

Time Series Analysis

This (not surprisingly) concerns the analysis of data collected over time ... weekly values, monthly values, quarterly values, yearly values, etc. Usually the intent is to discern whether there is some pattern in the values collected to date, with the intention of short term forecasting (to use as the basis of business decisions). We will write

$$y_t = \text{response of interest at time } t$$

(we usually think of these as equally spaced in clock time). Standard analyses of business time series involve:

- 1) smoothing/trend assessment

1

- 2) assessment of/accounting for seasonality

- 3) assessment of/exploiting "serial correlation"

These are usually/most effectively done on a scale where the "local" variation in y_t is approximately constant.

Smoothing Time Series

There are various fairly simple smoothing/averaging methods. Two are "ordinary moving averages" and "exponentially weighted moving averages."

2

Ordinary Moving Averages For a "span" of k periods,

$$\begin{aligned}\tilde{y}_t &= \text{moving average through time } t \\ &= \frac{y_t + y_{t-1} + y_{t-2} + \cdots + y_{t-k+1}}{k}\end{aligned}$$

Where seasonal effects are expected, it is standard to use

$$k = \text{number of periods per cycle}$$

Exponentially Weighted Moving Averages These weight observations less heavily as one moves back in time from the current period. They are typically computed "recursively" as

$$\begin{aligned}\tilde{y}_t &= \text{exponentially weighted moving average at time } t \\ &= w y_t + (1 - w) \tilde{y}_{t-1}\end{aligned}$$

3

(\tilde{y}_{t-1} is the EWMA from the previous period and the current EWMA is a compromise between the previous EWMA and the current observation.)

One must start this recursion somewhere and it's common to take $\tilde{y}_1 = y_1$.

Notice that $w = 1$ does no smoothing, while $w = 0$ smooths so much that the EWMA never changes (i.e. all the values are equal to the first).

Exercise/Example Table 13.1 (page 13-5) of the text gives quarterly retail sales for JC Penney, 1996-2001 (in millions of dollars). "By hand" 1) using $k = 4$ find the ordinary moving average for period 8, then 2) using (e.g.) $w = .3$, find the exponentially weighted moving average value for that period. (Finish the table.)

4

t	y_t	Span $k = 4$ MA	$w = .3$ EWMA
1	4452		4452
2	4507		$4469 = .3(4507) + .7(4452)$
3	5537		$4789 = .3(5537) + .7(4469)$
4	8157	$5663 = \frac{1}{4} \left(\begin{array}{l} 4452 + 4507 \\ + 5537 + 8157 \end{array} \right)$	$5799 = .3(8157) + .7(4789)$
5	6481	6170.5	6004
6	6420	6649	6129
7	7208	7066.5	6453
8	9509		

A plot of both the original time series and the $k = 4$ MA values for the JC Penney data is in Figure 13.13, page 13-28 of the text. Here is a JMP "Overlay Plot" version of this picture and an indication of how you can get JMP to make the MA's.

5

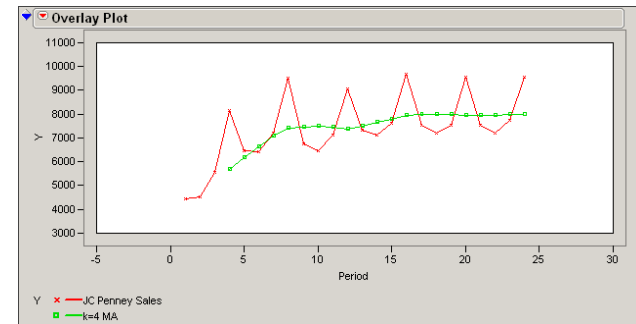


Figure 1: JC Penney Sales and $k = 4$ MA Series

6

$0.25 * (JC\ Penney\ Sales + Lag(JC\ Penney\ Sales, 1) + Lag(JC\ Penney\ Sales, 2) + Lag(JC\ Penney\ Sales, 3))$

Figure 2: JMP "Column Formula" for JC Penney MA's

Computation of EWMA's in JMP doesn't appear to be simple. Here is a plot of 3 EWMA series for the JC Penney sales data that shows the effect of changing w on how much smoothing is done. The most jagged plot is the (red) raw data plot ($w = 1.0$). The (green) $w = .5$ EWMA plot is smoother. The (blue) $w = .3$ EWMA plot is smoother still. The (black) $w = .1$ plot is smoothest.

7

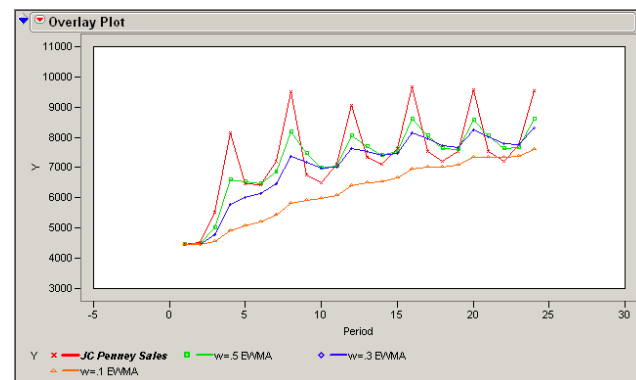


Figure 3: EWMA's for JC Penney Sales Data

8

There are other more sophisticated smoothing methods available in statistical software. JMP provides "splines."

