United States' Unemployment Rate Model

An iterative statistical modeling analysis of United States' unemployment rate data using time series, functional fitting, and linear regression modeling techniques.

Joel Iventosch

Friday, December 14, 2012

Applied Statistical Methods

Dr. Rahn

Final Project

Executive Summary

This project is intended to simulate life in the real world and how the statistical modeling techniques that we've learned in this course can be applied to specific scenarios that arise in business, government, and industry. As such, this report is written as if its intended audience were the decision makers at the hiring firm or organization who is sponsoring this research and statistical modeling analysis (with the exception of this introductory paragraph). As such, this report is written in a very business-oriented manner: clearly broken down by category and subcategory, concise and to the point, and more oriented towards what the models and results mean for the decision maker and his or her organization, rather than what they mean in an academic sense.

Statistical modeling and analysis is a highly iterative process in general, and this particular project is no exception. Three different modeling techniques were used throughout this process, each of which be briefly reported on individually within its own subsection. The first and third modeling techniques used, time series and linear regression respectively, required one or more iterations on their own. These iterations in the time series model will be reported on separately, but the numerous iterations of the linear regression model will not. These iterations will be discussed, but will not be reported on individually; only the final version of the linear regression model will be officially reported on, since the final model is the most accurate. In the concluding section, the three different modeling techniques will be briefly compared and a recommendation of which one to utilize will be provided to the hiring organization.

It is worth noting that this statistical modeling process, like most others, incorporates several fundamental and innate assumptions. These include, but are not limited to, the following: all data used in models is normally distributed; unemployment rate in the United States is ***dependent*** on one or more variable factors; time may or may not be one of these determining factors; the relation between these single or numerous independent factors and time may or may not be linear; the expected value of the random error for each model as a whole is zero.

This report will first present and discuss the results of the first modeling technique: non-seasonal time series. The seasonal time series model will be reported on next, followed by the functional fitting model, and lastly by the linear regression model.

Raw Data

It is worth making note of several different important points regarding the raw data used in this iterative modeling analysis. First, the data was collected was numerous sources, all of which are listed in the back of this report under the “References” section. Each reference includes a live link to the website of the source used.

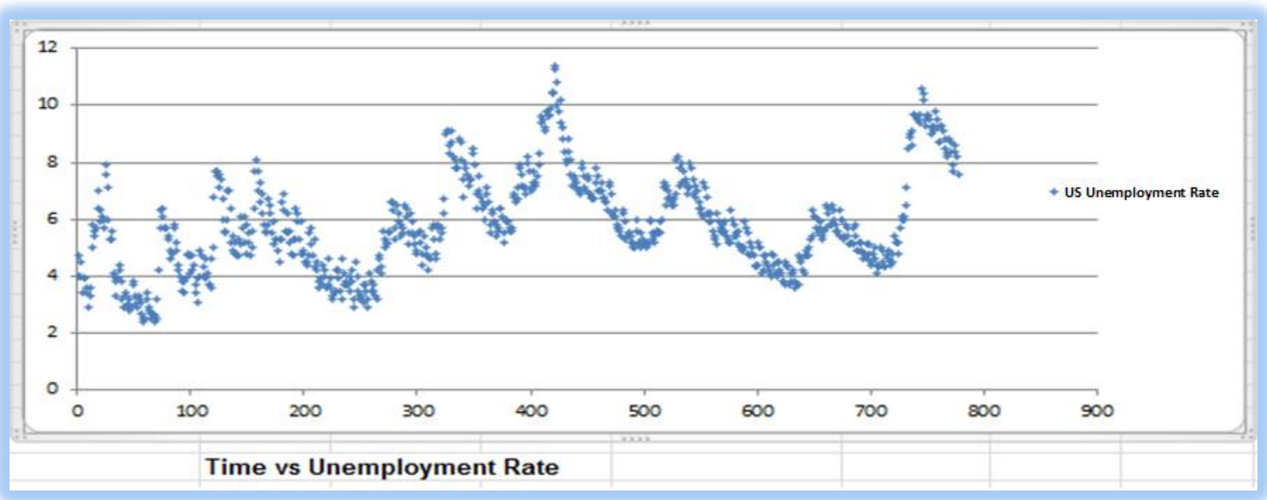
The data used as the dependent variable for this report are unemployment rates in the United States. Data was available on both unemployment ***levels*** – meaning actual numbers of people unemployed, as well as unemployment ***rates***. I elected to use unemployment ***rates***, because this a more constant view of the unemployment story than actual numbers of people. The actual number of unemployed people in the United States will change over time, even if the unemployment rate remains the exact same – due to an increasing population and several other factors. Hence, unemployment rates give a more accurate view into the state of unemployment in the country since they provide an interpretation of the unemployment level, somewhat isolated form the constantly changing population. That said, as will be apparent later in this report, the linear regression model shows that although population does not have the same level of direct effect on unemployment rates as it does numbers, it is still a very heavily weighted factor that does contribute to the model’s calculation of the predicted unemployment ***rate***.

