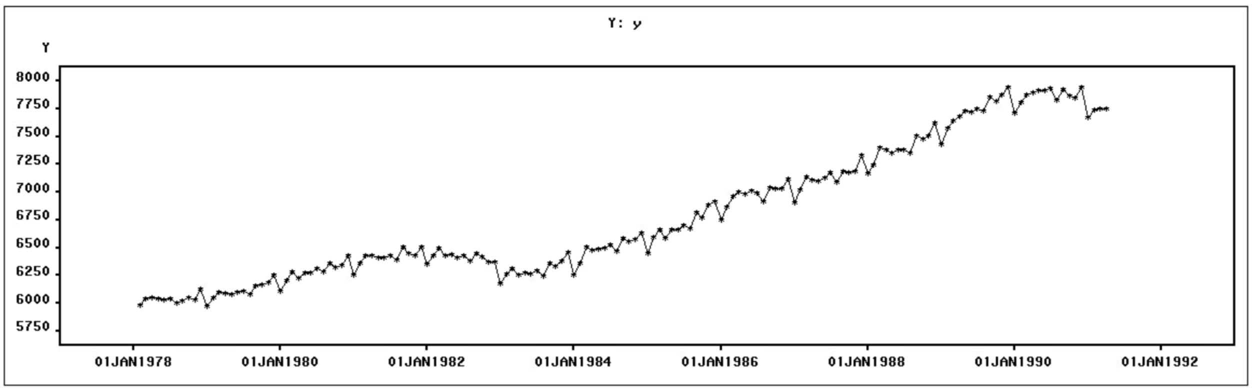
Justin Jiang

May 22, 2023

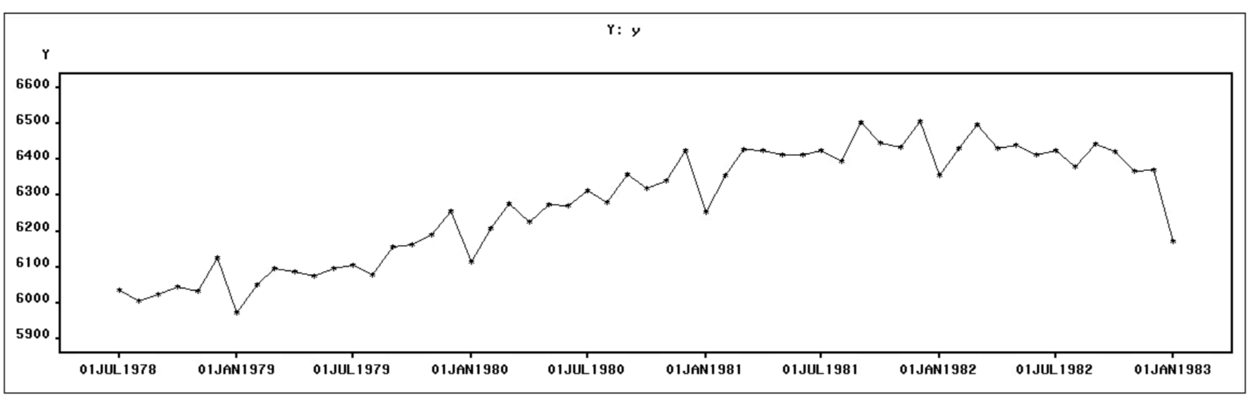
Time Series Analysis

<u>Step 1</u>: Analysis

The data set shows employment of people in Australia between February 1978 to April 1991. The data set contains 160 entries, so most likely the forecast will forecast in ranges of months. Importing the data to SAS to plot, this is one part of the plot:

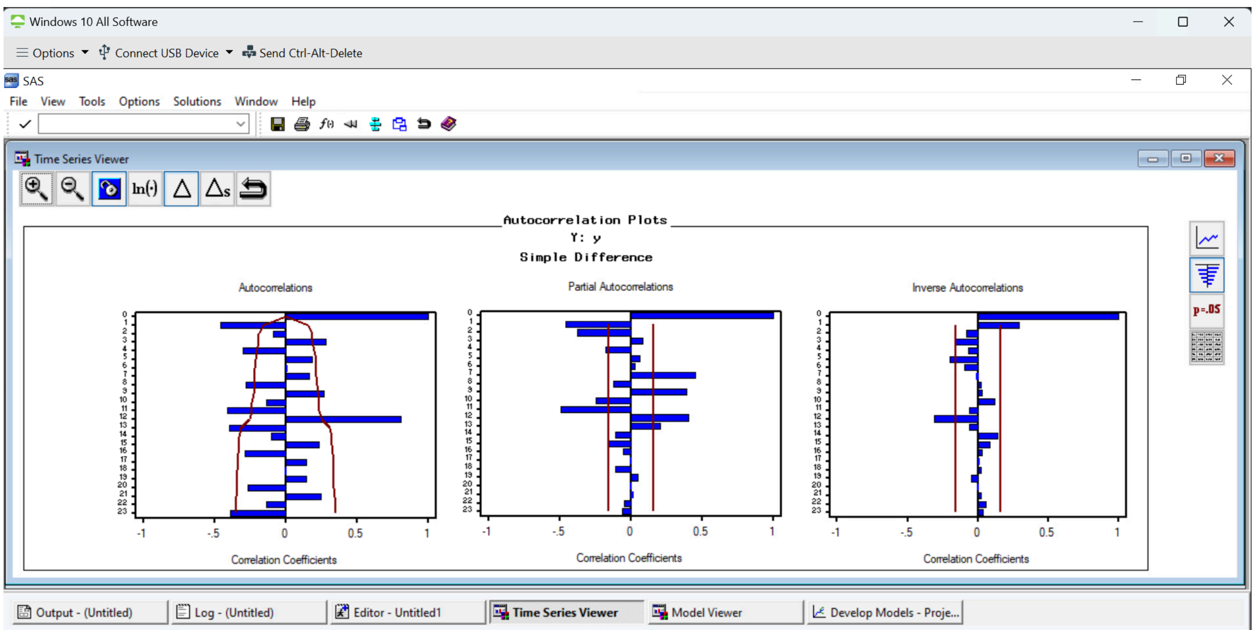


Looking at the plot, it can be assumed that the data is not stationary because the plot is showing signs of growth andno signs of being rigid. As seen from the graph below, there is some pattern as the time increases implying some form of seasonality The pattern can be seen between each year as there is the same steady increase, sharp decrease, steep increase and steady decrease..



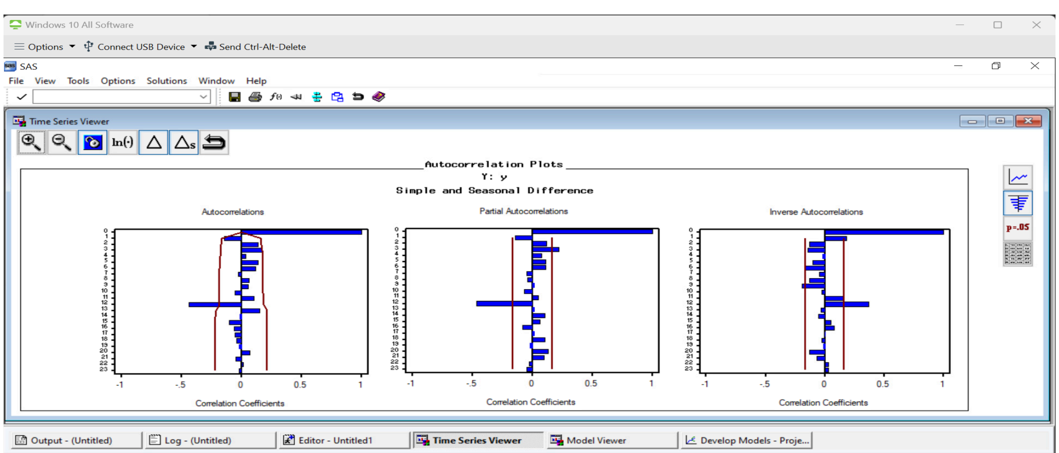
Thus, we can use some form of a trend curve such as a linear trend or exponential trend to determine what best fits the model because the data isn't stationary. Furthermore, since the data is seasonal, the use of seasonal differencing and seasonal dummies can be used in the model. Let's transform the data to gain an understanding of what factors like autocorrelation and moving average might be needed to consider in developing the model.