JMP Cubic Spline Smoothers These are available using the "Fit Y by X" procedure in JMP. They have a "stiffness knob" that lets one adjust how much "wiggling" the smoothed curve can do. Here are several splines fit to the JC Penney sales data. The "stiffness knob" is the parameter " λ ."

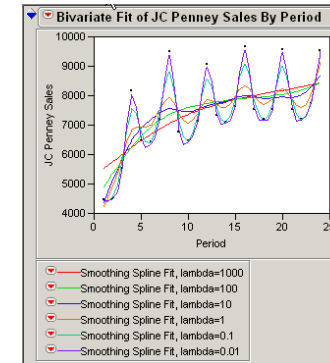


Figure 4: Splines Fit to the JC Penney Data

JMP will store the smoothed values obtained from these spline smoothers (just as it will store predicted values from regressions) in the original data table, if one clicks on the appropriate red triangle and chooses that option.

Typically one wants to "smooth" a time series in order to make forecasts/projections into the future. The MA, EWMA, and spline smoothers don't really provide forecasts beyond projecting a current value \tilde{y}_t to the next period, period $t+1$. A possibility for smoothing that provides forecasts other than a current smoothed value is to fit a simple curve to the series using regression, where the "x" variable is "t" (that is, the data vectors are $(1, y_1), (2, y_2), \dots$). It is particularly easy to fit "low order" polynomials (lines, parabolas, etc.) to such data using JMP. These provide extrapolations beyond the end of the data set.

JMP Fitting of (Low Order) Polynomial Trends to Time Series These are again conveniently available using the "Fit Y by X" procedure in JMP.

(Conceptually, one could also use the multiple regression procedure "Fit Model" after adding columns to the data table for powers of t . But we'll use the more elegant "Fit Y by X" method.)

Below is a JMP graphic for linear and quadratic (1st and second order polynomial) fits to the JC Penney time series. NOTICE that the extrapolations to a 25th period from these two polynomials will be quite different! The two fitted equations are

$$\hat{y}_t = 5903.2174 + 118.75261t$$

and

$$\hat{y}_t = 6354.9514 + 118.75261t - 9.4274932(t - 12.5)^2$$

and by clicking on the appropriate red triangles (next to the "Linear Fit" or "Polynomial Fit") one can save the predicted values into the data table. (If one

uses the "Fit Model" procedure, one can save the *formula* for fitted equation and get JMP to automatically compute forecasts into the future by adding rows to the data table with hypothetical *t*'s in them.)

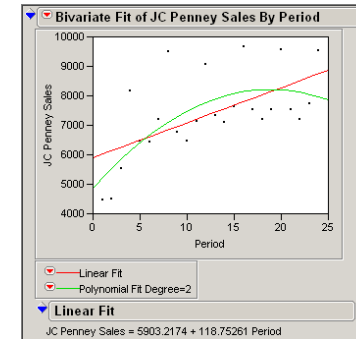


Figure 5: JMP Fit of Linear and Parabolic Trends to the JC Penney Data

As one moves from a line, to a parabola, to a cubic, etc., fitted polynomials will be allowed to be more and more wiggly, doing a better and better job of hitting the plotted points, but becoming less and less believable in terms of forecasts. The happiest circumstance is that where a simple straight line/linear trend seems to provide an adequate summary of the main movement of the time series.

Exercise/Example Compute "by hand" the linear and quadratic forecasts of y_{25} (the sales for the period immediately after the end of the data set) for the JC Penney sales based on the JMP fitted equations.

These (quite different) forecasts are

$$\begin{aligned}\hat{y}_{25} &= 5903.2174 + 118.75261(25) \\ &= 8872\end{aligned}$$

and

$$\begin{aligned}\hat{y}_{25} &= 6354.9514 + 118.75261(25) - 9.4274932(25 - 12.5)^2 \\ &= 7851\end{aligned}$$

Accounting for/Adjusting for Seasonality

The linear trend fit to the JC Penney data misses the seasonality in the data. Mostly, the straight line in Figure 5 "over-predicts" in the first 3 quarters of each year and "under-predicts" in the 4th quarter of each year. ($t = 1, 5, 9, 13, 17, 21$ are "first quarter" periods, $t = 2, 6, 10, 14, 18, 22$ are "second quarter" periods, etc.) It is well known that retail sales are typically best in the 4th quarter, where the Christmas season spurs consumer buying.

It makes sense in the analysis of business and economic time series to try to adjust smoothed values (and forecasts) in light of seasonal effects. Here we'll consider several ways of doing this.

