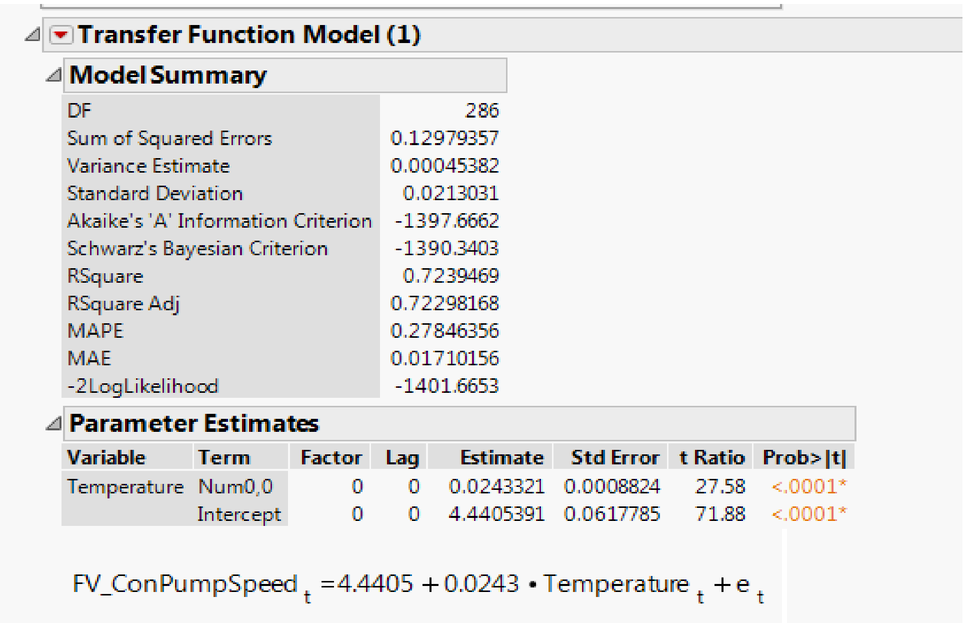
☑ Constrain fit

Forecast Periods 0

Prediction Interval 0.95

Select **Estimate.**

The model summary is shown below.



Transfer Function Model (1)

Model Summary

DF	286
Sum of Squared Errors	0.12979357
Variance Estimate	0.00045382
Standard Deviation	0.0213031
Akaike's 'A' Information Criterion	-1397.6662
Schwarz's Bayesian Criterion	-1390.3403
RSquare	0.7239469
RSquare Adj	0.72298168
MAPE	0.27846356
MAE	0.01710156
-2LogLikelihood	-1401.6653

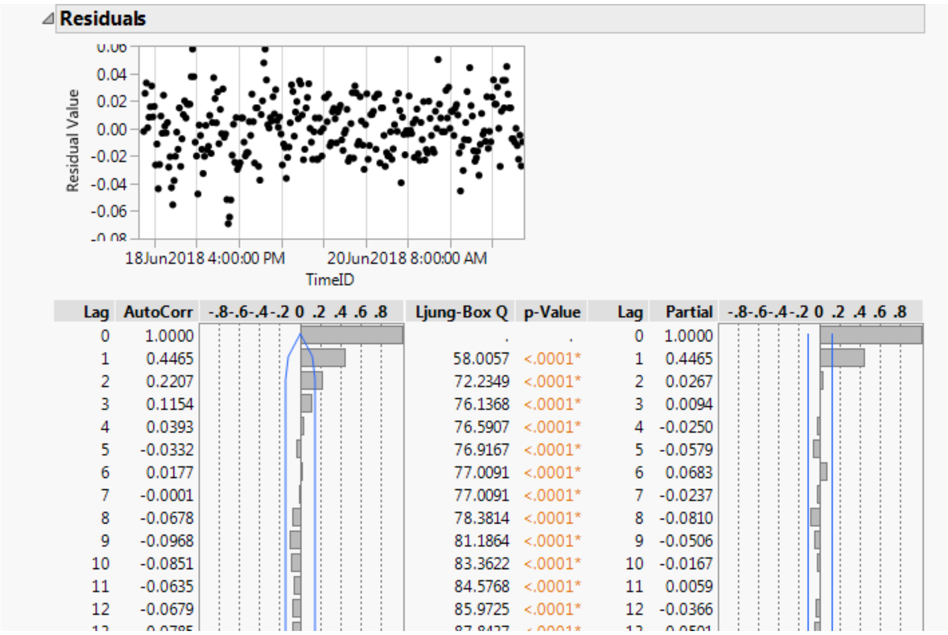
Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Temperature	Num0,0	0	0	0.0243321	0.0008824	27.58	<.0001*
	Intercept	0	0	4.4405391	0.0617785	71.88	<.0001*

$$FV\_ConPumpSpeed_t = 4.4405 + 0.0243 \cdot Temperature_t + e_t$$

The slope for Temperature is 0.0243 gallons per minute. For every increase of one degree, the average fill volume increases by 0.0243 ounces.

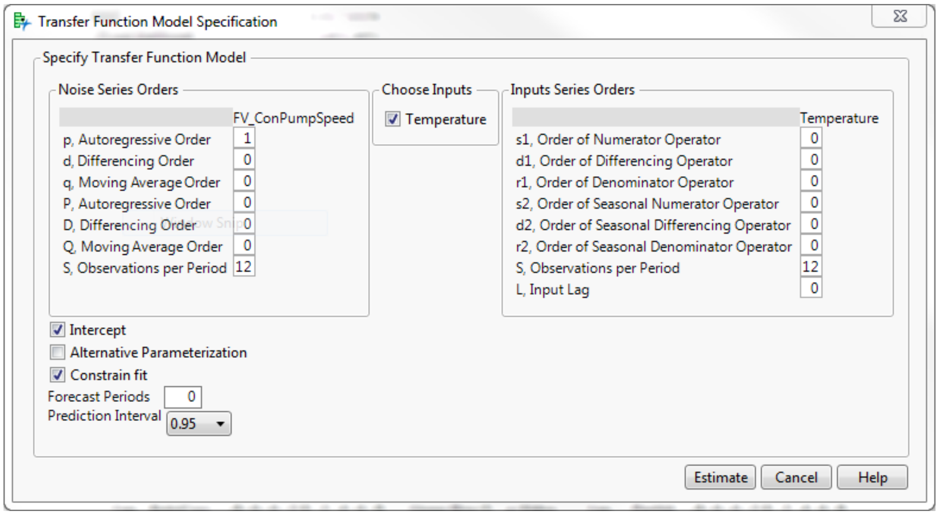
Examine the **Residuals**.



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.4465	58.0057	<.0001*	1	0.4465
2	0.2207	72.2349	<.0001*	2	0.0267
3	0.1154	76.1368	<.0001*	3	0.0094
4	0.0393	76.5907	<.0001*	4	-0.0250
5	-0.0332	76.9167	<.0001*	5	-0.0579
6	0.0177	77.0091	<.0001*	6	0.0683
7	-0.0001	77.0091	<.0001*	7	-0.0237
8	-0.0678	78.3814	<.0001*	8	-0.0810
9	-0.0968	81.1864	<.0001*	9	-0.0506
10	-0.0851	83.3622	<.0001*	10	-0.0167
11	-0.0635	84.5768	<.0001*	11	0.0059
12	-0.0679	85.9725	<.0001*	12	-0.0366
13	-0.0785	87.8427	<.0001*	13	-0.0501

There is no obvious pattern in the residual plot. However, the autocorrelation and partial autocorrelation plots suggest that an autoregressive term of order 1 (AR(1)) is necessary to model the autocorrelation in the residuals.

From the red triangle drop down menu beside **Transfer Function Analysis**, select **Transfer Function** again. Enter a one (1) for the **Autoregressive Order** as show below.



Transfer Function Model Specification

Specify Transfer Function Model

Noise Series Orders

	FV_ConPumpSpeed
p, Autoregressive Order	1
d, Differencing Order	0
q, Moving Average Order	0
P, Autoregressive Order	0
D, Differencing Order	0
Q, Moving Average Order	0
S, Observations per Period	12

Choose Inputs

☑ Temperature

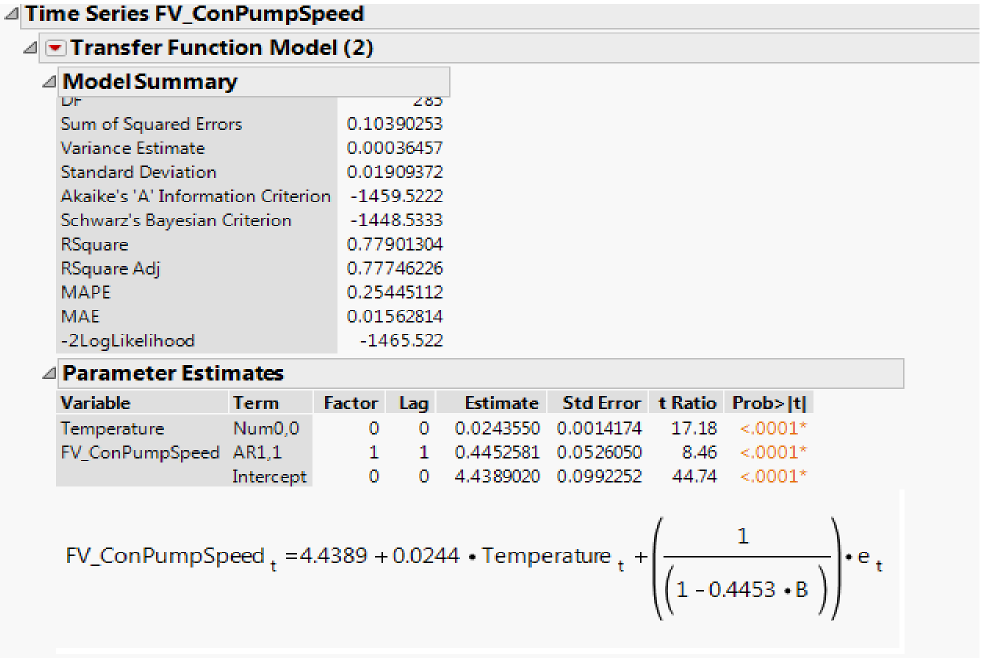
Inputs Series Orders

	Temperature
s1, Order of Numerator Operator	0
d1, Order of Differencing Operator	0
r1, Order of Denominator Operator	0
s2, Order of Seasonal Numerator Operator	0
d2, Order of Seasonal Differencing Operator	0
r2, Order of Seasonal Denominator Operator	0
S, Observations per Period	12
L, Input Lag	0

☑ Intercept
☐ Alternative Parameterization
☑ Constrain fit
Forecast Periods 0
Prediction Interval 0.95

Select **Estimate**.