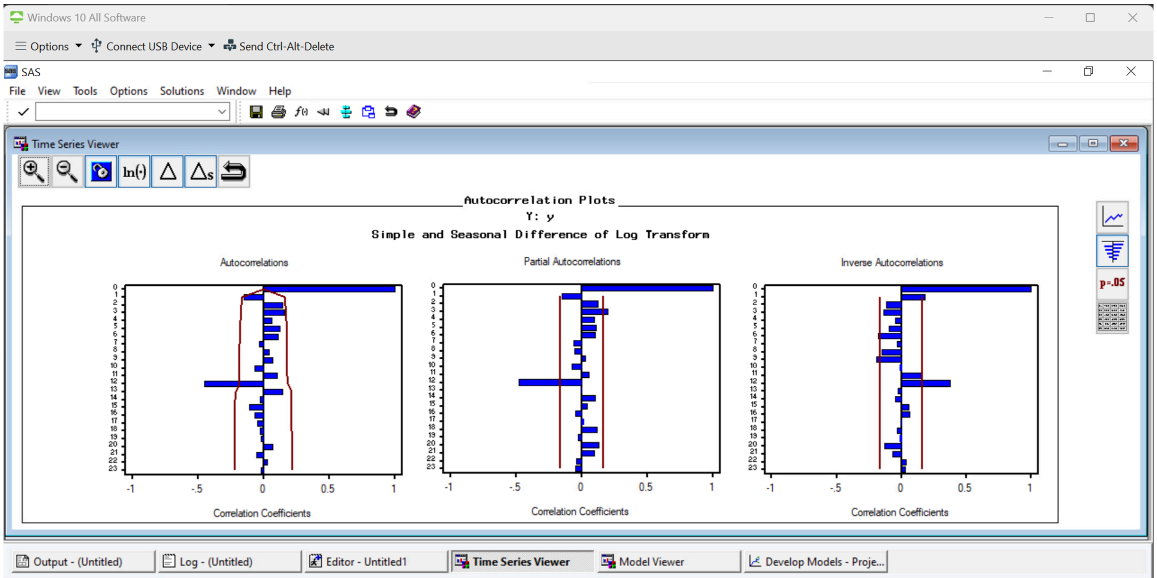
The model after simple differencing:



There is a pattern occurring in the autocorrelation so seasonal differencing is needed.

The model after seasonal differencing and log transform:



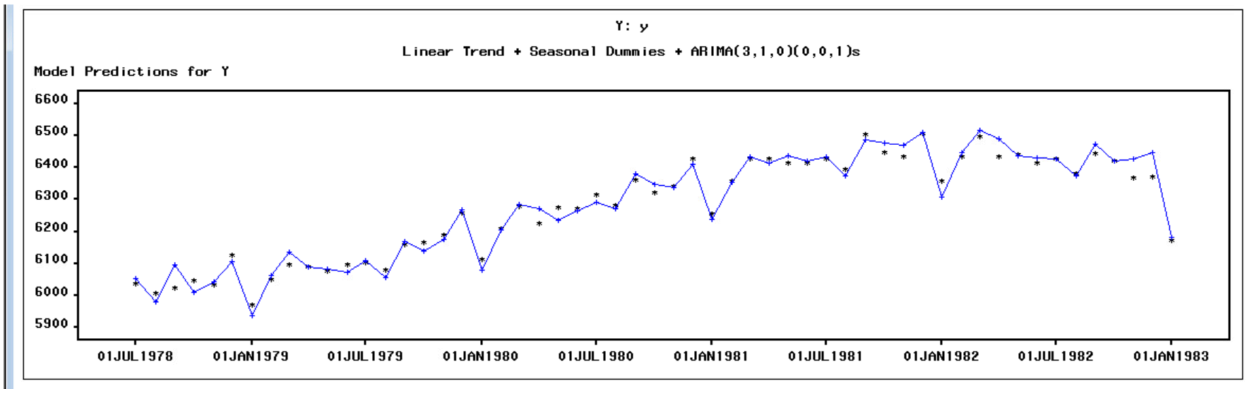


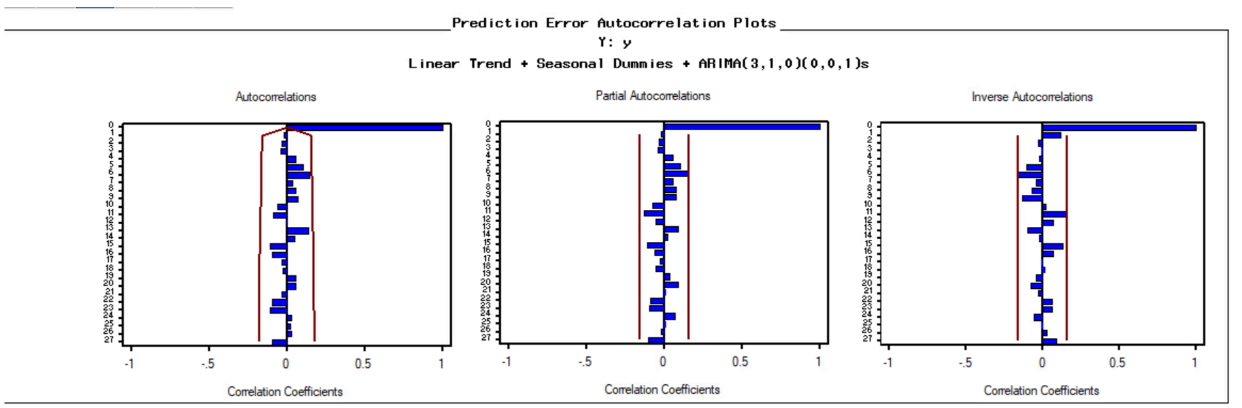
Looking at the transformation, it can be assumed that there is a moving average of 1 and 12 because coefficient 1 and 12 passes the bound in the autocorrelation and autocorrelation of 3 because coefficient 3 passes the boundary . Since the moving average is 1 and 12, moving average can be Q = 1 since the data is also seasonal. Models that can be attempted now can be a trend curve with ARIMA where AR = 3 and MA = 1, 12 with differencing and seasonal dummies. Winter's Smoothing and log transform on developing will be used to compare created models.

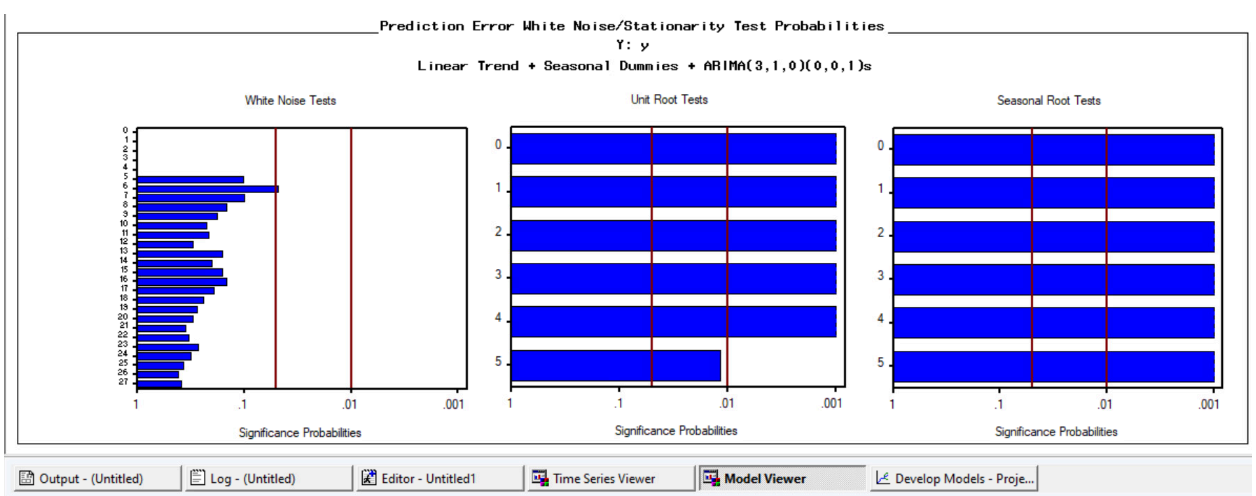
Step 2: Developing Models

For choosing our model, it can be deduced that the model being developed should be an ARIMA model with a trend curve with p = 3, d =1, and Q = 1 with seasonal dummies.