Simple Arithmetic and "Additive" Adjustment for Seasonal Effects One simple way of trying to account for seasonality is to look at all periods of a given type (e.g. 1st quarter periods where data are quarterly, or all June figures where data are monthly) and compute an average deviation of the original time series from the smoothed or fitted values in those periods. That average can then be *added* to smoothed values or forecasts from a smooth curve in order to account for seasonality.

Simple Arithmetic and "Multiplicative" Adjustment for Seasonal Effects A second simple way of trying to account for seasonality is to look at all periods of a given type (e.g. 1st quarter periods where data are quarterly, or

all June figures where data are monthly) and compute an average ratio of the actual values to the smoothed or fitted values in those periods. That average can then be used as a *multiplier* for smoothed values or forecasts from a smooth curve in order to account for seasonality.

Example The table below gives simple computation of "additive" and "multiplicative" seasonality factors for the 1st quarter JC Penney sales, based on the linear trend fit to the data and pictured in Figure 5.

Period, t	y_t	\hat{y}_t	$y_t - \hat{y}_t$	$\frac{y_t}{\hat{y}_t}$
1	4452	6022	-1570	.7393
5	6481	6497	-16	.9975
9	6755	6972	-217	.9685
13	7339	7447	-108	.9855
17	7528	7922	-394	.9503
21	7522	8397	-875	.8958
			-3180	5.5369

Then note that the average $y_t - \hat{y}_t$ is

$$\frac{-3180}{6} = -530$$

and the average y_t/\hat{y}_t is

$$\frac{5.5369}{6} = .9228$$

So fitted values or forecasts from the line fit to the JC Penney data could be adjusted by either *addition of -530 or multiplication by .9228*. For example, the forecast for period 25 (the first period after the end of the data in hand and a first quarter) from the linear fit in Figure 5 alone is 8872. This could be adjusted for the seasonality as either

$$\hat{y}_{25} = 8872 + (-530) = 8342$$

(making use of an "additive" seasonality adjustment) or as

$$\hat{y}_{25} = 8872(.9228) = 8187$$

(making use of a "multiplicative" seasonality adjustment).

Exercise The table below gives the 4th quarter values and fitted values from the line fit to the JC Penney data. Finish the calculations, get additive and multiplicative seasonality factors, and use them to make 4th quarter forecasts

for the year following the end of the data (this is period $t = 28$ and the linear fit alone projects sales of $\hat{y}_{28} = 5903.2174 + 118.75261(28) = 9228$).

Period, t	y_t	\hat{y}_t	$y_t - \hat{y}_t$	$\frac{y_t}{\hat{y}_t}$
4	8157	6378	1779	1.2789
8	9509	6853	2656	1.3876
12	9072	7328	1744	1.2380
16	9661	7803	1858	1.2381
20	9573	8278	1295	1.1564
24	9542	8753		

Making an additive adjustment

$$\hat{y}_{28} = 9228 + (\quad) =$$

and making a multiplicative adjustment

$$\hat{y}_{28} = 9229 \times (\quad) =$$

The U.S. government reports values of all kinds of economic time series. In many cases, both "raw" and "seasonally adjusted" versions of these are announced. That is, not only does the government announce a value of "housing starts," but it also announces a value of "seasonally adjusted housing starts." If SF is a multiplicative seasonality factor for the particular month under discussion, this means that both

housing starts

and

$$\text{seasonally adjusted housing starts} = \frac{\text{housing starts}}{SF}$$

are reported.

Using Dummy Variables in MRL to Account for Seasonality A more sophisticated and convenient means of creating (additive) seasonality adjustments is to employ dummy variables in a multiple linear regression. That is, if there are k seasons, one can think of making up $k - 1$ dummy variables x_1, x_2, \dots, x_{k-1} where for period t

$$x_{j,t} = \begin{cases} 1 & \text{if period } t \text{ is from season } j \\ 0 & \text{otherwise} \end{cases}$$

and then using these in a Multiple Linear Regression, fitting (for example)

$$y_t \approx b_0 + b_1 t + a_1 x_{1,t} + a_2 x_{2,t} + \dots + a_{k-1} x_{k-1,t}$$

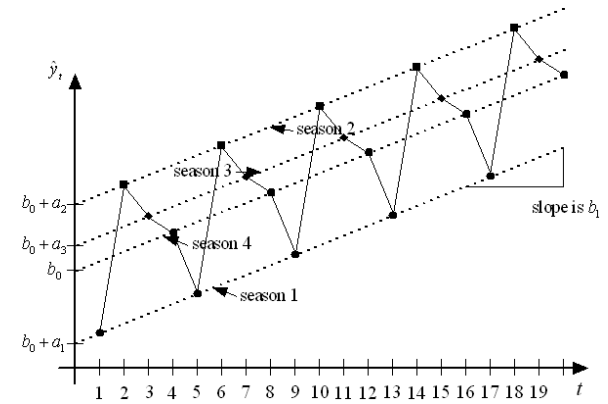
The next figure shows the set-up of a JMP data table for the JC Penney data to make use of this idea.