The raw data used in these models is for the time period ranging from January 1948 to the present (and it also includes the annual average unemployment rate for 1947 and is missing the most recent monthly unemployment rate for 2012). This data is non-seasonally adjusted unemployment rates is the United States. There was seasonally-adjusted data available, however I chose not to use this data as it is “adjusted” data and it thus already been run through the source’s own modeling analysis and is already saturated with its own set of implicit assumptions and thus inaccuracies. Every modeling process has to make certain underlying assumptions, and as a result, every model has its own error and bias. The more raw data is “processed” (meaning run through different models) the more it becomes embedded with these assumptions and inherent error. While processed data might be more interpretable – and is certainly useful for decision makers – it is not ideal to use when building you own model since these compounded assumptions and error can distort the results of the model. In light of this, I elected to use non-seasonally adjusted unemployment rate data for all of the models discussed in this report.

The raw data used in this modeling analysis was available in a monthly format and in an annual average format. The time series models and the functional fitting model both use the monthly data, while the linear regression model uses the annual average data. Lastly, there are a number of important assumptions that must be made in any statistical modeling process, and this is no exception. While not always ideal, these assumptions are a reality of the limitations that are imposed on us by the laws of mathematics and science. Assumptions made in this modeling process include, but are not limited to, the following:

- All data used in models is normally distributed
- Unemployment rate in the United States is ***dependent*** on one or more variable factors
- Time may or may not be one of these determining factors
- The relation between these single or numerous independent factors and time may or may not be linear (depending on which modeling technique we are using)
- The expected value of the random error for each model as a whole is zero.

FIGURE 1: SCATTER PLOT OF TIME VS. UNEMPLOYMENT RATE IN THE US



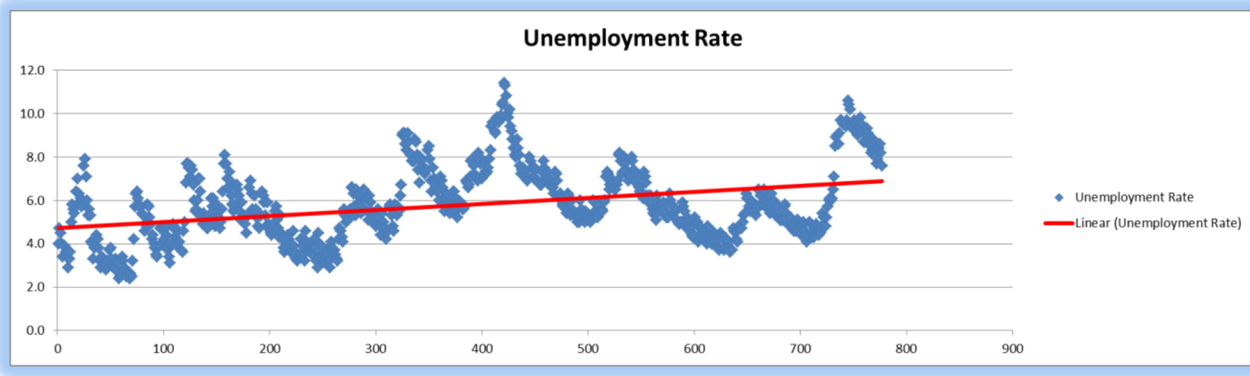
Model 1A: Time Series, non-seasonal

Model:

MODEL 1A: NON-SEASONAL TIME SERIES MODEL OF US UNEMPLOYMENT RATE

$Y_{hat} = 4.7149 + 0.0027838(\boldsymbol{t})$ where $\boldsymbol{t}$=0 for Jan, 1948
$\boldsymbol{t}$=1 for Feb, 1948
$\boldsymbol{t}$=12 for Jan, 1949 etc.

FIGURE 2: GRAPHIC VISUAL OF MODEL 1A



Methodology & Approach:

The modeling technique used to create this model is ***basic time series*** modeling. This modeling technique is essentially the same as a simple linear regression in which time is the independent variable, with a few key exceptions: with a time series all data must be in the correct order, since order has a direction relationship to time, and thus the resulting prediction value of the dependent variable; data must be contiguous, meaning there cannot be any gaps or missing data pairs in the series.

I chose to use this modeling technique because it seemed like a logical starting point, given the data at hand and the relationship between time and unemployment rate that seems to be apparent based on the data spread in Figure 1.

Analysis & Results:

COMPUTATIONAL ACCURACY:

The computational accuracy of this model is very precise. Note that this does not refer to the accuracy of the model as a predictor of unemployment rate (that is discussed in the "Significance and Accuracy of Model" section). Rather, this refers to the level of precision with which the model was calculated – in other words, are the underlying statistics behind the model accurately calculated, or were they estimated? Was rounding used to the extent at which accuracy was lost?

This model was calculated using a regression analysis toolpak in Microsoft Excel. As such, the calculations are very accurate – thus the computational accuracy of the model is very high.

ANALYSIS OF PARAMETERS:

The resulting coefficient of the parameter, ***Time Value***, used in this model is extremely significant in a statistical sense. This is such, due to the extremely small "P-value" of the parameter (see Figure 3). This p-value indicates that the parameter is highly significant (hence the model as a whole is extremely significant since it only has one parameter in this case), because it would pass a significance test at an alpha level of 0.001. In this case, the parameter would actually pass a significance test with an alpha level as small as $1\text{x}10^{-24}$. Thus we can reject the null hypothesis that there is no relation between time value and unemployment rate, and accept the alternative hypothesis that there is a relationship between the two, with only a $1\text{x}10^{-22}$ % probability of making a Type I error. A type I error occurs when the null hypothesis is true, but gets rejected anyways. In this scenario, a Type I error would occur if we assumed that there ***was*** a relation between time and unemployment rate when ***no such relation actually existed***. In simpler terms, the parameter is extremely significant because we can assume that there exists a relation between the parameter, Time Value, and the dependent variable, unemployment rate, with less than a $1\text{x}10^{-22}$ % probability of making a false assumption. In light of this, we conclude that the Time Value parameter is extremely significant in this model.