The ARIMA model with a linear trend is:







Parameter Estimates

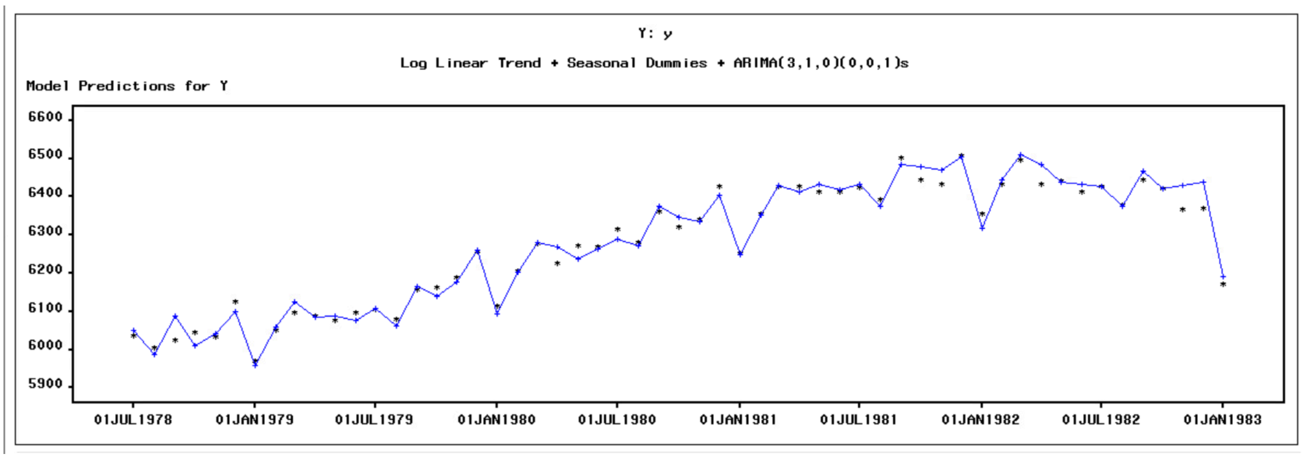
Y: y

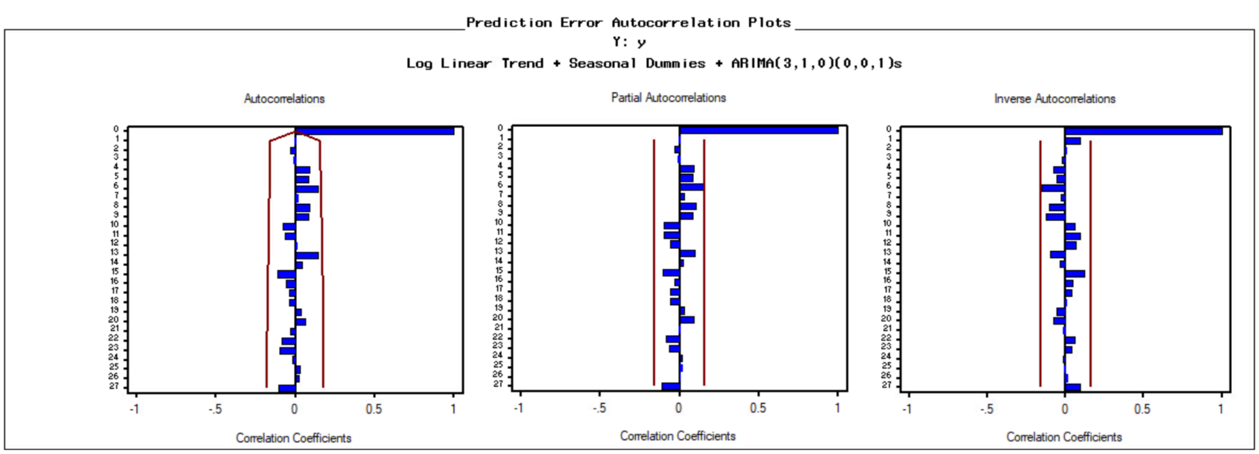
Linear Trend + Seasonal Dummies + ARIMA(3,1,0)(0,0,1)s

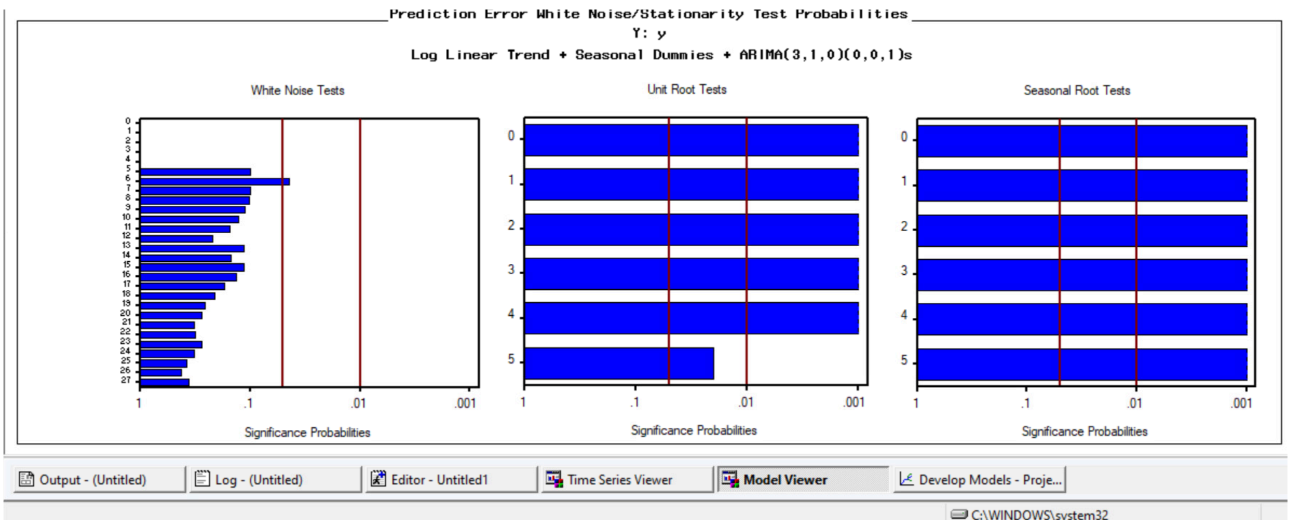
Model Parameter	Estimate	Std. Error	T	Prob>|T|
Seasonal Dummy 1	-264.62681	14.0394	-18.8489	<.0001
Seasonal Dummy 2	24.94383	11.9657	2.0846	0.0389
Seasonal Dummy 3	-1.31609	12.0242	-0.1095	0.9130
Seasonal Dummy 4	-90.76359	13.0085	-6.9772	<.0001
Seasonal Dummy 5	-68.26780	12.8087	-5.3298	<.0001
Seasonal Dummy 6	-73.29901	13.0202	-5.6296	<.0001
Seasonal Dummy 7	-55.64691	12.7804	-4.3541	<.0001
Seasonal Dummy 8	-125.63012	13.2682	-9.4685	<.0001
Seasonal Dummy 9	23.90585	12.1914	1.9609	0.0519
Seasonal Dummy 10	-101.23203	11.9960	-8.4388	<.0001
Seasonal Dummy 11	-60.35825	14.0317	-4.3016	<.0001
Model Variance (sigma squared)	914.63382	.	.	.

Fit Range: FEB1978 to APR1991

The ARIMA model with linear trend (Log transform):