Figure 6: JMP Data Table Prepared for Using MLR to Account for Seasonality

Period	JC Penney Sales	1st Quarter	2nd Quarter	3rd Quarter
1	4452	1	0	0
2	4507	0	1	0
3	5537	0	0	1
4	8157	0	0	0
5	6481	1	0	0
6	6420	0	1	0
7	7208	0	0	1
8	9509	0	0	0
9	6755	1	0	0
10	6483	0	1	0
11	7129	0	0	1
12	9072	0	0	0
13	7339	1	0	0
14	7104	0	1	0
15	7639	0	0	1
16	9661	0	0	0
17	7528	1	0	0
18	7207	0	1	0
19	7538	0	0	1
20	9573	0	0	0
21	7522	1	0	0
22	7211	0	1	0
23	7729	0	0	1
24	9542	0	0	0

What this method does is let the "intercept" of a linear trend in y_t change with period. A "cartoon" showing how this works for the case where there are $k = 4$ seasons is below.

Figure 7: Cartoon for Dummy Variables and Seasonality ($k = 4$ Seasons)



To fit such a "linear trend plus season dummies" model to time series data, one can employ a multiple linear regression program. JMP's "Fit Model" routine incorporates such a program. The JMP "Fit Model" dialogue box and resulting report for the JC Penney data follow.

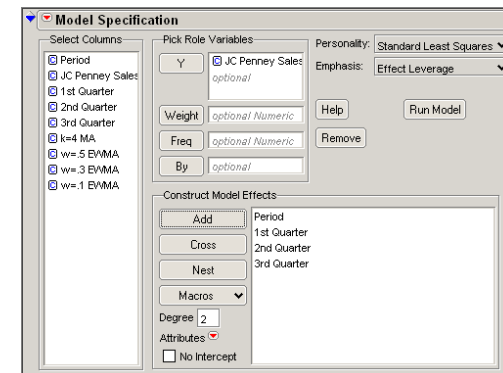


Figure 8: JMP "Fit Model" Dialogue Box for Using Dummies to Account for Seasonality

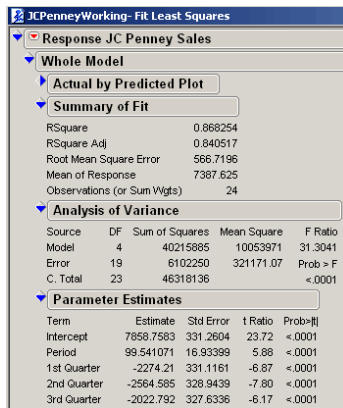


Figure 9: JMP Report for Fitting Linear Trend Plus Seasonal Dummies to the JC Penney Data

The report shows that fitted values for 4th quarter periods t are

$$\hat{y}_t = 7858.8 + 99.541t$$

and, for example, fitted values for 1st quarter periods are

$$\hat{y}_t = (7858.8 + (-2274.2)) + 99.541t$$

So, for example, 25th period (the first quarter immediately after the end of the data set) sales would be forecast as

$$\hat{y}_{25} = 7858.8 - 2274.2 + 99.541(25) = 8073$$

and 28th period sales (4th quarter sales for the year after the data ends) would be forecast as

$$\hat{y}_{28} = 7858.8 + 99.541(28) = 10,646$$

Using Serial Correlation (in Residuals) To Improve Predictions

Sometimes "trend plus seasonal effect" is all the information carried by a time series. But there are also many cases where yet more information can be extracted from the time series to improve on "trend plus seasonal effect" forecasts. This involves using residuals

$$e_t = y_t - \hat{y}_t$$

(for \hat{y}_t the "fitted trend plus seasonal effect" values for the data in hand).

If residuals look like random draws from a fixed universe, then there is nothing left in them to exploit. But sometimes they exhibit "serial correlation" that

allows us to effectively predict a given residual from previous ones. That is, sometimes the pairs

$$(e_{t-1}, e_t)$$

show some linear relationship that can be exploited. When that can be done, predictions of future residuals can be added to "trend plus seasonal" forecasts for future periods.

Figure 10 shows the residuals and "lag 1 residuals" for the linear trend plus seasonal fit to the JC Penney sales data in the data table.