By looking at the coefficients of each parameter in our model, we can obtain a logical interpretation of what each parameter in our model really means. Since the coefficient of the Time Value parameter is 0.002783808, we can interpret this as follows: with each additional month that passes, the unemployment rate in the US will increase by 0.002783808. Furthermore, the intercept indicates what level the unemployment rate should be (according to our model's predictions) when the Time Value is zero. Recalling that our unemployment was taken from January, 1948 through the present in this model, this means the intercept would indicate an unemployment level of 4.714933518 in January, 1948. The raw data shows that the actual unemployment rate in January, 1948 was 4.0%, indicating that our model's intercept seems reasonable, although perhaps not perfectly accurate. Lastly, notice the extremely small p-value of the intercept (see Figure 3). While this is not always a useful statistic (because often times we will ignore the meaning of a higher p-value for our intercept, since most models do not make sense if the intercept is discarded in the final model), it is a very good indicator of the significance of the intercept when the p-value is this extremely low. Thus we can place an extremely high level of confidence in the fact that there is a relationship between the intercept and the unemployment rate.

FIGURE 3: PARAMETER COEFFICIENTS, P-VALUES, AND INTERPRETATIONS FOR MODEL 1A

	Coefficients	P-value	Interpretation
Intercept	4.714933518	6.23E-196	Intercept is extremeley significant
Time Value	0.002783808	1.09E-25	Time Value perameter is extremely significant

SIGNIFICANCE AND ACCURACY OF MODEL:

As a whole, this model is extremely significant in a statistical sense. We can conclude this based on the p-value of F(model) shown in Figure 4 below. The F-test in statistics tests whether the dependent variable is significantly related to any of the independent variables. In this case, it tests whether the unemployment rate (dependent variable) is significantly related to the Time Value (our only independent variable in this case). The p-value of F(model) that results from a time series regression indicates the significance of the model as a whole, in that it indicates the probability of making a Type I error when we reject the null hypothesis (that our dependent variable is not related to ***any*** of our independent variables) in favor of the alterative hypothesis (that our dependent variable is related to ***at least one*** of our independent variables). As such, we can assume that our model relates Time Value to unemployment rate with only a $1.08967\text{x}10^{-23}$ % chance of making a false assumption. As such, we can conclude that our model is extremely significant.

By looking at the model's R square value (see Figure 4 below), we can get a sense how well our model explains the variations in the raw unemployment data. In other words, our model attempts to answer the question: "why does the unemployment rate change from month to month?" But the question with regard to the overall value of our model is: how well is our model able to account for the variation in the data? The model's R square value provides valuable insight into the answer to this question. The R square statistic is defined as the "explained variation"/"total variation". Thus, it provide a ratio of the variation our model can explain and incorporate into its forecasts, versus the total variation in the data, which includes both the explained variation ***and*** the random variation (or random error) that exists in the data but cannot be explained (hence cannot be forecasted) by our model.

Looking at Figure 4, we see that R square for this model is 0.1323, which means that our model only accounts for 13.23% of the total variation in the raw data. This is an extremely poor R square value, and as such we are forced to conclude that this model does not relate Time Value and unemployment rate in the US very accurately. Notice that this conclusion is made in spite of our model's very high level of overall significance (obtained from the p-value of F(Model)). Thus, our model leads us to conclude that there ***is*** a relation between Time Value and unemployment rate, but that this model is ***not very accurate in quantifying that relationship***.

To properly evaluate the accuracy of this model, three key questions must be answered:

1) **Does the model as a whole make logical sense?**

   Yes. The model makes logical sense in that there does appear to be a significant relation between Time Value and unemployment rate, and this model addresses it by linearly relating the two with a statistically obtained "best fit" line through the data.

2) **Is the model an accurate predictor of the unemployment rate in the US?**

   No. As the R square value makes clear, this model does not accurately predict the unemployment rate in the US. Aside from just using the R square value to make this conclusion, one need only look at Figure 2 to see that the best fit line obtained from this model is, at best, an extremely rough approximation of the raw data. It only makes logical sense that if the model provides very poor rough approximations of the data we have, then it will also provide very poor rough approximations of future unemployment values when it is used to forecast. Generally speaking, good fits provide good forecasts and rough fits provide very rough forecasts.

3) **Does the model answer the most important questions facing the decision maker?**

   No. Not only is this model inaccurate, but it also uses time as the sole predictor of unemployment rates. While this might be a useful tool for the hiring organization to have in some regard, it certainly does not provide the necessary level of information needed to be of great use to the decision maker who is attempting to reduce the unemployment level in the US. Even if this model was an extremely accurate predictor of unemployment rates, the decision maker has no control over time. In fact, nobody does! As such, this is not a very useful model for the decision make, granted it could certainly be a helpful tool in other respects if it were accurate.