Parameter Estimates

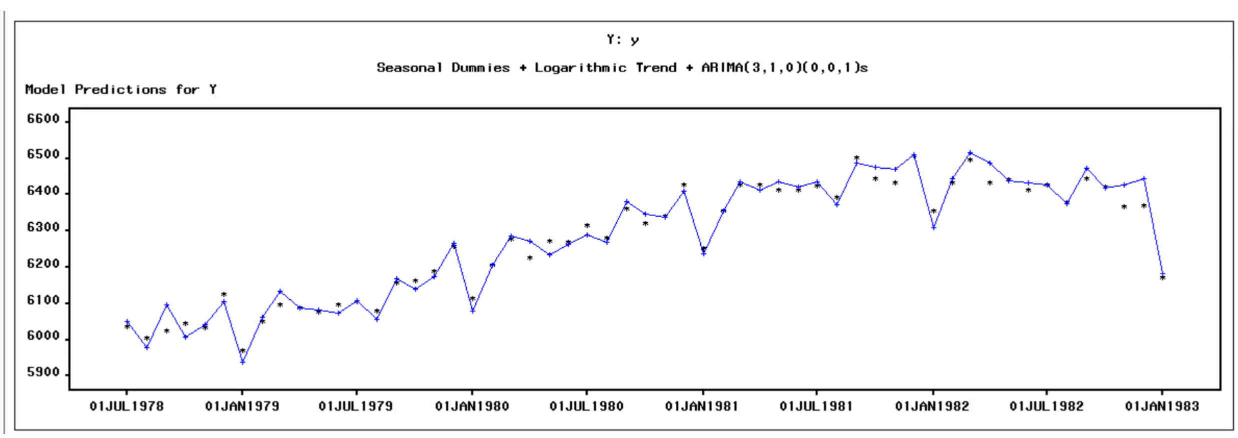
Y: y

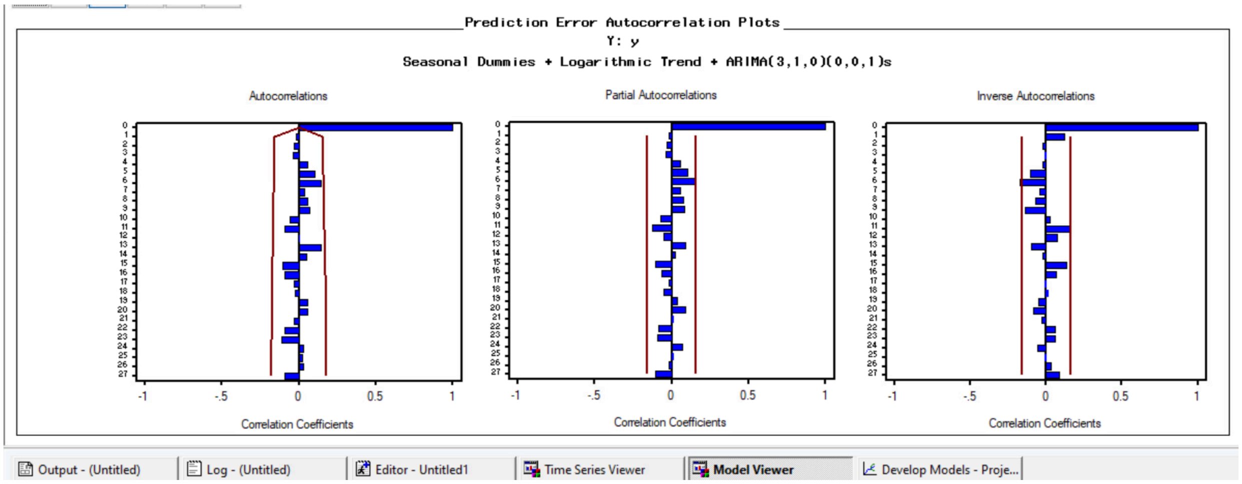
Log Linear Trend + Seasonal Dummies + ARIMA(3,1,0)(0,0,1)s

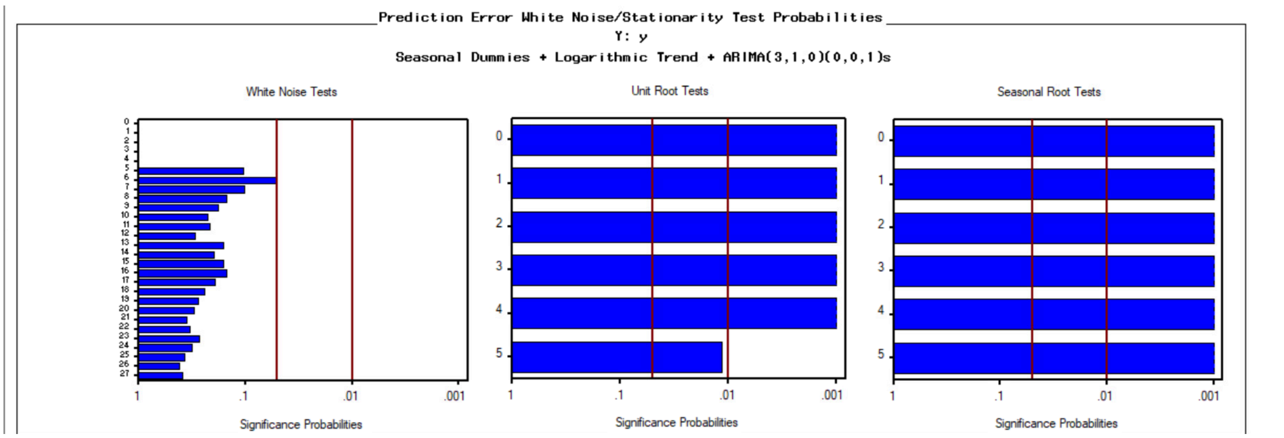
Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	0.01088	0.0015	7.3002	<.0001
Seasonal Moving Average, Lag 12	-0.00974	0.0904	-0.1078	0.9143
Autoregressive, Lag 1	-0.11198	0.0831	-1.3474	0.1800
Autoregressive, Lag 2	0.19083	0.0841	2.2703	0.0247
Autoregressive, Lag 3	0.18107	0.0863	2.0988	0.0376
Linear Trend	2.51093E-6	0.000010	0.2501	0.8029
Seasonal Dummy 1	-0.03860	0.0018	-20.9793	<.0001
Seasonal Dummy 2	0.00405	0.0016	2.5614	0.0115
Seasonal Dummy 3	0.0001015	0.0016	0.0633	0.9496
Seasonal Dummy 4	-0.01344	0.0017	-7.8387	<.0001
Seasonal Dummy 5	-0.00988	0.0017	-5.8086	<.0001
Seasonal Dummy 6	-0.01060	0.0017	-6.1424	<.0001
Seasonal Dummy 7	-0.00799	0.0017	-4.7091	<.0001
Seasonal Dummy 8	-0.01823	0.0017	-10.4291	<.0001
Seasonal Dummy 9	0.00366	0.0016	2.2523	0.0258
Seasonal Dummy 10	-0.01462	0.0016	-9.2182	<.0001

Fit Range: FEB1978 to APR1991

ARIMA model with Logarithmic Trend:







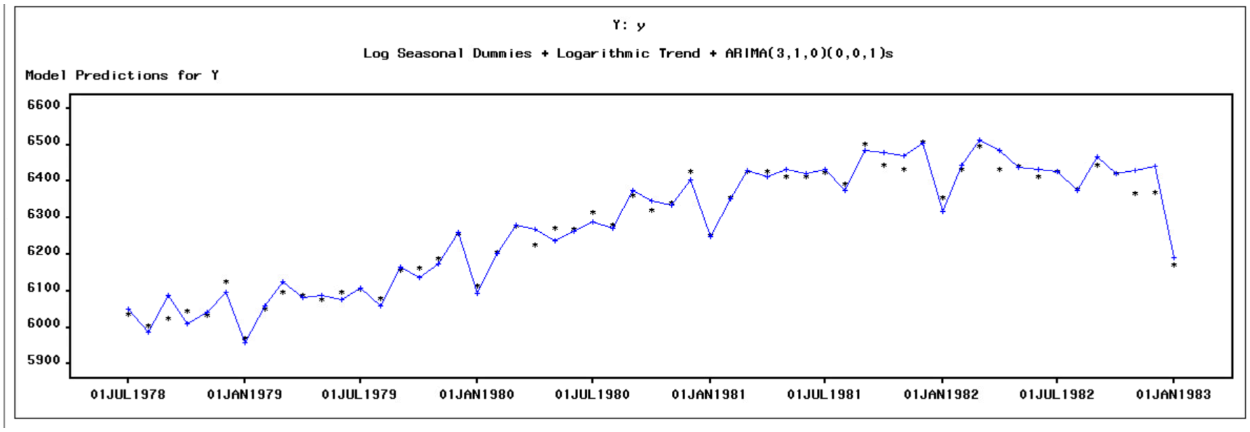
Y: y

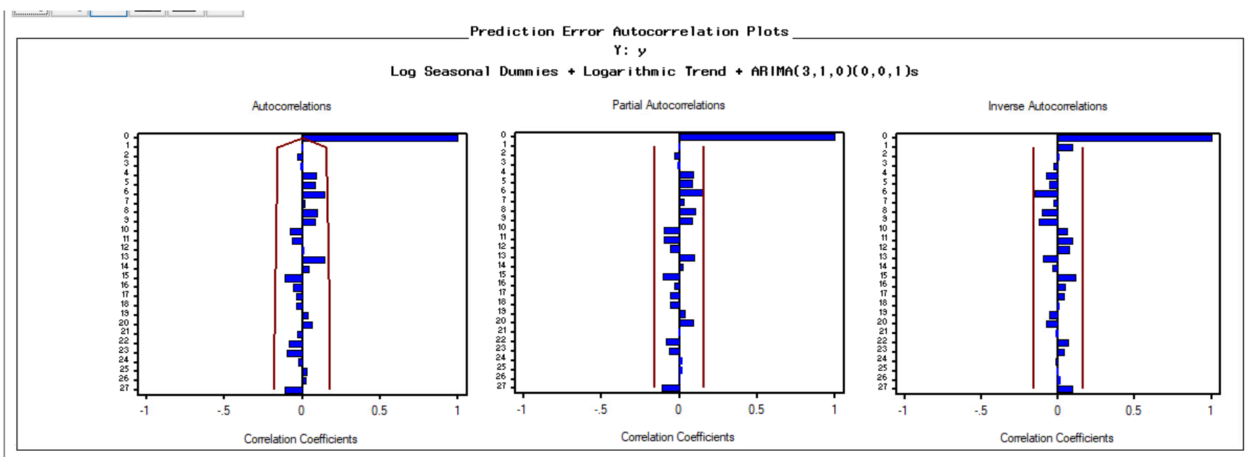
Seasonal Dummies + Logarithmic Trend + ARIMA(3,1,0)(0,0,1)s