Figure 10: Residuals e_t and Lag 1 Residuals e_{t-1} for the JC Penney Data

Period	JC Penney Sales	1st Quarter	2nd Quarter	3rd Quarter	Predicted JC Penney Sales	Residual JC Penney Sales	Lag 1 Residuals
1	4452	1	0	0	5684.08929	-1232.0893	
2	4507	0	1	0	5493.25595	-986.25595	-1232.0893
3	4537	0	0	1	6134.58929	-597.58929	-986.25595
4	8157	0	0	0	8256.92262	-99.922619	-597.58929
5	6481	1	0	0	6082.25357	398.746429	-99.922619
6	6420	0	1	0	5891.42024	528.579762	398.746429
7	7208	0	0	1	6532.75357	675.246429	528.579762
8	9509	0	0	0	8655.0869	853.913095	675.246429
9	6755	1	0	0	6480.41786	274.582143	853.913095
10	6483	0	1	0	6289.58452	193.415476	274.582143
11	7129	0	0	1	6930.91786	198.082143	193.415476
12	9072	0	0	0	9053.25119	18.7488095	198.082143
13	7339	1	0	0	6878.58214	460.417857	18.7488095
14	7104	0	1	0	6687.74881	416.25119	460.417857
15	7639	0	0	1	7329.08214	309.917857	416.25119
16	9861	0	0	0	9451.41548	209.584524	309.917857
17	7528	1	0	0	7276.74643	251.253571	209.584524
18	7207	0	1	0	7085.9131	121.086905	251.253571
19	7538	0	0	1	7727.24643	-189.24643	121.086905
20	9573	0	0	0	9849.57976	-276.57976	-189.24643
21	7522	1	0	0	7674.91071	-152.91071	-276.57976
22	7211	0	1	0	7484.07738	-273.07738	-152.91071
23	7729	0	0	1	8125.41071	-396.41071	-273.07738
24	9542	0	0	0	10247.744	-705.74405	-396.41071

33

Next, there are 3 plots. In the first e_t is plotted against t and in the second, e_t is plotted against e_{t-1} . These plots (in Figures 11 and 12) show the same thing in different terms. There is a time pattern in the residuals. So consecutive residuals tend to be big (positive) together and small (negative) together. That is because the fitted model over-predicts early in the data set and late in the data set, and under-predicts in the middle of the data set. That can also be seen if one looks carefully at the third plot of both y_t versus t and \hat{y}_t versus t (Figure 13).

34

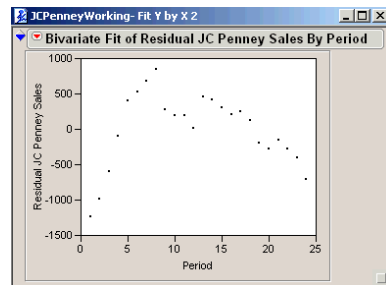


Figure 11: Plot of Residuals versus Period for the JC Penney Data

35

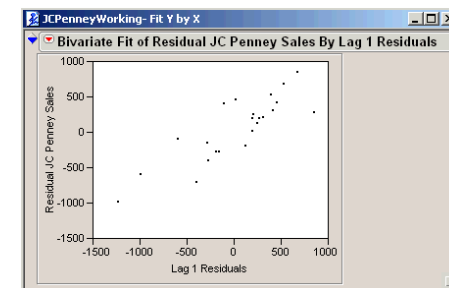


Figure 12: Plot of Residual e_t versus Lag 1 Residual e_{t-1} for the JC Penney Data

36

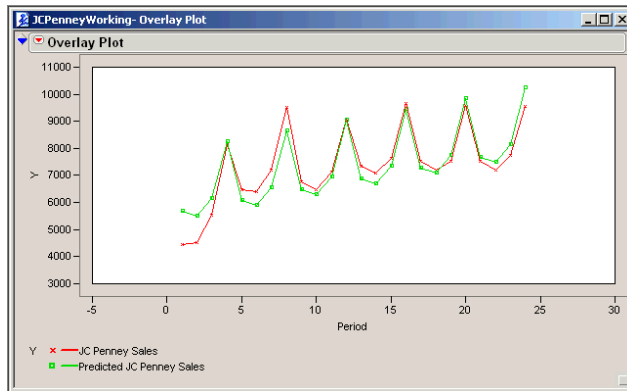


Figure 13: JC Penney Sales and Fitted Sales

The pattern in Figure 12 suggests that one might predict a residual from the immediately preceding residual using some form of regression. Figure 14 shows that using simple linear regression of residuals on lag 1 residuals gives a fitted equation

$$\hat{e}_t = 30.26762 + 0.7593887e_{t-1}$$

Notice that this means that from the last point in the JC Penney data set (period 24) it is possible to predict the residual at period 25, since the residual for period 24 will then be known! That is

$$\hat{e}_{25} = 30.26762 + 0.7593887e_{24}$$

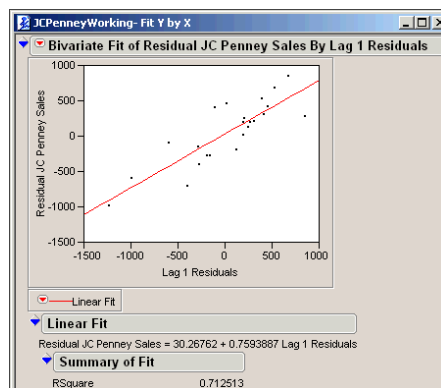


Figure 14: JMP Report for SLR of Residual on Lag 1 Residual

In fact, this line of thinking suggests that we can improve on the forecast of y_{25} based solely on linear trend plus seasonal ($\hat{y}_{25} = 8073$) by using