Figure 4: R Square Value and P-value of F(model) for Model 1A

Statistic:	Value:	Interpretation:
R Square	0.132324541	Very low correlation between time and unemployment rate
P-val of F(model)	1.08967E-25	Extremely high level of significance for model as a whole

Limitations & Proposed Improvements:

Based on the above section, it is apparent that this model has many limitations. First, it is not an accurate predictor. Secondly, it is not of much use to hiring organization since the decision makers have no control over changing the independent variable, time. Additionally, its linearity guarantees that this model will only be relevant within a certain domain, since the best fit line is a linearly increasing fit. In other words, when the Time Value is extremely large, the model makes no sense. For instance, a Time Value of 36,000 (which would be January of the year 4948) would yield an unemployment rate of 104.93 which is nonsensical since unemployment could never be above 100, by definition. Furthermore, this model is limited in that it does nothing to take into account the seasonality of the data. It includes no smoothing or moving averages, nor does it include any seasonal adjustment factor. Lastly, this model is inherently limited by the fact that it uses only one independent variable, time, as the sole determinant of unemployment rate. This is one of the primary and inseparable downfalls of the time series modeling technique.

To improve this model, I suggest (and will later in this report implement) several changes and adjustments. First, I would incorporate some way to take into account the seasonality of the data and hence provide for a seasonal adjustment factor that is built into the model (see Model 1B). Next, I would suggest using a non-linear time based modeling technique to obtain a better fit (used in Model 2). Lastly, I would try using a completely different modeling technique that is not based on time as the sole independent variable (used in Model 3).

Predictions & Recommendations

Based on the overall futility of this model as a predictor of unemployment rates (as discussed above), there is little value in making predictions and/or recommendations based on this model. This is especially true in light of the inevitability of time and the fact that time is the sole determinant of unemployment rates in this model. Nevertheless, the predicted unemployment rates for selected future time periods are forecasted in the table below:

FIGURE 4B: FORECASTED UNEMPLOYMENT RATES FOR SELECTED FUTURE TIME PERIODS (MODEL 1A)

Date	Time Value	Forecasted Unemployment Rate
Jan, 2014	792	**6.920**
Aug, 2020	871	**7.140**
Nov, 2040	1354	**8.484**
Jan, 2500	6624	**23.155**

Model 1B: Time Series, seasonal

Model:

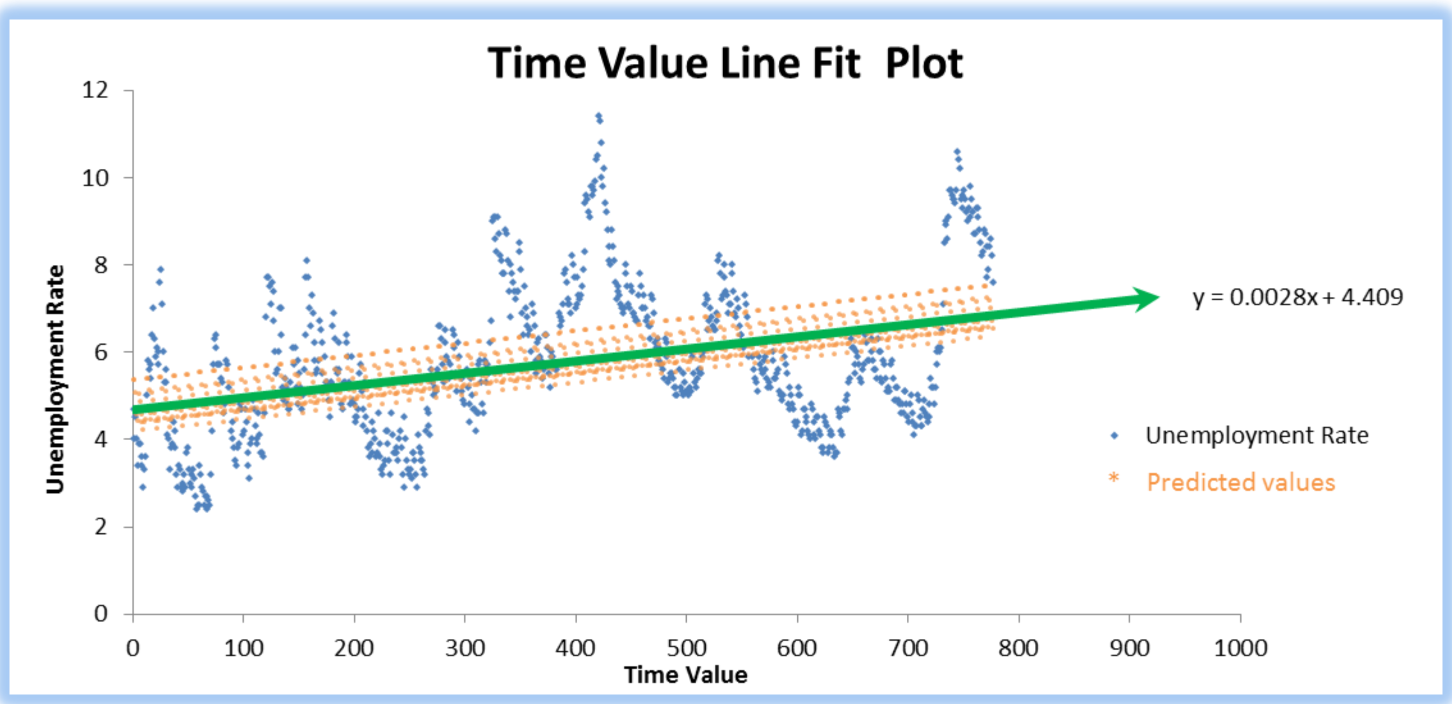
MODEL 1B: SEASONAL TIME SERIES MODEL OF US UNEMPLOYMENT RATE

$$\begin{aligned}Y_{hat} = 4.409 + \; & 0.0028012(\boldsymbol{t}) + 0.9448(\boldsymbol{JAN}) + 0.9435(\boldsymbol{FEB}) + \\ & 0.6653(\boldsymbol{MAR}) + 0.1563(\boldsymbol{APR}) - 0.0065(\boldsymbol{MAY}) + \\ & 0.6077(\boldsymbol{JUN}) + 0.4341(\boldsymbol{JUL}) + 0.1267(\boldsymbol{AUG}) - \\ & 0.0407(\boldsymbol{SEP}) - 0.2303(\boldsymbol{OCT}) - 0.0409(\boldsymbol{NOV})\end{aligned}$$

where ***t***=0 for Jan, 1948
t=1 for Feb, 1948
t=12 for Jan, 1949 etc. &

JAN, FEB, MAR, …, NOV = 1 if ***t*** corresponds to that month &
= 0 if ***t*** does NOT correspond to that month.

FIGURE 5: GRAPHIC VISUAL OF MODEL 1B



Methodology & Approach:

The modeling technique used to create this model is ***seasonally adjusted time series*** modeling. This modeling technique is essentially the same as a basic time series model in which time is the independent variable, with one addition: to add seasonality to the model, we attach n-1 dummy variables to our data set, where n is the number of periods within each full seasonal cycle in our model (i.e. for a monthly seasonal model n-1 = 11). We then set the dummy variables to their appropriate binary values: "1" if the specific value of ***t*** corresponds to that month, and "0" otherwise. This results in a series of $n_x n$ matrices being added to our data set (see Figure 6 below). We then regress the time series over all twelve of our independent variables (Time Value to plus n-1 dummy variables) just as we would with a linear regression.

The resulting parameter coefficients give a coefficient for ***t*** (our slope) plus a coefficient for each of the periods' dummy variables that amount to a constant additive seasonality adjustment. This constant gets added to the trend line output if the current value of ***t*** happens to correspond to that period (since the seasonal adjustment would be multiplied by 1 and then added to the intercept and trend) and all other seasonal coefficients get zeroed out since they will be multiplied by zero in the model, unless their month is current. The nth period gets no seasonal adjustment parameter, because it becomes the standard for all the other periods' coefficients to be based on. Thus, in Model 1B, the trend line provides the model's forecast as if every month were the standard month, December, and then appropriate seasonal adjustment is added or subtracted based on the coefficient of the dummy variable corresponding to which month it actually is. In this manner we are able to use the same basic time series modeling technique with the addition of period dummy variables in order to obtain the desired seasonal time series model.

I chose to use this modeling technique because it seemed like a logical next step in the iterative modeling process.

FIGURE 6: RAW DATA WITH 12 X 12 MATRICES OF DUMMY VARIABLES ADDED ON FOR MODEL 1B

Time Value	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Unemployment Rate
1	1	0	0	0	0	0	0	0	0	0	0	0	4.0
2	0	1	0	0	0	0	0	0	0	0	0	0	4.7
3	0	0	1	0	0	0	0	0	0	0	0	0	4.5
4	0	0	0	1	0	0	0	0	0	0	0	0	4.0
5	0	0	0	0	1	0	0	0	0	0	0	0	3.4
6	0	0	0	0	0	1	0	0	0	0	0	0	3.9
7	0	0	0	0	0	0	1	0	0	0	0	0	3.9
8	0	0	0	0	0	0	0	1	0	0	0	0	3.6
9	0	0	0	0	0	0	0	0	1	0	0	0	3.4
10	0	0	0	0	0	0	0	0	0	1	0	0	2.9
11	0	0	0	0	0	0	0	0	0	0	1	0	3.3
12	0	0	0	0	0	0	0	0	0	0	0	1	3.6
13	1	0	0	0	0	0	0	0	0	0	0	0	5.0
14	0	1	0	0	0	0	0	0	0	0	0	0	5.8
15	0	0	1	0	0	0	0	0	0	0	0	0	5.6
16	0	0	0	1	0	0	0	0	0	0	0	0	5.4
17	0	0	0	0	1	0	0	0	0	0	0	0	5.7
18	0	0	0	0	0	1	0	0	0	0	0	0	6.4
19	0	0	0	0	0	0	1	0	0	0	0	0	7.0
20	0	0	0	0	0	0	0	1	0	0	0	0	6.3
21	0	0	0	0	0	0	0	0	1	0	0	0	5.9
22	0	0	0	0	0	0	0	0	0	1	0	0	6.1
23	0	0	0	0	0	0	0	0	0	0	1	0	5.7
24	0	0	0	0	0	0	0	0	0	0	0	1	6.0

Analysis & Results:

COMPUTATIONAL ACCURACY:

The computational accuracy of this model is very precise. Note that this does not refer to the accuracy of the model as a predictor of unemployment rate (that is discussed in the "Significance and Accuracy of Model" section). Rather, this refers to the level of precision with which the model was calculated. This model was calculated using a regression analysis toolpak in Microsoft Excel. As such, the calculations are very accurate – thus the computational accuracy of the model is very high.