The second model summary is shown below.



Time Series FV_ConPumpSpeed

Transfer Function Model (2)

Model Summary

DF	285
Sum of Squared Errors	0.10390253
Variance Estimate	0.00036457
Standard Deviation	0.01909372
Akaike's 'A' Information Criterion	-1459.5222
Schwarz's Bayesian Criterion	-1448.5333
RSquare	0.77901304
RSquare Adj	0.77746226
MAPE	0.25445112
MAE	0.01562814
-2LogLikelihood	-1465.522

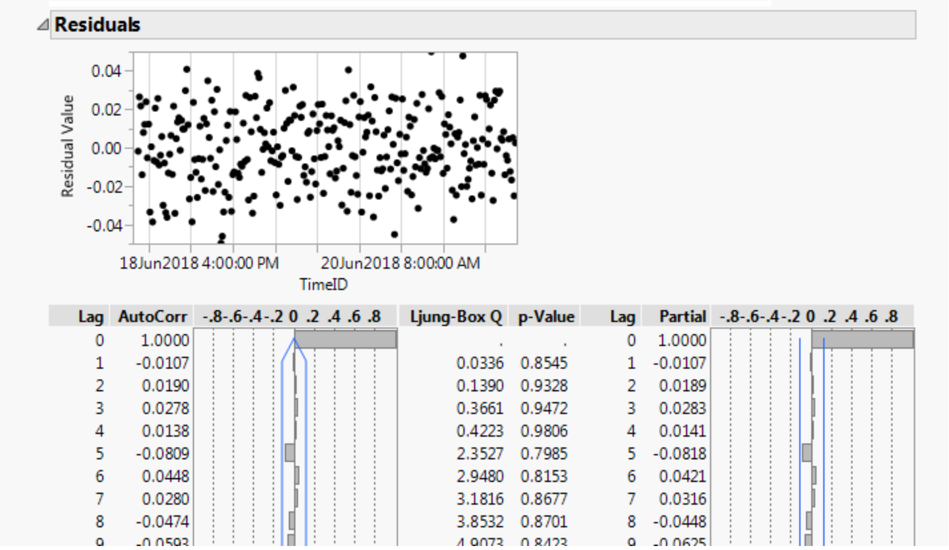
Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Temperature	Num0,0	0	0	0.0243550	0.0014174	17.18	<.0001*
FV_ConPumpSpeed	AR1,1	1	1	0.4452581	0.0526050	8.46	<.0001*
	Intercept	0	0	4.4389020	0.0992252	44.74	<.0001*

$$FV\_ConPumpSpeed_t = 4.4389 + 0.0244 \cdot Temperature_t + \left(\frac{1}{\left(1 - 0.4453 \cdot B\right)}\right) \cdot e_t$$

The factor in the last term reflects the AR(1) term in the model. The current value of fill volume depends on temperature but also the fill volume in the previous time interval.

Examine the **Residuals**.

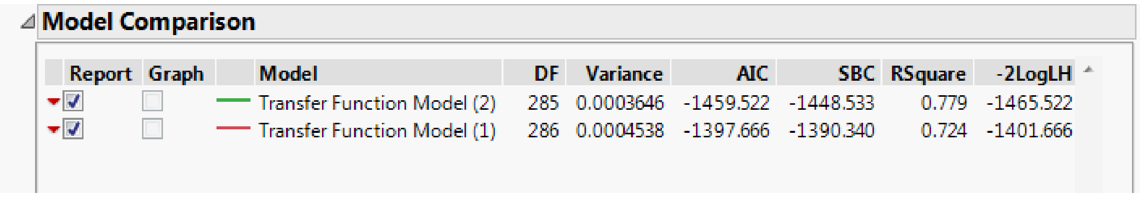


Residuals

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	-0.0107	0.0336	0.8545	1	-0.0107
2	0.0190	0.1390	0.9328	2	0.0189
3	0.0278	0.3661	0.9472	3	0.0283
4	0.0138	0.4223	0.9806	4	0.0141
5	-0.0809	2.3527	0.7985	5	-0.0818
6	0.0448	2.9480	0.8153	6	0.0421
7	0.0280	3.1816	0.8677	7	0.0316
8	-0.0474	3.8532	0.8701	8	-0.0448
9	-0.0593	4.9073	0.8423	9	-0.0625

These plots suggest that the AR(1) term is adequate for modeling the autocorrelation in the data.

Now examine the **Model Comparison** Table.

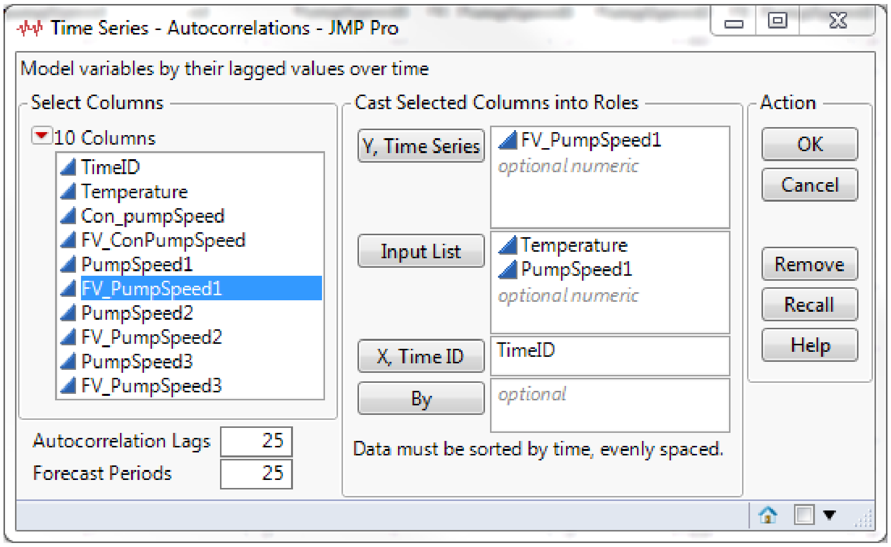


Model Comparison

Report	Graph	Model	DF	Variance	AIC	SBC	RSquare	-2LogLH
☑	□	🟩 Transfer Function Model (2)	285	0.0003646	-1459.522	-1448.533	0.779	-1465.522
☑	□	🟥 Transfer Function Model (1)	286	0.0004538	-1397.666	-1390.340	0.724	-1401.666

By all the metrics, the second model is superior.

To ascertain how changing pump speed effects fill volume, bring up a new instance of the **Time Series Platform** and enter assign the variables to their roles as show below.



Time Series - Autocorrelations - JMP Pro

Model variables by their lagged values over time

Select Columns

10 Columns

- TimeID
- Temperature
- Con_pumpSpeed
- FV_ConPumpSpeed
- PumpSpeed1
- FV_PumpSpeed1
- PumpSpeed2
- FV_PumpSpeed2
- PumpSpeed3
- FV_PumpSpeed3

Autocorrelation Lags 25

Forecast Periods 25

Cast Selected Columns into Roles

Y, Time Series: FV_PumpSpeed1 *optional numeric*

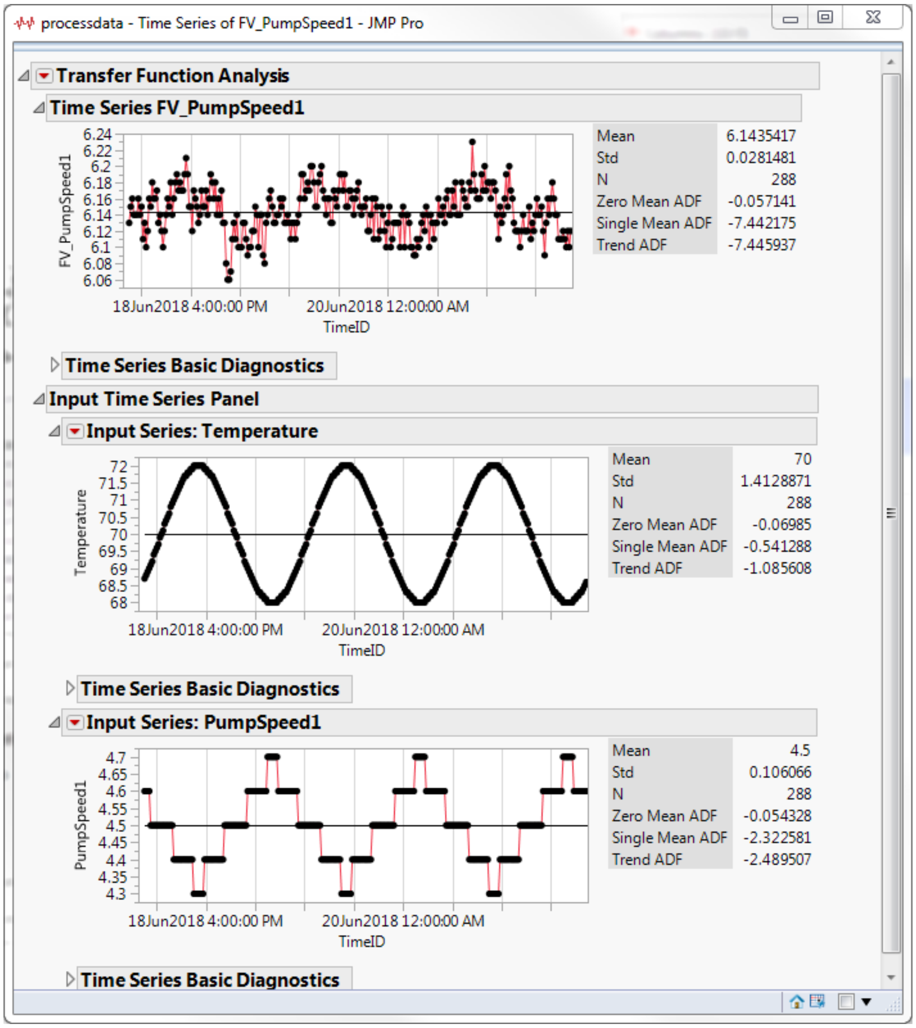
Input List: Temperature, PumpSpeed1 *optional numeric*

X, Time ID: TimeID

By: *optional*

Data must be sorted by time, evenly spaced.