$$\hat{y}_{25} + \hat{e}_{25}$$

Looking in the data table of Figure 10, we see that the residual in the final period of the data set is

$$e_{24} = -705.74405$$

and thus that

$$\hat{e}_{25} = 30.26762 + 0.7593887(-705.74405) = -506$$

so that what might be an improved forecast for period 25 is

$$8073 + (-506) = 7567$$

The basic idea of predicting residuals from previous residuals can be carried even further. One can try predicting a residual on the basis of not only the

immediately preceding one, but the immediately preceding two (or more). That is, it is possible to regress e_t on e_{t-1} and e_{t-2} in order to come up with a way of forecasting a next residual (and therefore improving a trend plus seasonal forecast). We will not show any details here (for one thing because the idea doesn't really offer any improvement in the JC Penney example), but the idea should be clear.

Case Study: JMP Airline Passenger Count Data

In the "Sample Data" provided with a JMP installation are some time series data. "Seriesg.jmp" gives 12 years worth of monthly airline passenger counts taken from the time series book of Box and Jenkins. (The data are from

January 1949 through December 1960 and the counts are in thousands of passengers.) This data set can be used to admirably demonstrate the topics discussed here. (Although we have made use of the JC Penney data set for illustrative purposes, it is far smaller than the minimum size that should really be used in a time series analysis. The length 144 airline passenger data set is closer to being of practical size for reliable development of forecasts.)

Figure 15 is a plot of the raw passenger counts versus time.

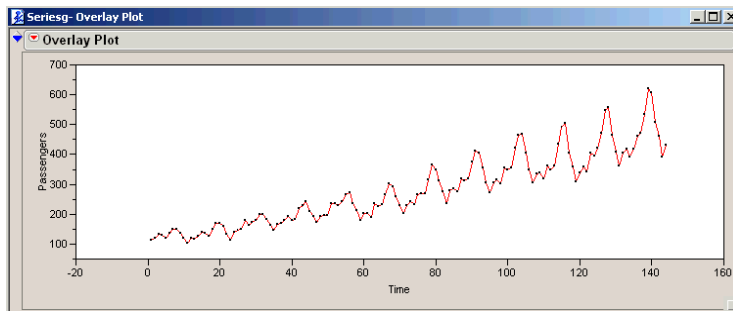


Figure 15: Airline Passenger Counts Time Series

Figure 15 has a feature that is common to many economic time series of any appreciable length. Namely, as time goes on, the "local" or short term variation seems to increase as the general level of the count increases. Besides, it looks like the general trend of count versus time may not be linear, but rather have

some upward curvature. It is far easier to fit and forecast series that don't have these features. So what we can do is to try to *transform* the raw counts, fit and forecast with the transformed series, and then "untransform" to make final interpretations. That is, we will analyze the (base 10) logarithms of passenger counts

$$y_t = \log_{10}(\text{passenger count at period } t)$$

Figure 16 is a plot of y_t and happily looks "better" than the original series in Figure 15 for purposes of fitting and forecasting.

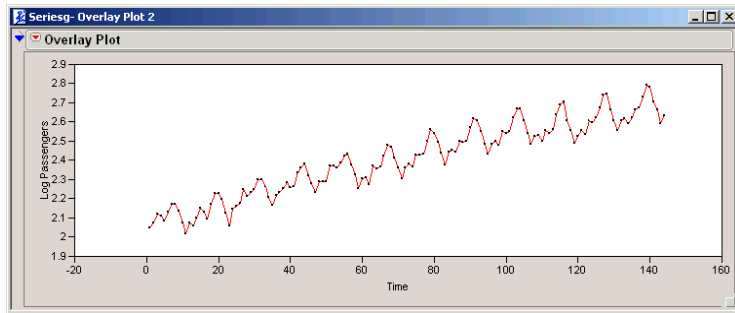


Figure 16: Logarithms of Passenger Counts

A first step in analysis of the y_t series is perhaps to see how a linear trend does at describing the data. We can use JMP to do SLR and fit a line to the (t, y_t) values and save the predictions. These can then be plotted using "Overlay Plot" along with the original series to get Figure 17.

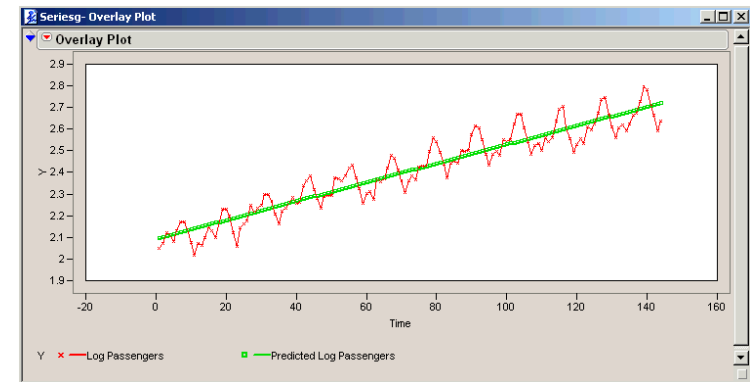


Figure 17: Linear Trend Fit to y_t Series

Of course, the linear trend ignores the seasonality in the times series. Since these are monthly data, we could define 11 monthly indicator variables. But that would be tedious, and happily the JMP data table (partially pictured in Figure 18) has the month information coded into it in the form of a "nominal" variable "Season." Since "Season" is a "nominal" variable (indicated by the red cap N) if we tell JMP's "Fit Model" routine to use it in a multiple regression, it will automatically use the single nominal variable to create $12 - 1 = 11$ dummy variables for all but one of the values of "Season." That is, we may fill in the "Fit Model" dialogue box as in Figure 19 to get fitted values for the "linear trend plus seasonal" model.