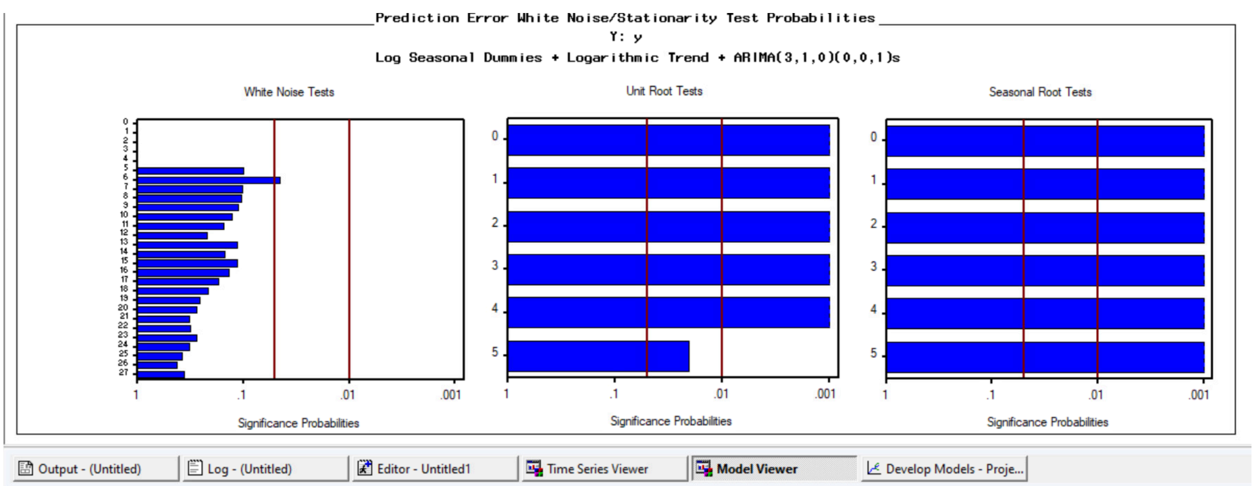
Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	68.55305	17.7323	3.8660	0.0002
Seasonal Moving Average, Lag 12	-0.07926	0.0908	-0.8731	0.3841
Autoregressive, Lag 1	-0.13402	0.0828	-1.6194	0.1076
Autoregressive, Lag 2	0.18747	0.0836	2.2426	0.0265
Autoregressive, Lag 3	0.20014	0.0850	2.3555	0.0199
Seasonal Dummy 1	-264.58859	14.0287	-18.8605	<.0001
Seasonal Dummy 2	24.91164	11.9638	2.0823	0.0391
Seasonal Dummy 3	-1.06799	12.0297	-0.0888	0.9294
Seasonal Dummy 4	-90.58617	13.0012	-6.9675	<.0001
Seasonal Dummy 5	-68.12598	12.8030	-5.3211	<.0001
Seasonal Dummy 6	-73.21066	13.0084	-5.6280	<.0001
Seasonal Dummy 7	-55.58885	12.7716	-4.3525	<.0001
Seasonal Dummy 8	-125.59586	13.2543	-9.4759	<.0001
Seasonal Dummy 9	23.91599	12.1840	1.9629	0.0516
Seasonal Dummy 10	-101.22790	11.9932	-8.4405	<.0001
Seasonal Dummy 11	-60.37841	14.0206	-4.3064	<.0001

Fit Range: FEB1978 to APR1991

ARIMA model with logarithmic trend (Log transform):







Parameter Estimates

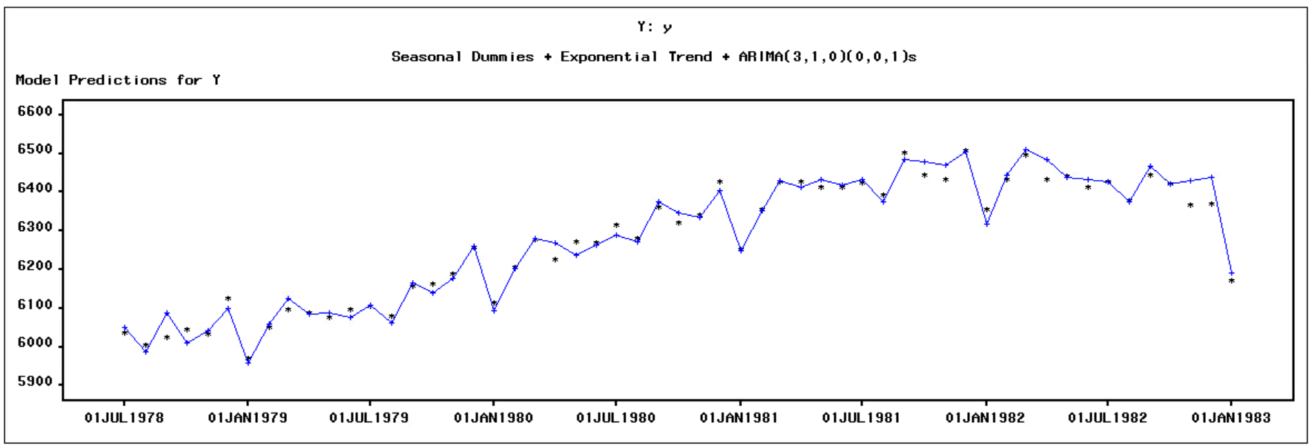
Y: y

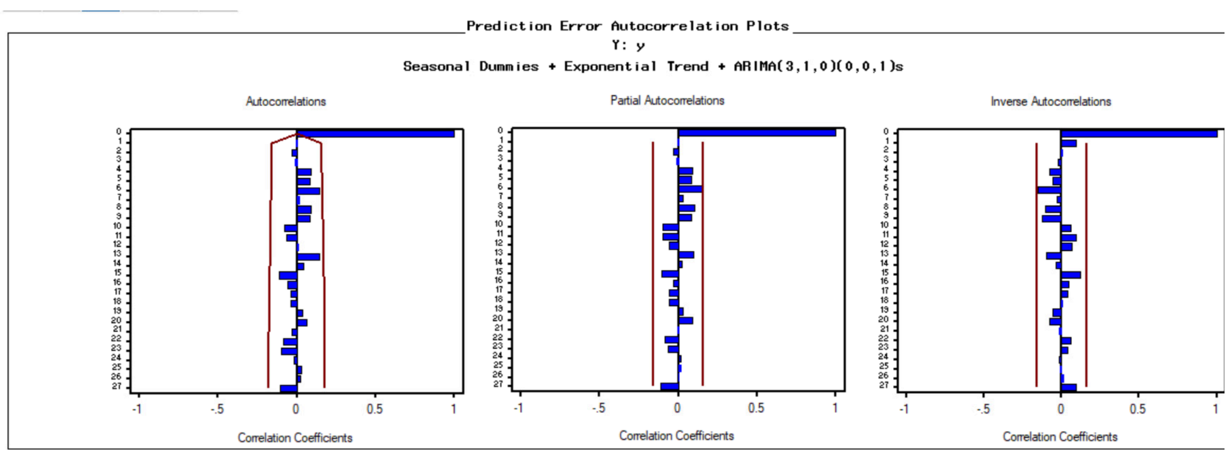
Log Seasonal Dummies + Logarithmic Trend + ARIMA(3,1,0)(0,0,1)s