Select **OK**. The following plots appear.



processdata - Time Series of FV_PumpSpeed1 - JMP Pro

Transfer Function Analysis

Time Series FV_PumpSpeed1

Mean	6.1435417
Std	0.0281481
N	288
Zero Mean ADF	-0.057141
Single Mean ADF	-7.442175
Trend ADF	-7.445937

Time Series Basic Diagnostics

Input Time Series Panel

Input Series: Temperature

Mean	70
Std	1.4128871
N	288
Zero Mean ADF	-0.06985
Single Mean ADF	-0.541288
Trend ADF	-1.085608

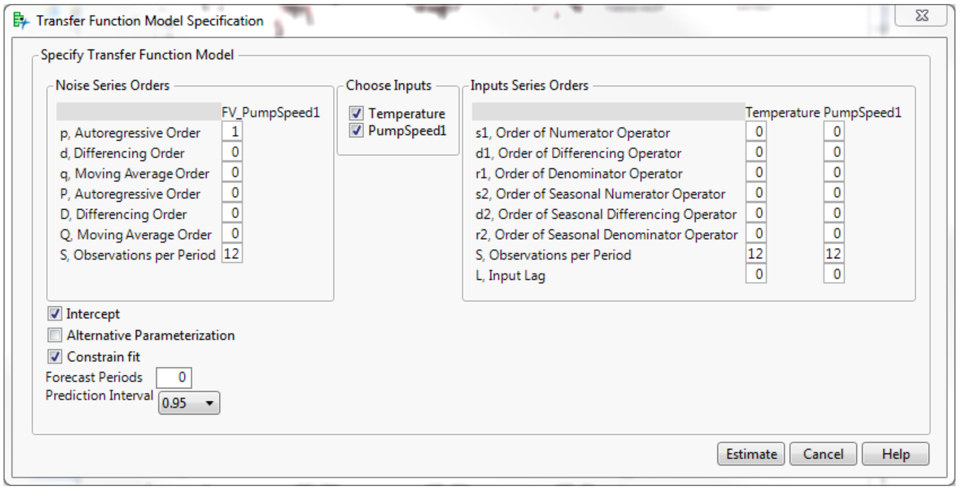
Time Series Basic Diagnostics

Input Series: PumpSpeed1

Mean	4.5
Std	0.106066
N	288
Zero Mean ADF	-0.054328
Single Mean ADF	-2.322581
Trend ADF	-2.489507

Time Series Basic Diagnostics

From the Red Triangle beside **Transfer Function Analysis** drop down menu, select **Transfer Function**. In the **Time Series Model Specification** window includes an AR(1) term as shown below.



Transfer Function Model Specification

Specify Transfer Function Model

Noise Series Orders

	FV_PumpSpeed1
p, Autoregressive Order	1
d, Differencing Order	0
q, Moving Average Order	0
P, Autoregressive Order	0
D, Differencing Order	0
Q, Moving Average Order	0
S, Observations per Period	12

Choose Inputs

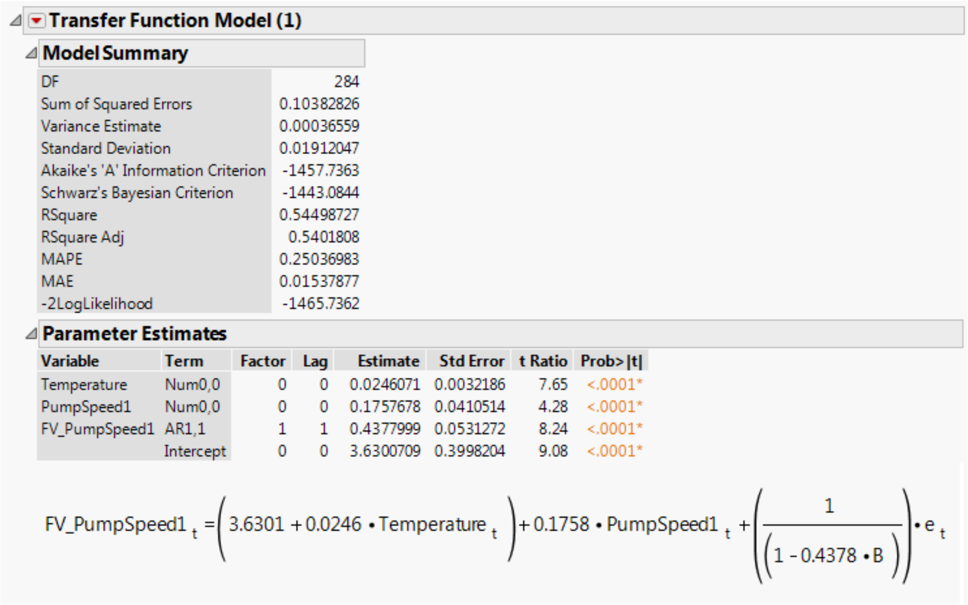
☑ Temperature
☑ PumpSpeed1

Inputs Series Orders

	Temperature	PumpSpeed1
s1, Order of Numerator Operator	0	0
d1, Order of Differencing Operator	0	0
r1, Order of Denominator Operator	0	0
s2, Order of Seasonal Numerator Operator	0	0
d2, Order of Seasonal Differencing Operator	0	0
r2, Order of Seasonal Denominator Operator	0	0
S, Observations per Period	12	12
L, Input Lag	0	0

☑ Intercept
☐ Alternative Parameterization
☑ Constrain fit
Forecast Periods 0
Prediction Interval 0.95

Select **Estimate**. The model summary appears.



Transfer Function Model (1)

Model Summary

DF	284
Sum of Squared Errors	0.10382826
Variance Estimate	0.00036559
Standard Deviation	0.01912047
Akaike's 'A' Information Criterion	-1457.7363
Schwarz's Bayesian Criterion	-1443.0844
RSquare	0.54498727
RSquare Adj	0.5401808
MAPE	0.25036983
MAE	0.01537877
-2LogLikelihood	-1465.7362

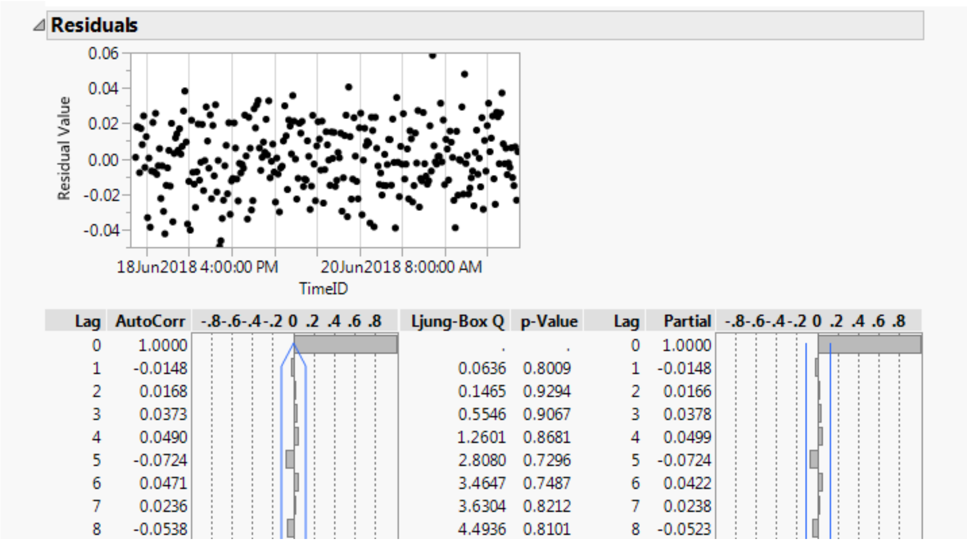
Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Temperature	Num0,0	0	0	0.0246071	0.0032186	7.65	<.0001*
PumpSpeed1	Num0,0	0	0	0.1757678	0.0410514	4.28	<.0001*
FV_PumpSpeed1	AR1,1	1	1	0.4377999	0.0531272	8.24	<.0001*
	Intercept	0	0	3.6300709	0.3998204	9.08	<.0001*

$$\text{FV\_PumpSpeed1}_t = \left(3.6301 + 0.0246 \bullet \text{Temperature}_t\right) + 0.1758 \bullet \text{PumpSpeed1}_t + \left(\frac{1}{\left(1 - 0.4378 \bullet B\right)}\right) \bullet e_t$$

The model indicates that for every one unit increase in temperature fill volume increases by 0.0246 fluid ounces and for every one unit increase in pump speed, fill volume increases by 0.1758 ounces.

Examine the **Residuals**.



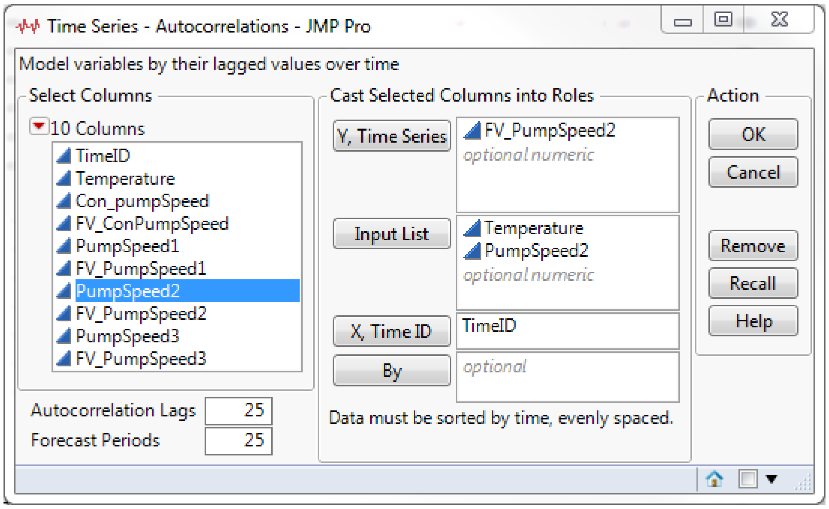
Residuals

Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	-0.0148		0.0636	0.8009	1	-0.0148	
2	0.0168		0.1465	0.9294	2	0.0166	
3	0.0373		0.5546	0.9067	3	0.0378	
4	0.0490		1.2601	0.8681	4	0.0499	
5	-0.0724		2.8080	0.7296	5	-0.0724	
6	0.0471		3.4647	0.7487	6	0.0422	
7	0.0236		3.6304	0.8212	7	0.0238	
8	-0.0538		4.4936	0.8101	8	-0.0523	

The residuals indicate that the AR(1) term is adequate for modeling the autocorrelation.