ANALYSIS OF PARAMETERS:

The resulting coefficient of the parameter, ***Time Value***, used in this model is extremely significant in a statistical sense. This is such, due to the extremely small "P-value" of the parameter (see Figure 7). This p-value indicates that the parameter is extremely significant, because it would pass a significance test at an alpha level of 0.001. In this case, the parameter would actually pass a significance test with an alpha level as small as $1x10^{-26}$. Thus we can reject the null hypothesis that there is no relation between time value and unemployment rate, and accept the alternative hypothesis that there is a relationship between the two, with only a $1x10^{-23}$ % probability of making a false assumption. In light of this, we conclude that the Time Value parameter is extremely significant in this model.

Looking at Figure 7 below, we see that only four of our eleven dummy variable parameters have a p-value less than 0.05. This means that only these four parameters would pass a significance test with an alpha value of 0.05, and the other seven would fail. In simpler terms, we can only conclude for these four parameters (and Time Value and the intercept) that there is a significant relationship between the given parameter and the unemployment rate, while maintaining a 5% or less probability of being wrong. This leads to the conclusion that five of the parameters and the intercept are significant (hence important) to our model, while the other seven are not.

By looking at the coefficients of each parameter in our model, we can obtain a logical interpretation of what each parameter in our model really means. Since the coefficient of the Time Value parameter is 0.0028012, we can interpret this as follows: with each additional month that passes, the unemployment rate in the US will increase by 0.0028012. Furthermore, the intercept indicates what level the unemployment rate should be (according to our model's predictions) when the Time Value is zero. Recalling that our unemployment was taken from January, 1948 through the present in this model, this means the intercept would indicate an unemployment level of 4.409 in January, 1948. The raw data shows that the actual unemployment rate in January, 1948 was 4.0%, indicating that our model's intercept seems reasonable, although perhaps not perfectly accurate. Lastly, notice the extremely small p-value of the intercept (see Figure 7). This is a very good indicator of the significance of the intercept when the p-value is this extremely low. Thus we can place an extremely high level of confidence in the fact that there is a relationship between the intercept and the unemployment rate.

Lastly, looking at the coefficient for each months' dummy variable, we can see that a month whose coefficient is ***x*** greater than zero indicates that the unemployment rate in that month is, on average, ***x*** greater than the forecasted rate would be if it were for December (the standard month). Similarly, a dummy variable whose coefficient is negative indicates that the unemployment rate in that month is less on average than it is for an equivalent forecast if it were in December. As such, the dummy variable coefficients can each be interpreted as the seasonal adjustment factor for that particular month each year. This seems to make logical sense since the months leading up to and around the holidays (October, November, and December) have negative or neutral (for December) seasonal adjustments and thus lower unemployment rates based on our model. Common knowledge tells us that this makes sense since companies tend to hire more employees during these months to prepare for and handle the holiday rush. We see from Figure 7 that the seasonal adjustments for January, February, March, and April are much higher, thus signaling higher unemployment in these months, which also makes sense since many businesses reduce their temporary workforce after the holidays when they are no longer needed.

In light of Model 1B's parameter p-values and coefficients, we can conclude that the parameters make sense logically, but are not all significant.

FIGURE 7: PARAMETER COEFFICIENTS, P-VALUES, AND INTERPRETATIONS FOR MODEL 1B

	Coefficients	*P-value*	*Significant at α = .05?*
Intercept	4.40909363	3.76E-73	YES
Time Value	0.002801202	3.76E-27	YES
tJAN	0.944751203	0.000619	YES
tFEB	0.943488463	0.00063	YES
tMARCH	0.665302645	0.015729	YES
tAPR	0.156347597	0.569633	NO
tMAY	-0.00645361	0.981273	NO
tJUNE	0.607668269	0.02734	YES
tJULY	0.434097836	0.114663	NO
tAUG	0.126681249	0.644999	NO
tSEP	-0.04073534	0.882219	NO
tOCT	-0.2303351	0.404093	NO
tNOV	-0.0409488	0.882058	NO

SIGNIFICANCE AND ACCURACY OF MODEL:

As a whole, this model is extremely significant in a statistical sense. We can conclude this based on the p-value of F(model) shown in Figure 8 below. The F-test in statistics tests whether the dependent variable is significantly related to any of the independent variables. In this case, it tests whether the unemployment rate (dependent variable) is significantly related to the Time or any of our seasonal dummy variables. Since the p-value for F(model) is $1.84636X10^{-27}$, we can assume that our model relates at least one of our independent variables to unemployment rate with only a $1.84636x10^{-25}$ % chance of making a false assumption. As such, we can conclude that our model is extremely significant, ***as a whole***. That said, we know from looking at the p-values of each parameter, that the model is not entirely significant. This conflict in the statistics regarding the parameters and the model as a whole is an indicator that we should be very cautious excepting this model – and probably should not accept and/or use it, barring a lack of other viable alternatives.