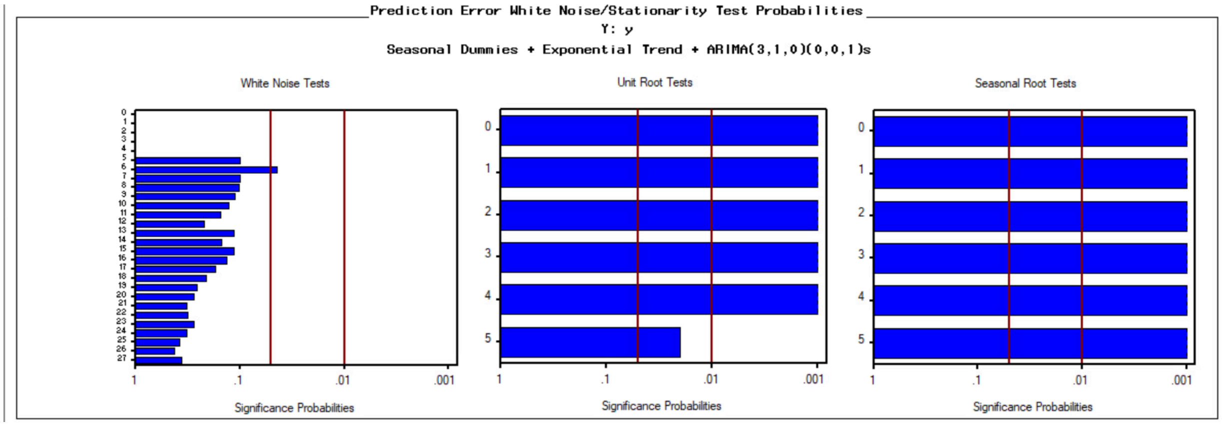
Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	0.01028	0.0024	4.2737	<.0001
Seasonal Moving Average, Lag 12	-0.00916	0.0901	-0.1016	0.9192
Autoregressive, Lag 1	-0.11262	0.0831	-1.3548	0.1776
Autoregressive, Lag 2	0.18892	0.0841	2.2475	0.0262
Autoregressive, Lag 3	0.17973	0.0862	2.0839	0.0390
Seasonal Dummy 1	-0.03860	0.0018	-20.9921	<.0001
Seasonal Dummy 2	0.00405	0.0016	2.5600	0.0115
Seasonal Dummy 3	0.0001266	0.0016	0.0789	0.9372
Seasonal Dummy 4	-0.01342	0.0017	-7.8315	<.0001
Seasonal Dummy 5	-0.00987	0.0017	-5.8014	<.0001
Seasonal Dummy 6	-0.01059	0.0017	-6.1413	<.0001
Seasonal Dummy 7	-0.00799	0.0017	-4.7080	<.0001
Seasonal Dummy 8	-0.01822	0.0017	-10.4359	<.0001
Seasonal Dummy 9	0.00366	0.0016	2.2547	0.0257
Seasonal Dummy 10	-0.01462	0.0016	-9.2197	<.0001
Seasonal Dummy 11	-0.00872	0.0018	-4.7424	<.0001

Fit Range: FEB1978 to APR1991

ARIMA model with exponential curve:







Parameter Estimates

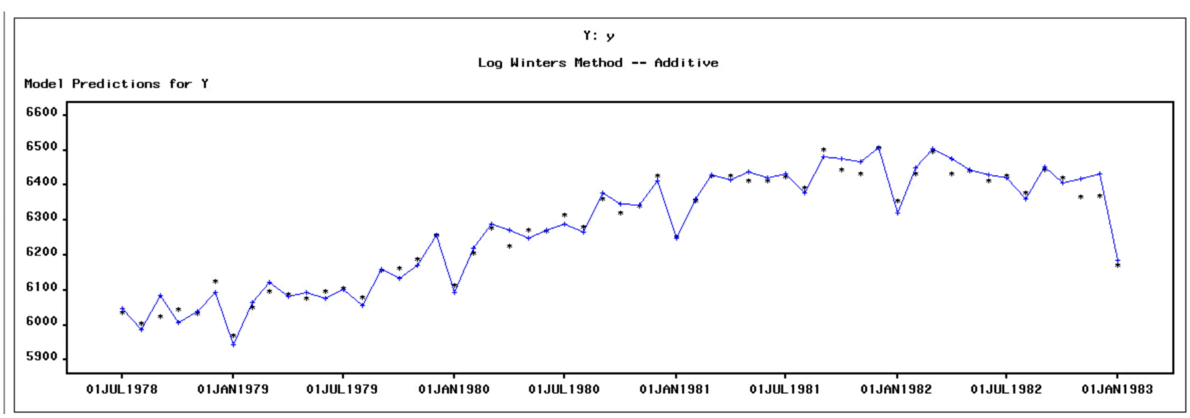
Y: y

Seasonal Dummies + Exponential Trend + ARIMA(3,1,0)(0,0,1)s

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	0.01088	0.0015	7.3002	<.0001
Seasonal Moving Average, Lag 12	-0.00974	0.0904	-0.1078	0.9143
Autoregressive, Lag 1	-0.11198	0.0831	-1.3474	0.1800
Autoregressive, Lag 2	0.19083	0.0841	2.2703	0.0247
Autoregressive, Lag 3	0.18107	0.0863	2.0988	0.0376
Seasonal Dummy 1	-0.03860	0.0018	-20.9793	<.0001
Seasonal Dummy 2	0.00405	0.0016	2.5614	0.0115
Seasonal Dummy 3	0.0001015	0.0016	0.0633	0.9496
Seasonal Dummy 4	-0.01344	0.0017	-7.8387	<.0001
Seasonal Dummy 5	-0.00988	0.0017	-5.8086	<.0001
Seasonal Dummy 6	-0.01060	0.0017	-6.1424	<.0001
Seasonal Dummy 7	-0.00799	0.0017	-4.7091	<.0001
Seasonal Dummy 8	-0.01823	0.0017	-10.4291	<.0001
Seasonal Dummy 9	0.00366	0.0016	2.2523	0.0258
Seasonal Dummy 10	-0.01462	0.0016	-9.2182	<.0001
Seasonal Dummy 11	-0.00872	0.0018	-4.7381	<.0001

Fit Range: FEB1978 to APR1991

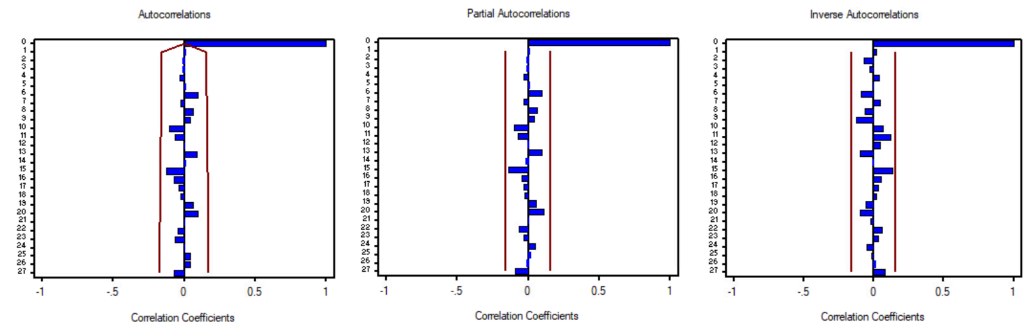
Winter Smoothing with log transform:



Prediction Error Autocorrelation Plots

Y: y

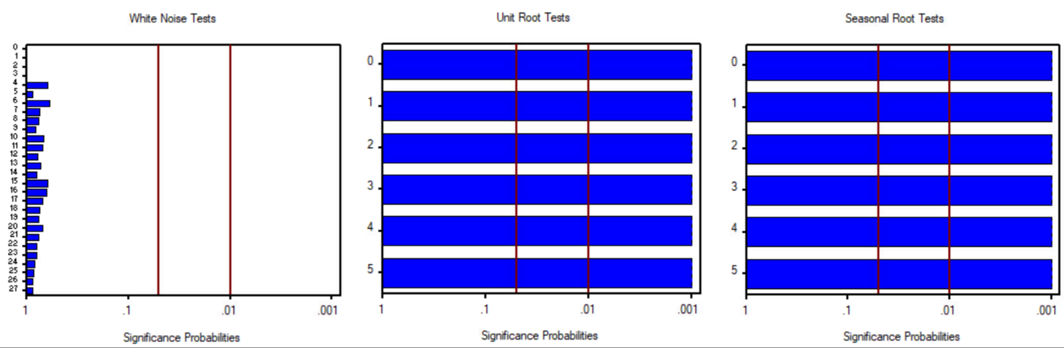
Log Winters Method -- Additive



Prediction Error White Noise/Stationarity Test Probabilities

Y: y

Log Winters Method -- Additive



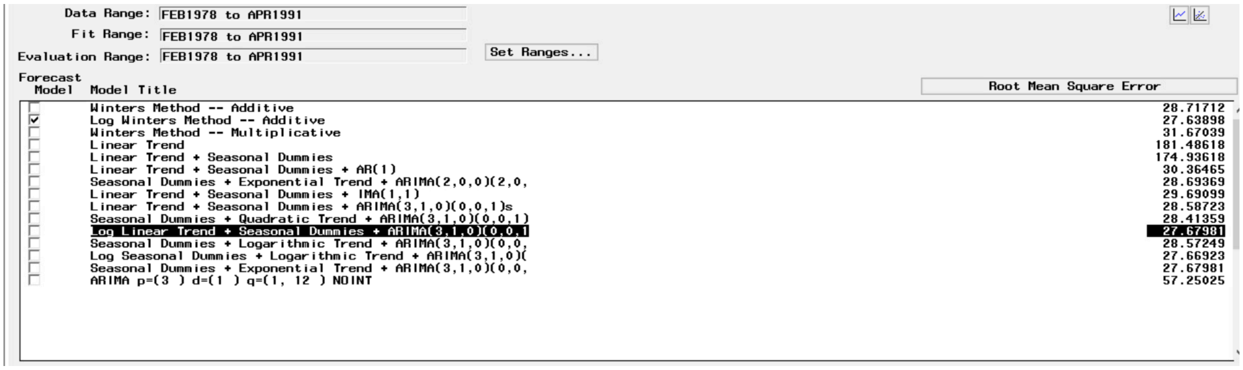
Parameter Estimates

Y: y

Log Winters Method -- Additive

Model Parameter	Estimate	Std. Error	T	Prob>|T|
LEVEL Smoothing Weight	0.68202	0.0553	12.3384	<.0001
TREND Smoothing Weight	0.24456	0.0506	4.8362	<.0001
SEASONAL Smoothing Weight	0.00100	0.0445	0.0225	0.9821
Residual Variance (sigma squared)	0.0000166	.	.	.
Smoothed Level	8.95211	.	.	.
Smoothed Trend	-0.00272	.	.	.
Smoothed Seasonal Factor 1	-0.01893	.	.	.
Smoothed Seasonal Factor 2	-0.00366	.	.	.
Smoothed Seasonal Factor 3	0.00572	.	.	.
Smoothed Seasonal Factor 4	0.00155	.	.	.
Smoothed Seasonal Factor 5	0.00186	.	.	.
Smoothed Seasonal Factor 6	0.0005965	.	.	.
Smoothed Seasonal Factor 7	0.00187	.	.	.
Smoothed Seasonal Factor 8	-0.00703	.	.	.
Smoothed Seasonal Factor 9	0.00591	.	.	.
Smoothed Seasonal Factor 10	0.0005692	.	.	.

RSME:



Data Range: FEB1978 to APR1991
Fit Range: FEB1978 to APR1991
Evaluation Range: FEB1978 to APR1991
Set Ranges...

Forecast Model	Model Title	Root Mean Square Error
	Winters Method -- Additive	28.71712
✔	Log Winters Method -- Additive	27.63898
	Winters Method -- Multiplicative	31.67039
	Linear Trend	181.48618
	Linear Trend + Seasonal Dummies	174.93618
	Linear Trend + Seasonal Dummies + AR(1)	30.36465
	Seasonal Dummies + Exponential Trend + ARIMA(2,0,0)(2,0,	28.69369
	Linear Trend + Seasonal Dummies + IMA(1,1)	29.69099
	Linear Trend + Seasonal Dummies + ARIMA(3,1,0)(0,0,1)s	28.58723
	Seasonal Dummies + Quadratic Trend + ARIMA(3,1,0)(0,0,1)	28.41359
	Log Linear Trend + Seasonal Dummies + ARIMA(3,1,0)(0,0,1	27.67981
	Seasonal Dummies + Logarithmic Trend + ARIMA(3,1,0)(0,0,	28.57249
	Log Seasonal Dummies + Logarithmic Trend + ARIMA(3,1,0)(	27.66923
	Seasonal Dummies + Exponential Trend + ARIMA(3,1,0)(0,0,	27.67981
	ARIMA p=(3) d=(1) q=(1, 12) NOINT	57.25025

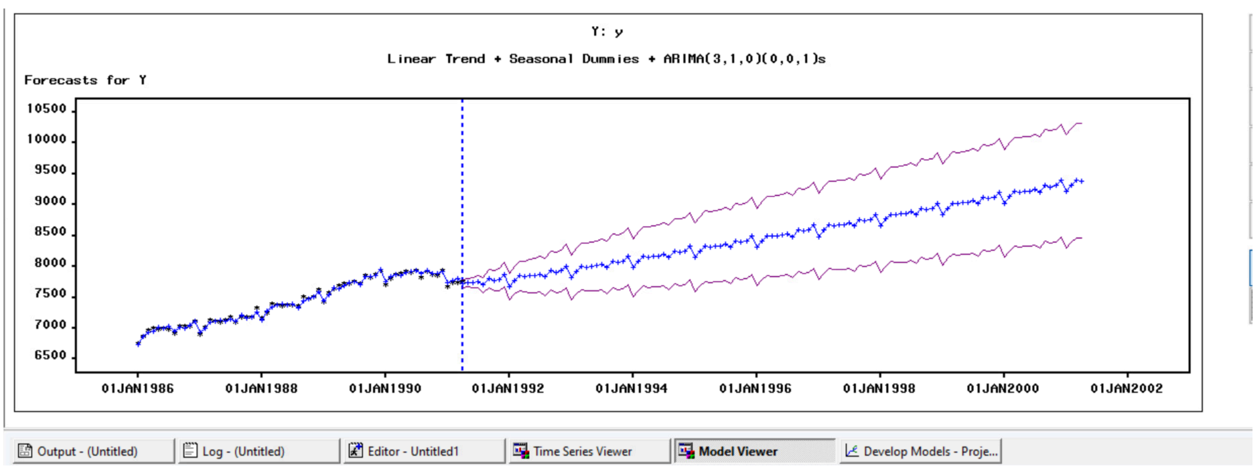
<u>Choosing the model</u>:

The model I would choose that reflects this model is a logarithmic trend with ARIMA p = 3, and Q = 1 with differencing and seasonal dummies. I choose this model because in the context of the dataset, the number of employment have shown a steady increase as time increases. This means that the forecast should be either linear for constant increase or logarithmic for a steady increase. Comparing the model with the same ARIMA and use of differencing and seasonal dummies, these models (exponential, linear, linear with log transform, and logarithmic with log transform) fail the White Noise Test because lag 6 crosses the lower boundary of failing the White Noise test. All the models, except Winter Additive, have seasonal dummy 3 not being significant since it is greater than 0.05. Comparing the logarithmic with the linear trend model is difficult and can be said that both are doable. The model with the logarithmic trend passes the White Noise Test. The intercept has moving average lag 12, and autocorrrelation lag 1 are not significant. Seasonal dummy 3 is also not significant. With the linear trend, the model has a higher RSME (root square error) by around 0.01 and having seasonal dummy 9 not being significant. Since logarithmic trend passes the White Noise test and has a smaller RSME. The Logarithmic Trend is chosen over Linear Trend on log transform because the logarithmic trend on log transform doesn't pass the White Noise Test although it does have a smaller RSME by around 0.9.

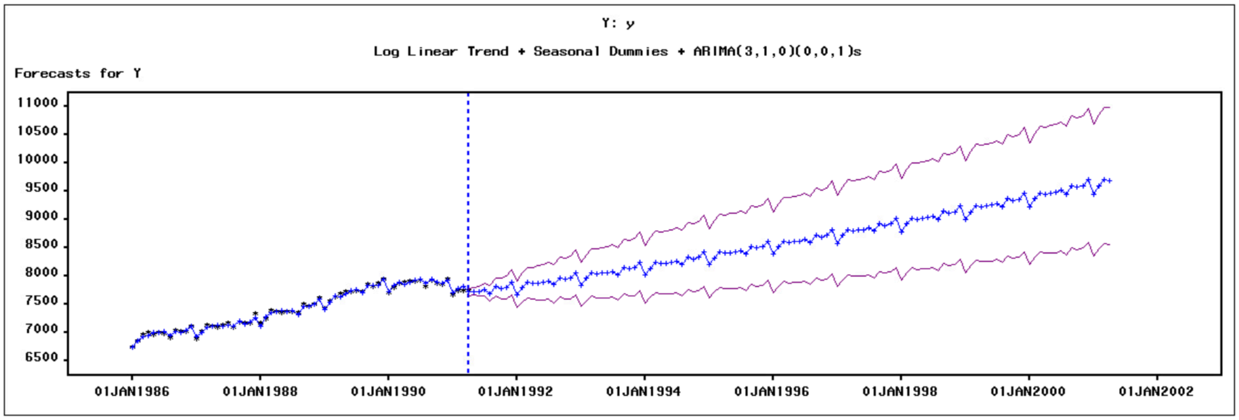
Comparing the logarithmic trend with its log transform, the log transform doesn't also pass the White Noise Test, but has the smaller RSME by around 0.9. The exponential trend when being compared to logarithmic trend, doesn't pass the White Noise Test but has a smaller RSME around 0.9. For the Log Winter's Additive Method, it flawlessly passes the White Noise Test and has a smaller RSME by around 1. However, the seasonal smoothing osn'yt significant. Despite how some of these models can by favored by having a smaller RSME, further elaboration about choosing the Logastic trend will discussed when choosing the forecast model.