Now that the relationship among fill volume, temperature, and pump speed has been quantified, by adjusting pump speed appropriately, the variability of fill volume can be reduced.

Open a new instance of the **Time Series Platform**. Assign the variables to their roles as shown below.



Time Series - Autocorrelations - JMP Pro

Model variables by their lagged values over time

Select Columns

10 Columns

- TimeID
- Temperature
- Con_pumpSpeed
- FV_ConPumpSpeed
- PumpSpeed1
- FV_PumpSpeed1
- PumpSpeed2
- FV_PumpSpeed2
- PumpSpeed3
- FV_PumpSpeed3

Cast Selected Columns into Roles

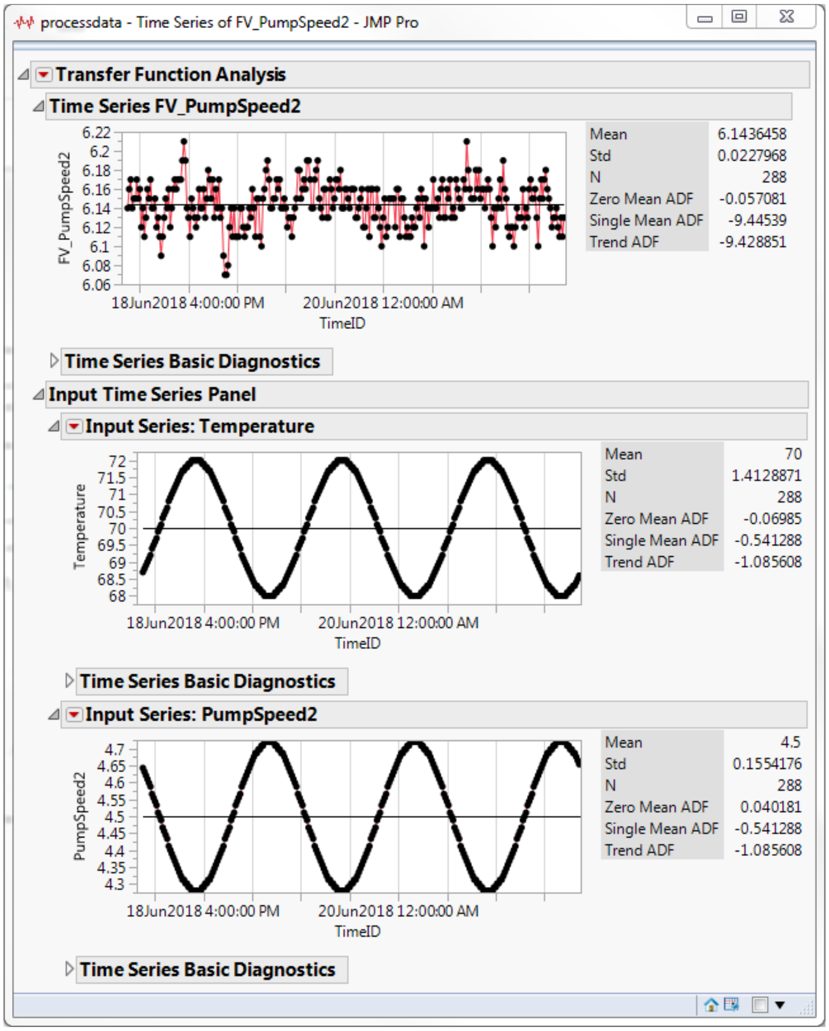
Y, Time Series	FV_PumpSpeed2 *optional numeric*
Input List	Temperature PumpSpeed2 *optional numeric*
X, Time ID	TimeID
By	*optional*

Data must be sorted by time, evenly spaced.

Autocorrelation Lags 25

Forecast Periods 25

Select **OK**. The following graphs appear.



processdata - Time Series of FV_PumpSpeed2 - JMP Pro

Transfer Function Analysis

Time Series FV_PumpSpeed2

Mean	6.1436458
Std	0.0227968
N	288
Zero Mean ADF	-0.057081
Single Mean ADF	-9.44539
Trend ADF	-9.428851

Time Series Basic Diagnostics

Input Time Series Panel

Input Series: Temperature

Mean	70
Std	1.4128871
N	288
Zero Mean ADF	-0.06985
Single Mean ADF	-0.541288
Trend ADF	-1.085608

Time Series Basic Diagnostics

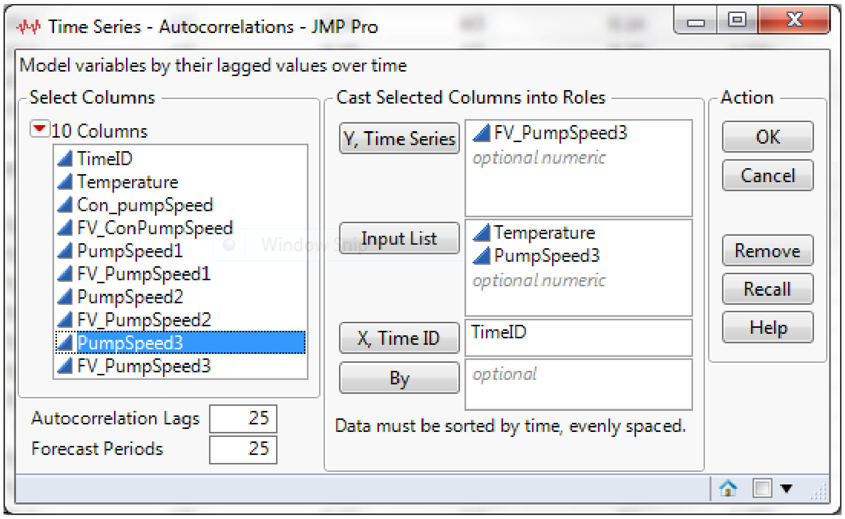
Input Series: PumpSpeed2

Mean	4.5
Std	0.1554176
N	288
Zero Mean ADF	0.040181
Single Mean ADF	-0.541288
Trend ADF	-1.085608

Time Series Basic Diagnostics

The mean of fill volume has not changed but the standard deviation has dropped from 0.404 to 0.0228, a decrease of about 45 percent. Consequently, the target fill volume can be reduced by reducing the average pump speed to 4.1 gallons per minute.

Open another instance of the **Time Series Platform** and assign variables and roles as shown below.


Time Series - Autocorrelations - JMP Pro
Select Columns
Cast Selected Columns into Roles
Action
OK
Cancel
Remove
Recall
Help

Model variables by their lagged values over time

10 Columns

- TimeID
- Temperature
- Con_pumpSpeed
- FV_ConPumpSpeed
- PumpSpeed1
- FV_PumpSpeed1
- PumpSpeed2
- FV_PumpSpeed2
- PumpSpeed3
- FV_PumpSpeed3

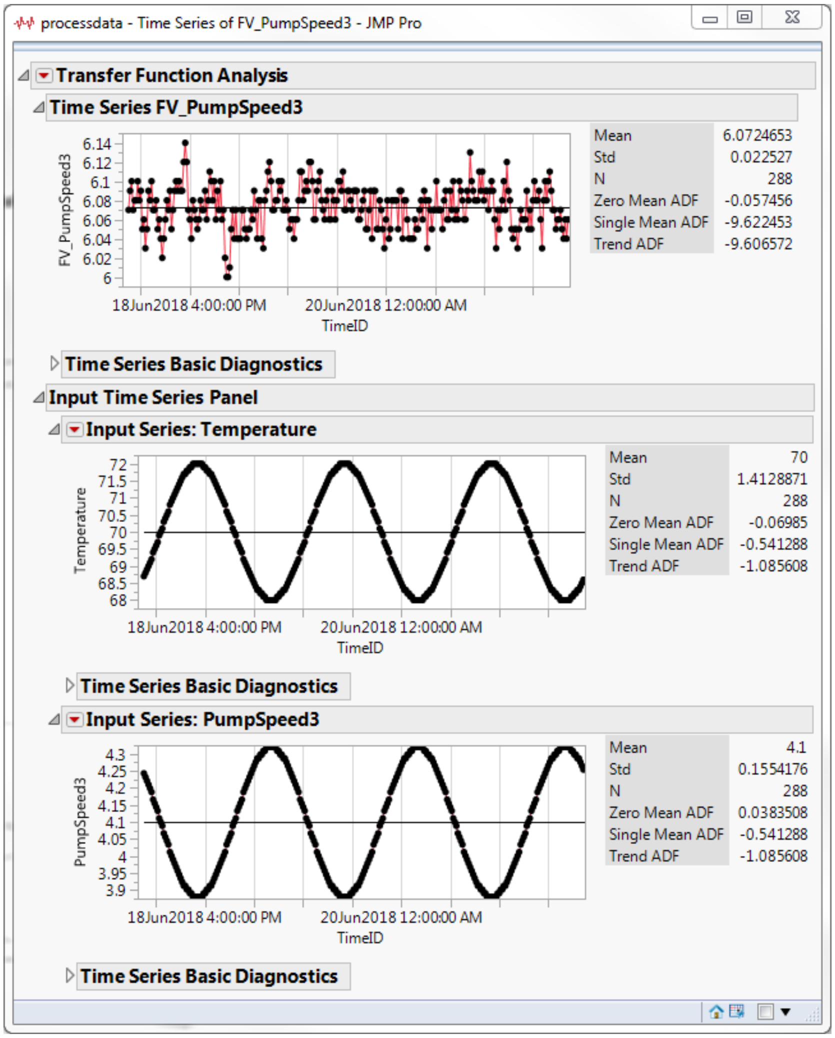
Y, Time Series	FV_PumpSpeed3 *optional numeric*
Input List	Temperature PumpSpeed3 *optional numeric*
X, Time ID	TimeID
By	*optional*

Autocorrelation Lags 25

Forecast Periods 25

Data must be sorted by time, evenly spaced.