	Passengers	Time	Season	Log Passengers
1	112	1	1	2.04921802
2	118	2	2	2.07188201
3	132	3	3	2.12057383
4	129	4	4	2.11058971
5	121	5	5	2.08278537
6	135	6	6	2.13033377
7	148	7	7	2.17026172
8	148	8	8	2.17026172
9	136	9	9	2.13353891
10	119	10	10	2.07554696
11	104	11	11	2.01703334
12	118	12	12	2.07188201
13	115	13	1	2.06069784
14	126	14	2	2.10037055
15	141	15	3	2.14921911

Figure 18: Partial JMP Data Table for the Airline Passenger Data

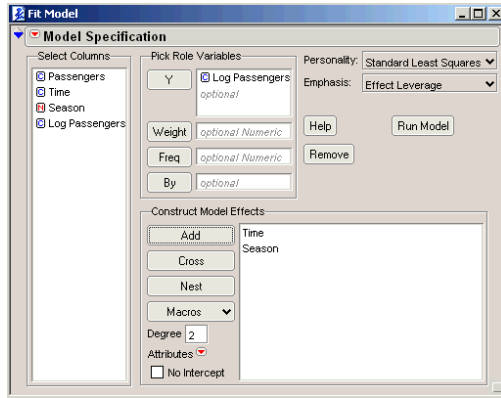


Figure 19: JMP "FIT Model" Dialogue Box for Linear Trend Plus Seasonal Fit

A partial JMP report for the fitting indicated in Figure 19 is shown in Figure 20. A plot of the fitted values for the linear trend plus seasonal model is shown in Figure 21.

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	2.0899065	0.004326	483.08	<.0001
Time	0.0043728	0.000052	84.40	<.0001
Season(1)	-0.037092	0.007124	-5.21	<.0001
Season(2)	-0.04667	0.007122	-6.55	<.0001
Season(3)	0.0098867	0.007121	1.39	0.1674
Season(4)	-0.003693	0.00712	-0.52	0.6048
Season(5)	-0.004724	0.007119	-0.66	0.5082
Season(6)	0.0483238	0.007118	6.79	<.0001
Season(7)	0.0934654	0.007118	13.13	<.0001
Season(8)	0.0894287	0.007119	12.56	<.0001
Season(9)	0.0266147	0.00712	3.74	0.0003
Season(10)	-0.033387	0.007121	-4.69	<.0001
Season(11)	-0.095802	0.007122	-13.45	<.0001

Figure 20: Partial JMP Report for Linear Trend Plus Seasonal Fit to y_t

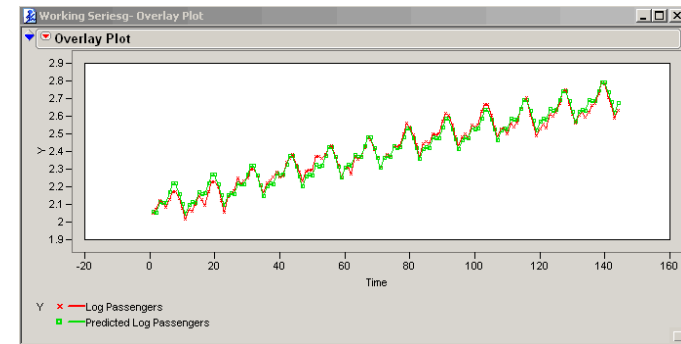


Figure 21: Linear Trend Plus Seasonal Fit to Logarithms of Passenger Counts

Of course, the fit indicated in Figure 21 is better than the one in Figure 17. And the forecasts provided by the regression model can be extended into the "future" (beyond $t = 144$ that represents the last point in the data set). But there is even more that can be done if one considers the nature of the residuals from the regression fit. Figure 22 shows a plot of the residuals $e_t = y_t - \hat{y}_t$ versus t and Figure 23 shows that there is a fair amount of correlation between residuals and lagged residuals. (This is no surprise given the nature of the plot in Figure 22 where "slow" trends in the residuals make ones close together in time similar in value.)