Step 3: Choosing the forecast

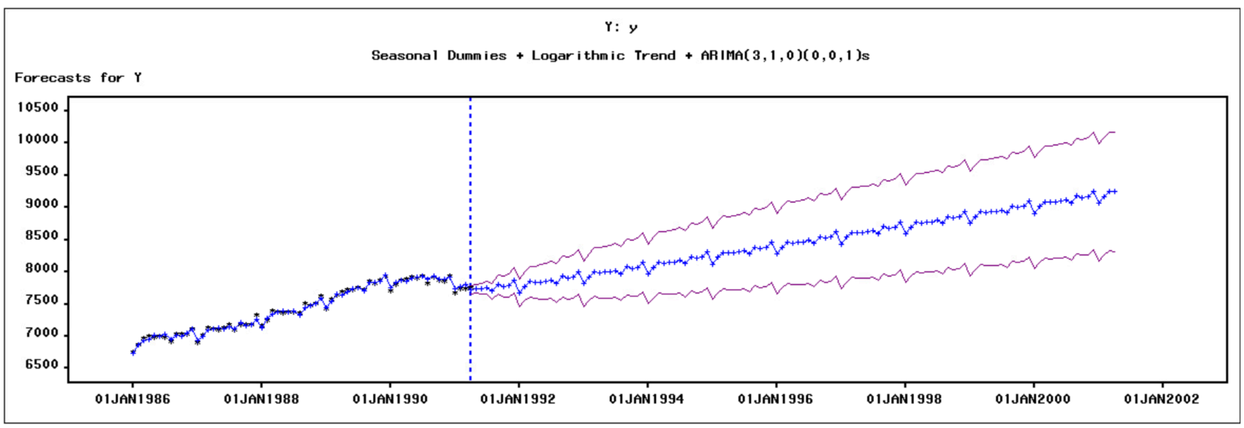
ARIMA model with Linear Trend



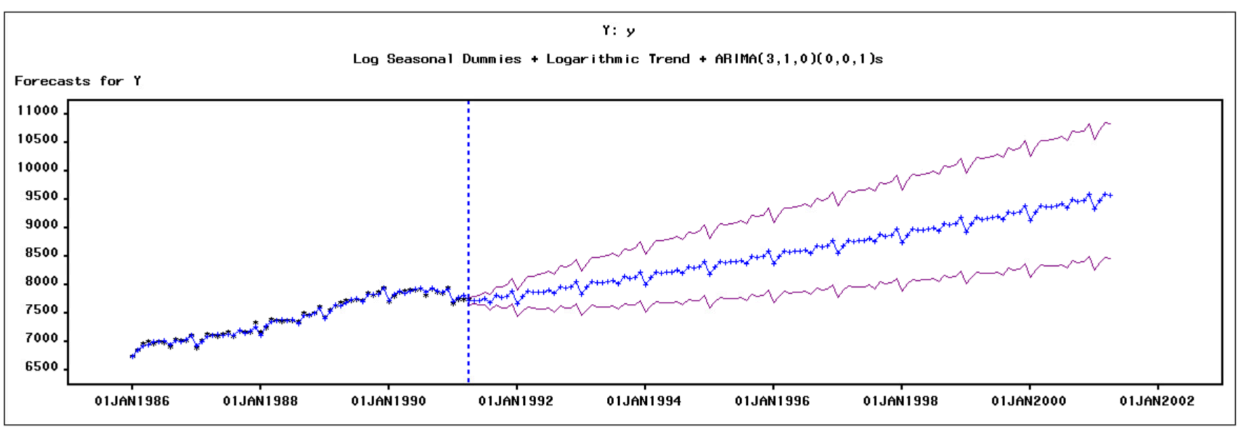
ARIMA model with Linear Trend (Log Transform)



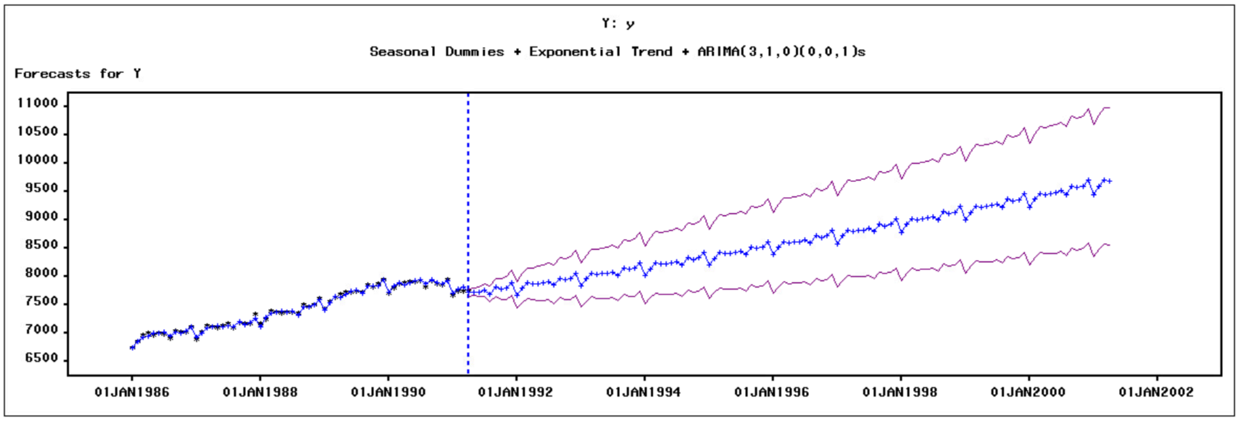
ARIMA model with Logarithmic Trend



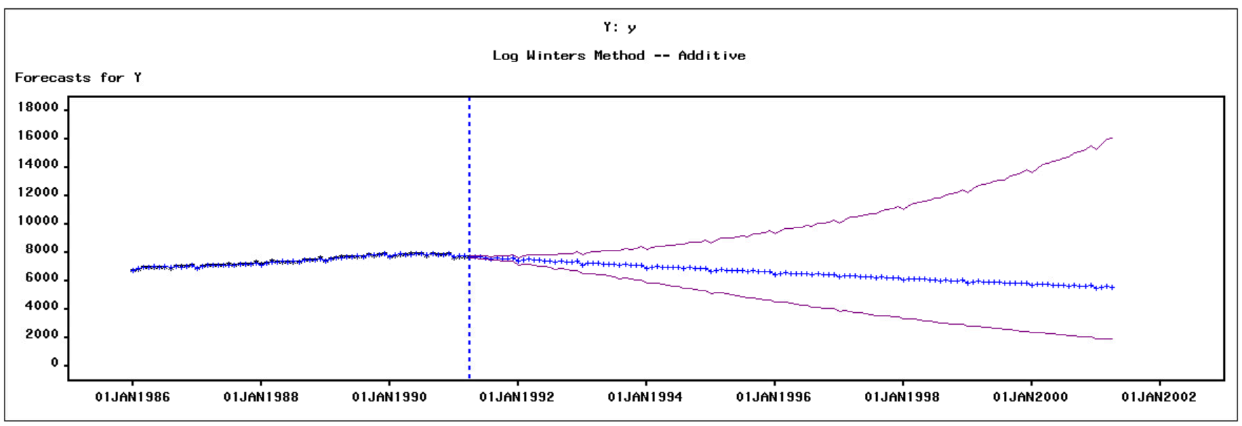
ARIMA model with Logarithmic Trend (Log Transform)



ARIMA model with Exponential Trend



Winter's Additive (Log Transform)



<u>Best Forecast</u>:

When choosing the forecasting model, it is best to try to eliminate what models aren't going to be chosen. The goal is to find the confidence interval that is the most ridgid and narrow. Despite the good qualities of the Log Winter Additive, the forecast shows the confidence intervals exponentially growing causing the interval to be wide. The exponential trend and log transform of linear trend and logarithmic trend will not be chosen because the confidence interval is wider than the linear and logarithmic trend. Furthermore, the log transform will not be used because in the context of the data set, employment shouldn't necessarily increase that fast as more years are forecasted. The difficulty of choosing the forecasting models is choosing between the logarithmic and linear trend. Both have similar confidence intervals and under the same context, plotting the employment over time shows a linear or logarithmic like growth between 1978 to 1991. This means elements discussed in part 2 will have to help decide what forecasting model to choose. As stated from part 2, logarithmic model will be chosen because it passes the White Noise Test and have a smaller RSME.