Select **OK**.



Mean	6.0724653
Std	0.022527
N	288
Zero Mean ADF	-0.057456
Single Mean ADF	-9.622453
Trend ADF	-9.606572

Mean	70
Std	1.4128871
N	288
Zero Mean ADF	-0.06985
Single Mean ADF	-0.541288
Trend ADF	-1.085608

Mean	4.1
Std	0.1554176
N	288
Zero Mean ADF	0.0383508
Single Mean ADF	-0.541288
Trend ADF	-1.085608

The decrease in average pump speed to 4.1 gallons per minute reduces the average fill volume from 6.144 to 6.072 fluid ounces, a savings of only 0.072 fluid ounces per bottle. However, when the company produces 5 million bottles a year, the total savings is 360,000 fluid ounces and $69,000.

Market Mix Model

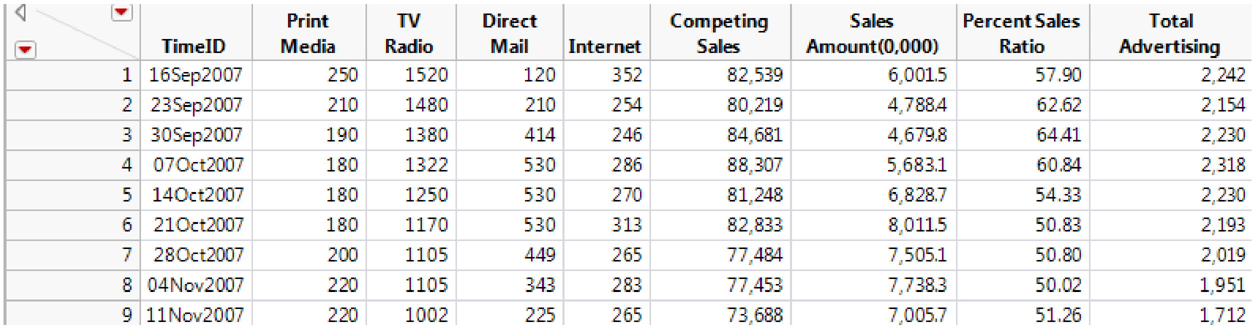
Business Application Description

A market mix model is a multiple regression model with the response sales in dollars and inputs the amount of dollars spent on advertising by different media. Because the data is a time series, usually autocorrelation is present so the assumption of independent errors is violated. It is often also the case that there is a delay in advertising spend and impact on sales. Thus spend in some media is a leading indicator of sales in the future.

A market mix model can be used to quantify the return on investment for each dollar spent by the different media. Once the return on investment is quantified, it can be used to allocate future advertising dollars. The model can also be used to plan future advertising strategies by examining different allocation of a total advertising budget.

Market Mix Model Demonstration.

The market mix data table is shown below. It contains the weekly sales and advertising spend for a nonperishable product. The data is based on real data but has been modified for demonstration purposes

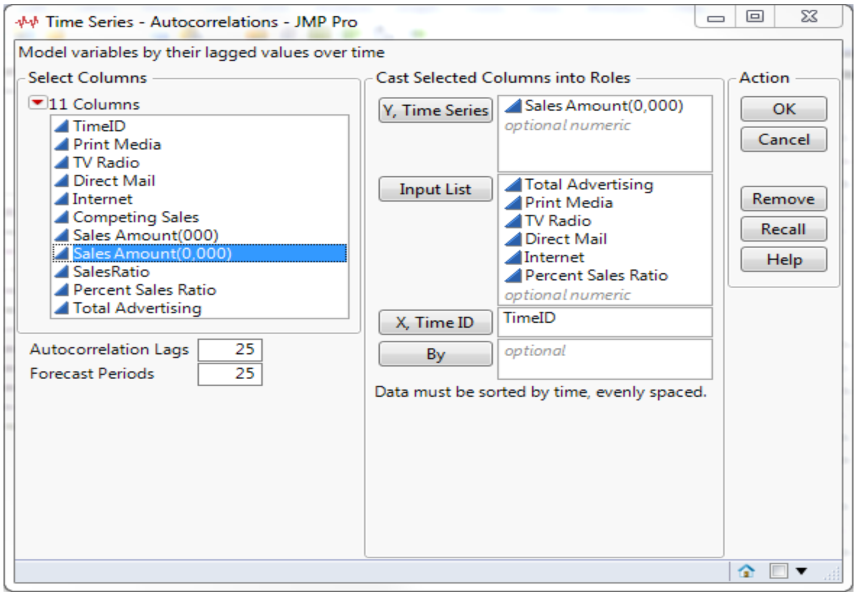


	TimeID	Print Media	TV Radio	Direct Mail	Internet	Competing Sales	Sales Amount(0,000)	Percent Sales Ratio	Total Advertising
1	16Sep2007	250	1520	120	352	82,539	6,001.5	57.90	2,242
2	23Sep2007	210	1480	210	254	80,219	4,788.4	62.62	2,154
3	30Sep2007	190	1380	414	246	84,681	4,679.8	64.41	2,230
4	07Oct2007	180	1322	530	286	88,307	5,683.1	60.84	2,318
5	14Oct2007	180	1250	530	270	81,248	6,828.7	54.33	2,230
6	21Oct2007	180	1170	530	313	82,833	8,011.5	50.83	2,193
7	28Oct2007	200	1105	449	265	77,484	7,505.1	50.80	2,019
8	04Nov2007	220	1105	343	283	77,453	7,738.3	50.02	1,951
9	11Nov2007	220	1002	225	265	73,688	7,005.7	51.26	1,712

The interval for the TimeID is week. The responses is Sales Amount(0,000) in units of $10,000. There are five candidate inputs. The columns Print Media, TV Radio, Direct Mail, and Internet are dollar spend in units of $1,000. The fifth is Percent Sales Ratio that is a function of Competing Sales and Sales Amount. It is used in the model to account for how competition affects sales. Total Advertising is the sum of spend over all media so is not a viable input if the four media are included.

It is very important when interpreting the model coefficients that one keeps in mind the scale of the response and the inputs

Select **Analyze**, then **Specialized Modeling** and then **Time Series** to open the **Time Series Platform** Dialog. Assign the variables to their roles as shown below.



Time Series - Autocorrelations - JMP Pro

Model variables by their lagged values over time

Select Columns

11 Columns
- TimeID
- Print Media
- TV Radio
- Direct Mail
- Internet
- Competing Sales
- Sales Amount(000)
- Sales Amount(0,000)
- SalesRatio
- Percent Sales Ratio
- Total Advertising

Autocorrelation Lags 25

Forecast Periods 25

Cast Selected Columns into Roles

Y, Time Series: Sales Amount(0,000)

Input List: Total Advertising, Print Media, TV Radio, Direct Mail, Internet, Percent Sales Ratio

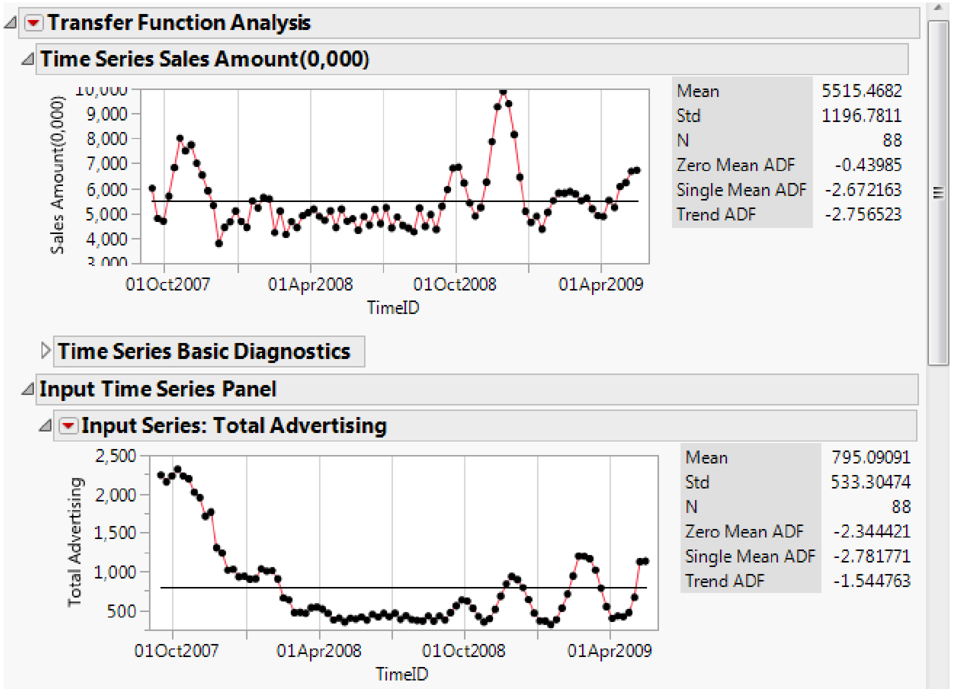
X, Time ID: TimeID

By: optional

Data must be sorted by time, evenly spaced.

Select **OK**.

Plots of the response and candidate inputs by week are shown below.