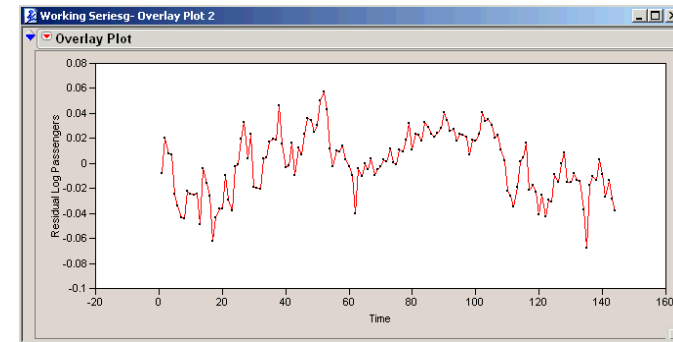


Figure 22: Plot of Residuals versus t for the Log Passenger Counts

	Residual Log Passengers	Lag 1 Residuals	Lag 2 Residuals
Residual Log Passengers	1.0000	0.7896	0.6722
Lag 1 Residuals	0.7896	1.0000	0.7832
Lag 2 Residuals	0.6722	0.7832	1.0000

Figure 23: Residuals, Lag 1 Residuals, and Lag 2 Residuals for Log Passenger Counts

It is possible (by examining regressions of residuals on lagged residuals) to come to the conclusion that in terms of predicting residuals from earlier residuals it suffices to simply use the single previous one (nothing important is gained by using the two previous ones). And in fact, for this problem, an appropriate prediction equation (coming from SLR of e_t on e_{t-1}) is

$$\hat{e}_t = -0.000153 + 0.7918985e_{t-1}$$

This can be used to adjust the fits/predictions from the linear trend plus seasonal model of log counts as

$$(\text{adjusted fit})_t = \hat{y}_t + \hat{e}_t$$

These are plotted along with the original series and the earlier fitted values \hat{y}_t in Figure 24. There is a small, but clearly discernible improvement in the quality of the modeling provided by this adjustment for serial correlation in the residuals.

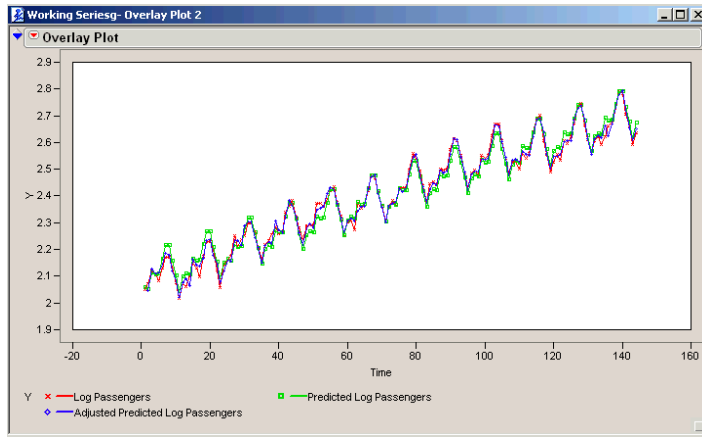


Figure 24: Original Values y_t , Fitted Values \hat{y}_t , and Adjusted Fitted Values $\hat{y}_t + \hat{e}_t$

Notice then that an adjusted forecast of log passenger count for period $t = 145$ (the January/Season 1 following the end of the data set) becomes

$$\begin{aligned} \hat{y}_{145} + \hat{e}_{145} &= (2.0899065 + (-0.037092) + .0043728(145)) \\ &\quad + (-0.000153 + 0.7918985(-0.0377583)) \\ &= 2.65682038 \end{aligned}$$

This figure is (of course) on a log scale. We may "untransform" this value in order to get a forecast for a passenger count (as opposed to a log passenger count). This is

$$10^{2.65682038} = 454$$

In fact, it is worthwhile to see a final plot, that compares the original series of counts to the whole set of values

$$10^{\hat{y}_t + \hat{e}_t}$$

that function as fitted values on the original (count) scale. This is shown in Figure 25 (including the value for period 145, whose plotted symbol is larger than the others, and represents a forecast beyond the end of the original data set).

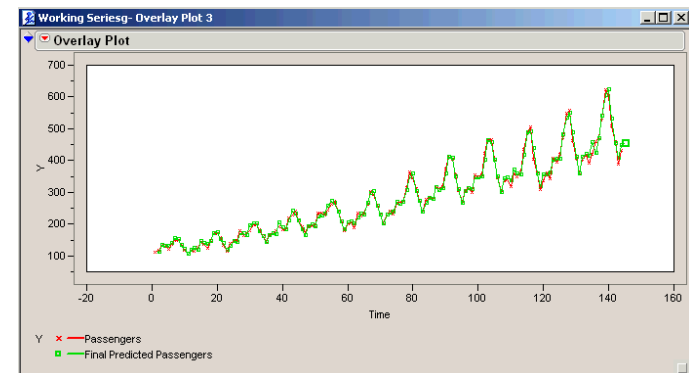


Figure 25: Plot of Passenger Counts and Final Fitted Values $10^{\hat{y}_t + \hat{e}_t}$ (Including a Forecast for $t = 145$)