Transfer Function Analysis

Time Series Sales Amount(0,000)

Mean	5515.4682
Std	1196.7811
N	88
Zero Mean ADF	-0.43985
Single Mean ADF	-2.672163
Trend ADF	-2.756523

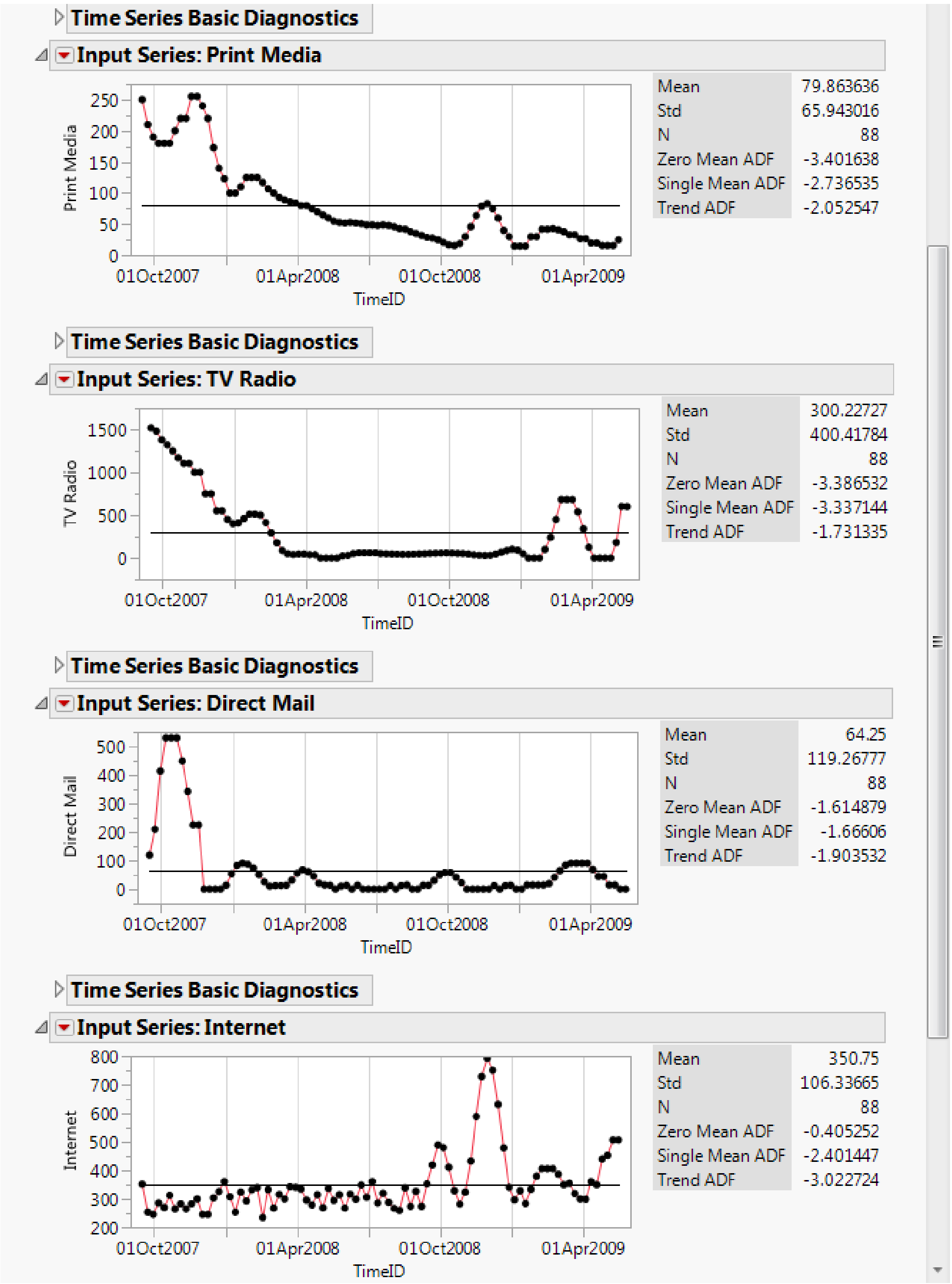
Time Series Basic Diagnostics

Input Time Series Panel

Input Series: Total Advertising

Mean	795.09091
Std	533.30474
N	88
Zero Mean ADF	-2.344421
Single Mean ADF	-2.781771
Trend ADF	-1.544763

Notice that Sales Amount is rather flat except for a few spikes in late 2007 and late in 2008. Total advertising spend has declined dramatically since late 2007. There appears to be several campaigns in late 2008 and early 2009. Examine the advertising spend by different media in the plots below.



▷ Time Series Basic Diagnostics

◿ Input Series: Print Media

Mean	79.863636
Std	65.943016
N	88
Zero Mean ADF	-3.401638
Single Mean ADF	-2.736535
Trend ADF	-2.052547

▷ Time Series Basic Diagnostics

◿ Input Series: TV Radio

Mean	300.22727
Std	400.41784
N	88
Zero Mean ADF	-3.386532
Single Mean ADF	-3.337144
Trend ADF	-1.731335

▷ Time Series Basic Diagnostics

◿ Input Series: Direct Mail

Mean	64.25
Std	119.26777
N	88
Zero Mean ADF	-1.614879
Single Mean ADF	-1.66606
Trend ADF	-1.903532

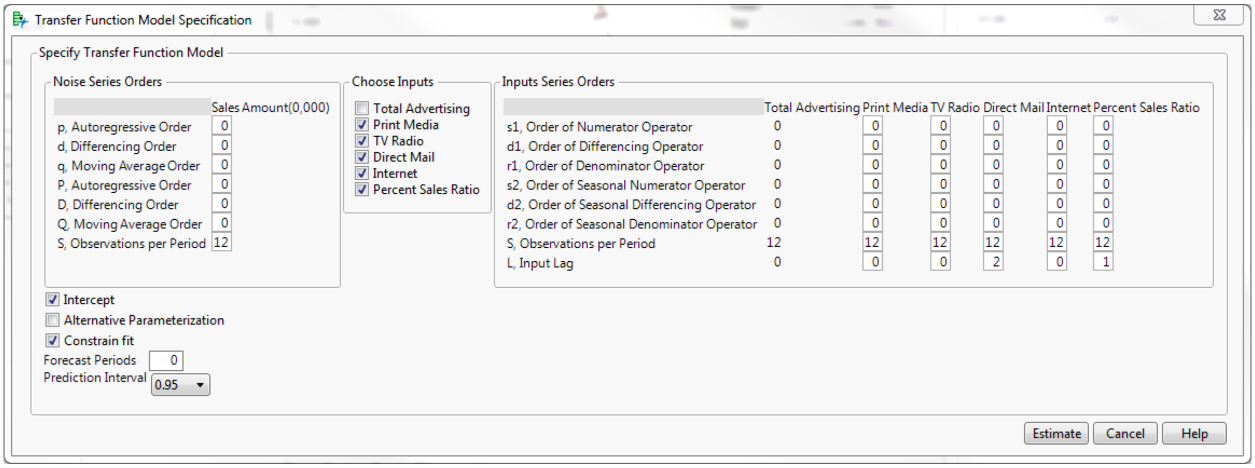
▷ Time Series Basic Diagnostics

◿ Input Series: Internet

Mean	350.75
Std	106.33665
N	88
Zero Mean ADF	-0.405252
Single Mean ADF	-2.401447
Trend ADF	-3.022724

It appears that the internet became a more important part of their advertising strategy starting in late 2008.

To fit the first model, from the red triangle beside **Transfer Function Analysis** drop down menu, select **Transfer Function** to bring up the **Specify Transfer Function Model** window. Configure the options as shown below.



Transfer Function Model Specification

Specify Transfer Function Model

Noise Series Orders

	Sales Amount(0,000)
p, Autoregressive Order	0
d, Differencing Order	0
q, Moving Average Order	0
P, Autoregressive Order	0
D, Differencing Order	0
Q, Moving Average Order	0
S, Observations per Period	12

Choose Inputs

- ☐ Total Advertising
- ☑ Print Media
- ☑ TV Radio
- ☑ Direct Mail
- ☑ Internet
- ☑ Percent Sales Ratio

Inputs Series Orders

	Total Advertising	Print Media	TV Radio	Direct Mail	Internet	Percent Sales Ratio
s1, Order of Numerator Operator	0	0	0	0	0	0
d1, Order of Differencing Operator	0	0	0	0	0	0
r1, Order of Denominator Operator	0	0	0	0	0	0
s2, Order of Seasonal Numerator Operator	0	0	0	0	0	0
d2, Order of Seasonal Differencing Operator	0	0	0	0	0	0
r2, Order of Seasonal Denominator Operator	0	0	0	0	0	0
S, Observations per Period	12	12	12	12	12	12
L, Input Lag	0	0	0	2	0	1

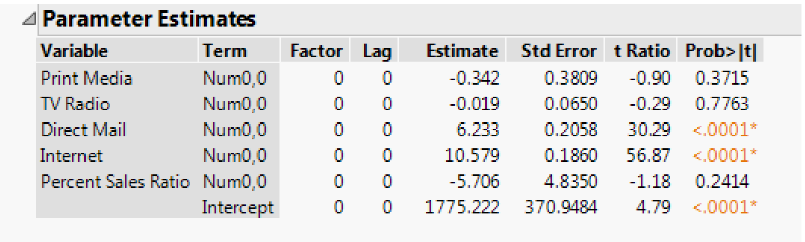
☑ Intercept
☐ Alternative Parameterization
☑ Constrain fit
Forecast Periods 0
Prediction Interval 0.95

Be sure to uncheck Total Advertising. A lag of two (2) was chosen for Direct Mail based on business knowledge. A lag of one (1) was chosen based on business knowledge and some trial and error.

There are analytical methods to help chose a transfer function form and the lag for inputs that involve a process called Prewhitening. (Box and Jenkins, Chapter 11.). Usually, some trial and error is still needed to arrive at a “best” model.

Select **Estimate**.

The **Parameter Estimates** Table is below.



Parameter Estimates

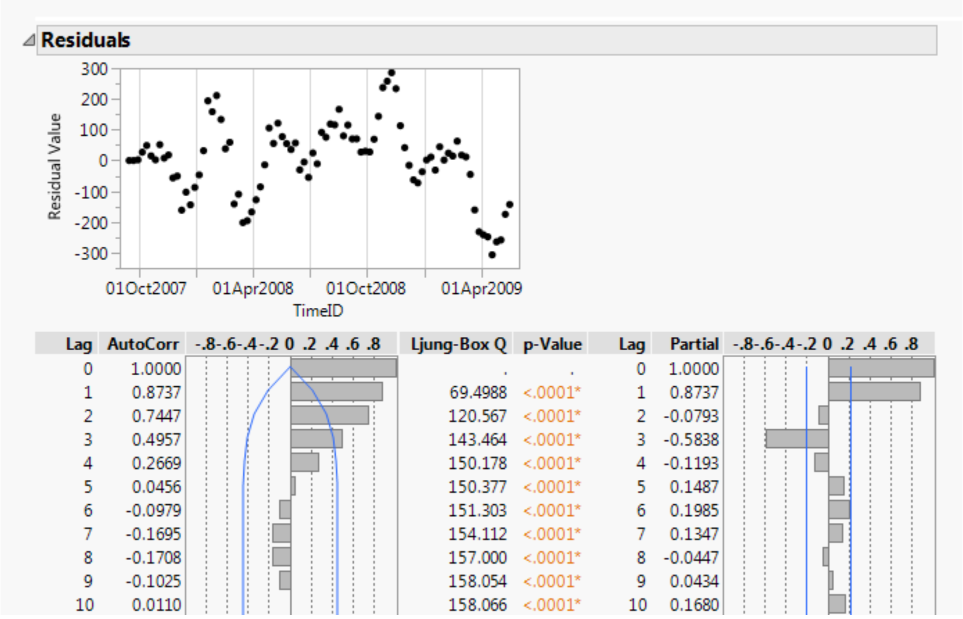
Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Print Media	Num0,0	0	0	-0.342	0.3809	-0.90	0.3715
TV Radio	Num0,0	0	0	-0.019	0.0650	-0.29	0.7763
Direct Mail	Num0,0	0	0	6.233	0.2058	30.29	<.0001*
Internet	Num0,0	0	0	10.579	0.1860	56.87	<.0001*
Percent Sales Ratio	Num0,0	0	0	-5.706	4.8350	-1.18	0.2414
	Intercept	0	0	1775.222	370.9484	4.79	<.0001*

Some of the parameter estimates are not significant. However, because the possible autocorrelation has not been accounted for, it is premature to conclude they are not important. Now examine the model equation.

$$\text{Sales Amount(0,000)}_t = \left(\left(\left(\left(\left(1775.222 - 0.3423 \bullet \text{Print Media}_t\right) - 0.0185 \bullet \text{TV Radio}_t\right) + 6.2327 \bullet \text{Direct Mail}_{t-2}\right) + 10.579 \bullet \text{Internet}_t\right) - 5.7062 \bullet \text{Percent Sales Ratio}_{t-1}\right) + e_t$$

The subscript for Direct Mail of (t-2) reflects the lag of two (2) specified. The subscript of (t-1) for Percent Sales Ratio reflects the lag or one (1) specified.

Examine the **Residuals**.

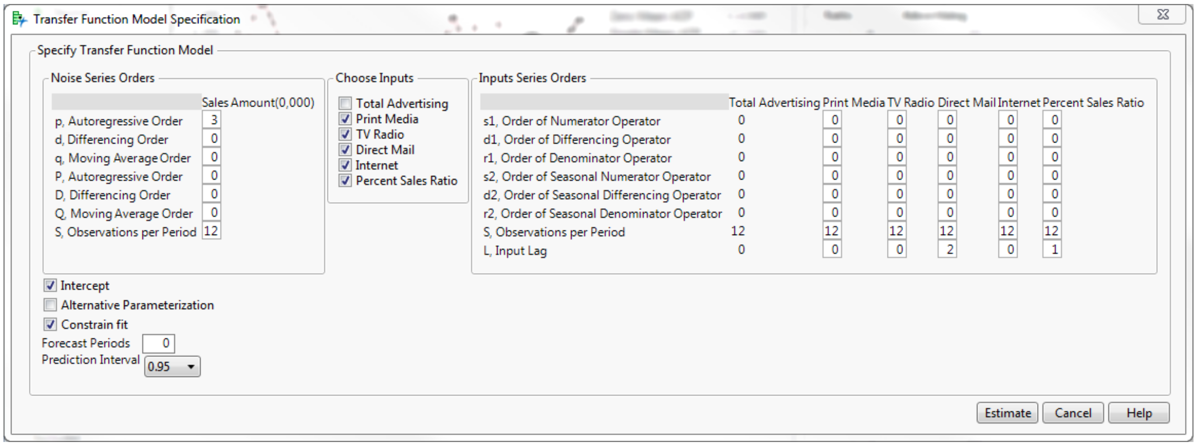


Residuals

Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.8737		69.4988	<.0001*	1	0.8737	
2	0.7447		120.567	<.0001*	2	-0.0793	
3	0.4957		143.464	<.0001*	3	-0.5838	
4	0.2669		150.178	<.0001*	4	-0.1193	
5	0.0456		150.377	<.0001*	5	0.1487	
6	-0.0979		151.303	<.0001*	6	0.1985	
7	-0.1695		154.112	<.0001*	7	0.1347	
8	-0.1708		157.000	<.0001*	8	-0.0447	
9	-0.1025		158.054	<.0001*	9	0.0434	
10	0.0110		158.066	<.0001*	10	0.1680	

The pattern of the residual plot and in the autocorrelation and partial autocorrelation plots indicate that there is strong autocorrelation in the residuals. The large spike at lag three (3) in the partial autocorrelation plot suggest that an autoregressive term of order three (3) (AR(3)) is needed.

Bring up a second instance of the **Transfer Function Model Specification** window and specify the options as shown below.



Transfer Function Model Specification

Specify Transfer Function Model

Noise Series Orders

	Sales Amount(0,000)
p, Autoregressive Order	3
d, Differencing Order	0
q, Moving Average Order	0
P, Autoregressive Order	0
D, Differencing Order	0
Q, Moving Average Order	0
S, Observations per Period	12

Choose Inputs

□ Total Advertising
☑ Print Media
☑ TV Radio
☑ Direct Mail
☑ Internet
☑ Percent Sales Ratio

Inputs Series Orders

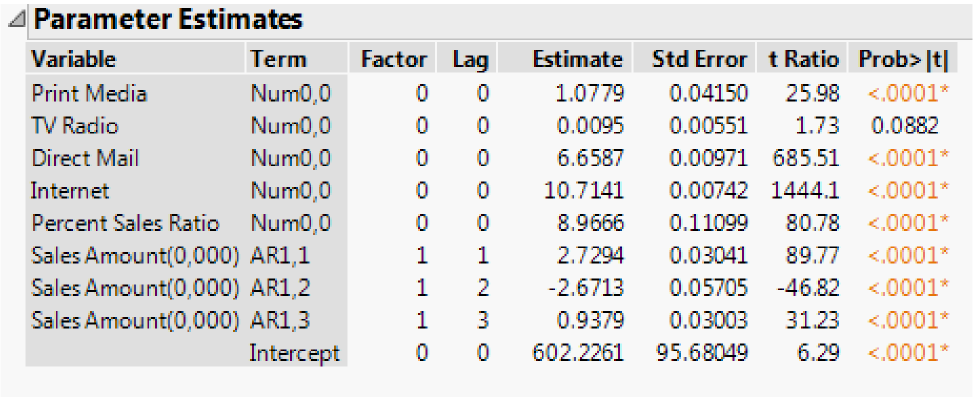
	Total Advertising	Print Media	TV Radio	Direct Mail	Internet	Percent Sales Ratio
s1, Order of Numerator Operator	0	0	0	0	0	0
d1, Order of Differencing Operator	0	0	0	0	0	0
r1, Order of Denominator Operator	0	0	0	0	0	0
s2, Order of Seasonal Numerator Operator	0	0	0	0	0	0
d2, Order of Seasonal Differencing Operator	0	0	0	0	0	0
r2, Order of Seasonal Denominator Operator	0	0	0	0	0	0
S, Observations per Period	12	12	12	12	12	12
L, Input Lag	0	0	0	2	0	1

☑ Intercept
□ Alternative Parameterization
☑ Constrain fit
Forecast Periods 0
Prediction Interval 0.95

The only change from the previous model is the inclusion of the AR(3) specification.

Select **Estimate**.

The parameter estimates for the second model are shown below.

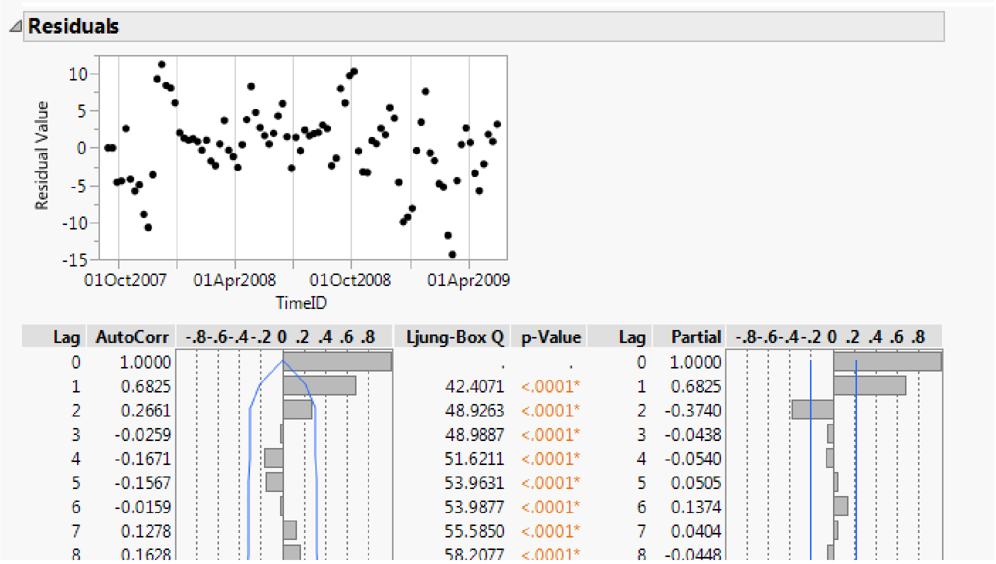


Parameter Estimates

Variable	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Print Media	Num0,0	0	0	1.0779	0.04150	25.98	<.0001*
TV Radio	Num0,0	0	0	0.0095	0.00551	1.73	0.0882
Direct Mail	Num0,0	0	0	6.6587	0.00971	685.51	<.0001*
Internet	Num0,0	0	0	10.7141	0.00742	1444.1	<.0001*
Percent Sales Ratio	Num0,0	0	0	8.9666	0.11099	80.78	<.0001*
Sales Amount(0,000)	AR1,1	1	1	2.7294	0.03041	89.77	<.0001*
Sales Amount(0,000)	AR1,2	1	2	-2.6713	0.05705	-46.82	<.0001*
Sales Amount(0,000)	AR1,3	1	3	0.9379	0.03003	31.23	<.0001*
	Intercept	0	0	602.2261	95.68049	6.29	<.0001*

Print Media and Percent Sales Ratio parameters are now significant. TV Radio parameter is still not significant but its p-value is much smaller than before.

Expand the **Residuals** window to ascertain if autocorrelation remains in the residuals.



Residuals

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.6825	42.4071	<.0001*	1	0.6825
2	0.2661	48.9263	<.0001*	2	-0.3740
3	-0.0259	48.9887	<.0001*	3	-0.0438
4	-0.1671	51.6211	<.0001*	4	-0.0540
5	-0.1567	53.9631	<.0001*	5	0.0505
6	-0.0159	53.9877	<.0001*	6	0.1374
7	0.1278	55.5850	<.0001*	7	0.0404
8	0.1628	58.2077	<.0001*	8	-0.0448

The plots indicate that there is still autocorrelation in the residuals. The plots suggest that higher order autoregressive terms maybe needed.

After trying AR(4) and AR(5) terms, the algorithm that estimates the parameters failed to converge. Convergence problems can happen when trying to fit time series models with a large number of parameters. A given data set only will support models up to a specific level of complexity.

Though this model may not be perfect, it is useful. The model equation is shown below.

$$\text{Sales Amount(0,000)}_{t} = \left(\left(\left(\left(602.2261 + 1.0779 \cdot \text{Print Media}_{t}\right) + 0.0095 \cdot \text{TV Radio}_{t}\right) + 6.6587 \cdot \text{Direct Mail}_{t-2}\right) + 10.7141 \cdot \text{Internet}_{t}\right) + 8.9666 \cdot \text{Percent Sales Ratio}_{t-1}$$

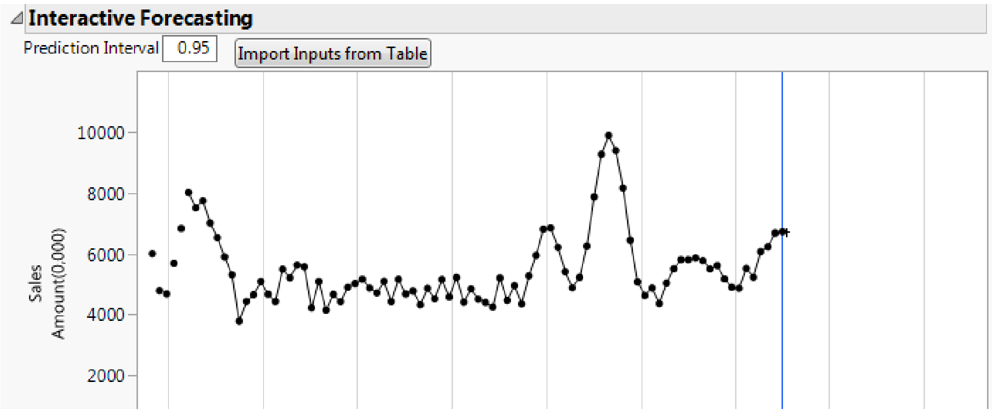
$$+\left(\frac{1}{\left(\left(\left(1 - 2.7294 \cdot B\right) + 2.6713 \cdot B^{2}\right) - 0.9379 \cdot B^{3}\right)}\right) \cdot e_{t}$$

The regression coefficients in the model and be interpreted in the following way.

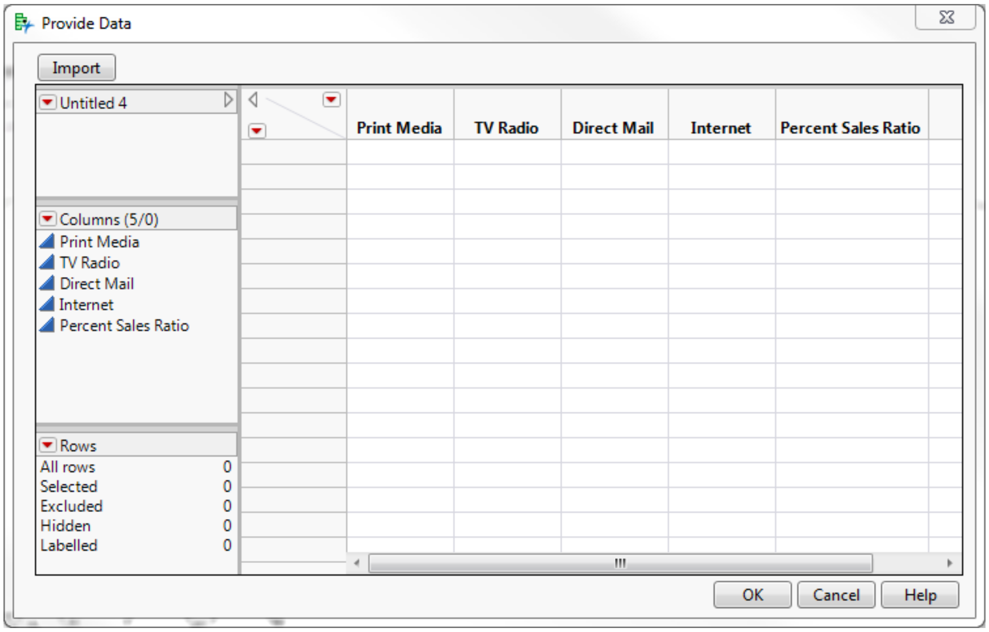
The Print Media coefficient is 1.078. Keeping in mind the difference in scale, for every dollar spent on Print Media advertising returns $10.78. For Direct Mail, for every dollar spent two weeks before increase sales by $66.59 in the current week. The other advertising coefficients can be interpreted similarly.

Interpreting the actual values of the autoregressive terms is difficult. The most important fact is that sales in the current week not only depend on the recent advertising spend and also on actual sales in the prior three weeks.

The **Interactive Forecasting** feature can be used to illustrate the usefulness of the model.

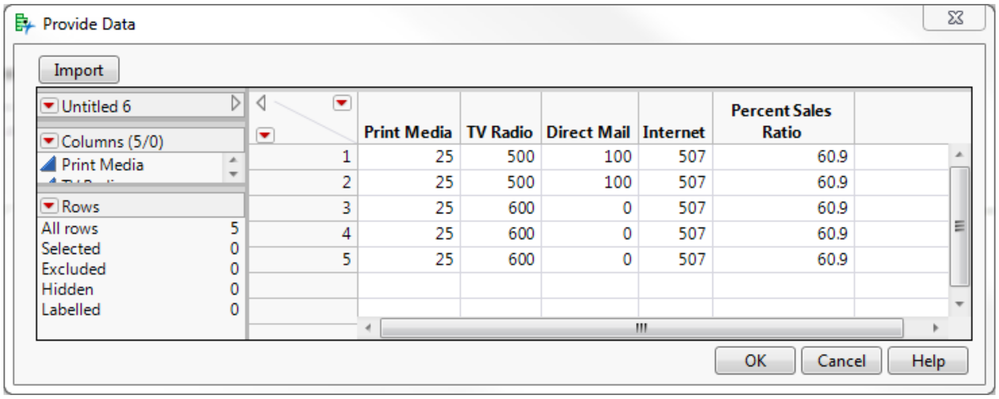


Select **Import Inputs from Table**.



A data table appears with columns named for the inputs in the model.

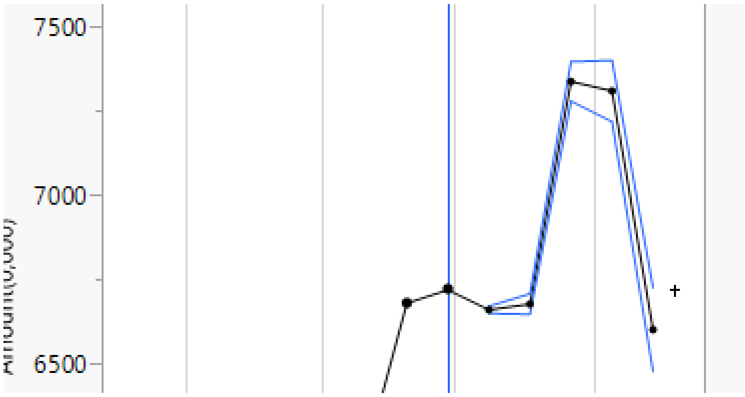
Future sales as a function of the inputs can be produced by allocating the advertising budget to the different media. For example, reallocate $100,000 from TV radio to Direct Mail as shown below. Recall that the units of advertising spend are in thousands.



Rows	
All rows	5
Selected	0
Excluded	0
Hidden	0
Labelled	0

	Print Media	TV Radio	Direct Mail	Internet	Percent Sales Ratio
1	25	500	100	507	60.9
2	25	500	100	507	60.9
3	25	600	0	507	60.9
4	25	600	0	507	60.9
5	25	600	0	507	60.9

The forecast based on these values is shown below.



The three-step ahead forecast reflects the two-week lag in the effect of the increase in Direct Mail spend. The blue lines are the 95% confidence bounds

Conclusions.

The **JMP® Time Series Platform** was used to fit dynamic regression models that are useful in industry and business. A dynamic regression model was used to reduce the variability of a manufacturing process that resulted in substantial cost savings. A dynamic regression models was used to develop a business market mix model that allows one to examine the impact on sales of different allocations of marketing dollars to different media.

Acknowledgements.

I thank my wife, Rosemary Batten Lucas for he helpful comments in reviewing my manuscript. I also thank SAS Education for allowing me to use the market mix data. I also thank my anonymous friend for taking the time to describe the manufacturing process and review my simulation for realism. I also thank the Discovery Conference Steering Committee for accepting my talk.

References

Box, G. E. P. and Jenkins, G. M. 1976. *Time Series Analysis forecasting and control*. Holden-Day San Francisco.

SAS Institute Inc. 2018. *JMP® 14 Predictive and Specialized Modeling*. Cary, NC: SAS Institute Inc.

Contact Information:

Email: robertlucas1972@gmail.com