By looking at the model's R square value (see Figure 8 below), we can get a sense how well our model explains the variations in the raw unemployment data. One key question with regard to the overall value of our model is: how well is our model able to account for the variation in the data? The model's R square value provides important insight to help answer this question. Looking at Figure 8, we see that R square for this model is 0.185027413, which means that our model only accounts for about 18.5% of the total variation in the raw data. Although this R square value is approximately 0.05 better than that of Model 1A, it is still an extremely poor R square value, and as such we are forced to conclude that this model does not relate Time Value (with seasonal adjustment dummy variables) and unemployment rate in the US very accurately. Similar to Model 1A, this conclusion is made in spite of Model 1B's very high level of overall significance (obtained from the p-value of F(model)). Thus, our model leads us to conclude that there ***is*** a relation between Time Value and unemployment rate, but that this model is ***still not very accurate in quantifying that relationship, even with the improvements made from Model 1A*** (namely, the addition of seasonal adjustment).

To properly evaluate the accuracy of this model, three key questions must be answered:

1) **Does the model as a whole make logical sense?**
   Yes. The model makes logical sense in that there does appear to be a significant relation between Time Value and unemployment rate, and also a significant relationship between ***some*** of the monthly dummy variables and unemployment rate. This model addresses it by linearly relating the Time Value and unemployment rate with with a statistically obtained "best fit" line through the data, and then adding or subtracting a seasonal adjustment factor to the forecast resulting from the best fit line. This does make logical sense, especially in light of the fact that the monthly seasonal adjustment factors seem to align with what we would anticipate them to be based on common knowledge.
2) **Is the model an accurate predictor of the unemployment rate in the US?**

No. As the R square value makes clear, this model does not accurately predict the unemployment rate in the US. Aside from just using the R square value to make this conclusion, one need only look at Figure 5 to see that the best fit line obtained from this model is still an extremely rough approximation of the raw data, even after seasonal adjustments are factored in. As such, I conclude that this model is not an accurate predictor of the unemployment rate in the US.

3) **Does the model answer the most important questions facing the decision maker?**

No. Not only is this model inaccurate, but it also uses time (including seasonality, granted) as the sole predictor of unemployment rates. Similar to the conclusion reached regarding the utility of Model 1A, this model might be a useful tool for the hiring organization to have in some regard, but it certainly does not provide the necessary level of information needed to be of great use to the decision maker who is attempting to reduce the unemployment level. Since the decision maker does not have any control over the independent variable, time, this is not a very useful model overall.

FIGURE 8: R SQUARE VALUE AND P-VALUE OF F(MODEL) FOR MODEL 1B

Statistic:	Value:	Interpretation:
R Square	0.185027413	Still very low, but 0.05 better than non-seasonal model
F(Model)	1.84636E-27	Even higher level of significance for the model as a whole

LIMITATIONS & PROPOSED IMPROVEMENTS:

Based on the above section, it is apparent that this model has many limitations. First off, over half of its parameters are insignificant. Secondly, it is not an accurate predictor. Furthermore, it is not of much use to hiring organization since the decision makers have no control over changing the independent variable, time. Additionally, its linearity guarantees that this model will only be relevant within a certain domain, since the best fit line is a linearly increasing fit. In other words, when the Time Value is extremely large, the model makes no sense. Lastly, this model is inherently limited by the fact that it uses only time (and derivations of time: the monthly dummy variables), as the determinant of unemployment rate.

To improve this model, I suggest (and will later in this report implement) several changes and adjustments. First, I would suggest using a non-linear time based modeling technique to obtain a better fit (used in Model 2). Next, I would try using a completely different modeling technique that is not based on time as the sole independent variable (used in Model 3).

PREDICTIONS & RECOMMENDATIONS

Based on the overall futility of this model as a predictor of unemployment rates (as discussed above), there is little value in making predictions and/or recommendations based on this model. Nevertheless, the predicted unemployment rates for selected future time periods are forecasted in the table below to demonstrate the differences and improvements between Model 1A and Model 1B:

FIGURE 9: FORECASTED UNEMPLOYMENT RATES FOR SELECTED FUTURE TIME PERIODS (MODEL 1A VS. MODEL 1B)

Date	Time Value	Forecasted Unemployment Rate	
		Model 1A	*Model 1B*
Jan, 2014	792	**6.920**	**7.572**
Aug, 2020	871	**7.140**	**6.976**
Nov, 2040	1354	**8.484**	**8.161**
Jan, 2500	6624	**23.155**	**23.909**

Model 2: Functional Fitting

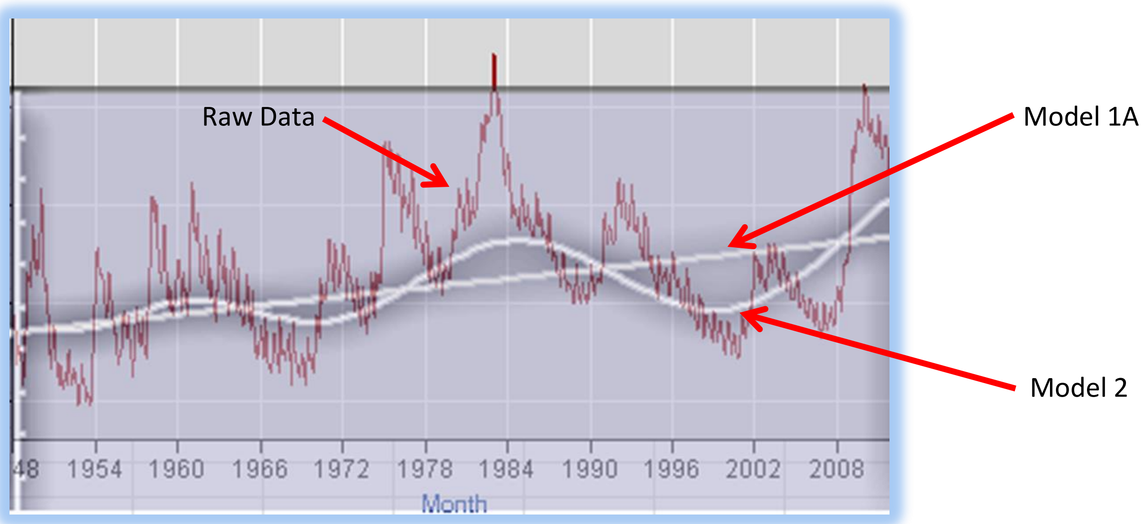
Model:

MODEL 2: FUNCTIONAL FITTING MODEL OF US UNEMPLOYMENT RATE

$$Y_{hat} = (.002\boldsymbol{t})*\cos((\boldsymbol{t}+75)^{\wedge}(1/2.125))+0.0028\boldsymbol{t}+4.409$$

where $\boldsymbol{t}$=0 for Jan, 1948
$\boldsymbol{t}$=1 for Feb, 1948
$\boldsymbol{t}$=12 for Jan, 1949 etc.

FIGURE 10: GRAPHIC VISUAL OF MODEL 1A & MODEL 2

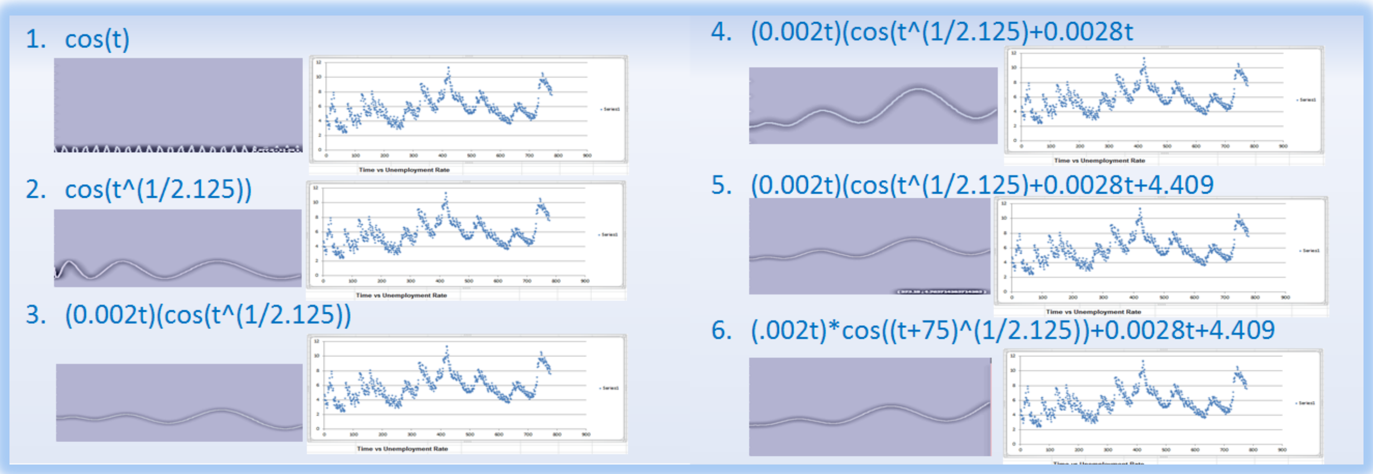


Methodology & Approach:

The modeling technique used to create this model is ***functional fitting***. This is a visual and mathematical modeling technique that involves analyzing a scatter plot of the raw data as a function of time and essentially "fitting" the model function to the data curve. In other words, this method starts with making a best guess as to what mathematical function would best match the graph formed by the raw data, and then making adjustments and improvements to the fit from there. In this sense this modeling technique is extremely iterative in that it requires numerous adjustments and modifications to the model's function. For this model in particular, Model 2, I began with f(x) = cos(t) as my initial model function, and then used my previous knowledge of mathematics and guess and check methodologies to adjust and improve the function until reaching its final version. Although this process involved well over twenty iterations to find the best possible fitting functional model (and undoubtedly a better one yet still exists), the process I used can be approximated by the six iterations shown in Figure 11 below.

I chose to use this modeling technique because it seemed like a non-linear function might be a much more accurate model of unemployment as a function of time than would be any linear function.

FIGURE 11: VISUAL REPRESENTATION OF FUNCTIONAL FITTING MODELING TECHNIQUE USED FOR MODEL 2



Analysis & Results:

COMPUTATIONAL ACCURACY:

The computational accuracy of this model is much lower than that of Models 1A and 1B. While Models 1A and 1B were calculated using a regression analysis toolpak in Microsoft Excel, this model was derived by way of functional fitting, which is inherently a less accurate methodology. This particular usage of functional fitting was particular imprecise because it relied so heavily on graphical visuals and guess and check techniques. As such, the "calculations" involved in this model are approximate in nature. Note, however, this does not mean that this model will not be as, or more, accurate than the others – in fact, it likely will be.

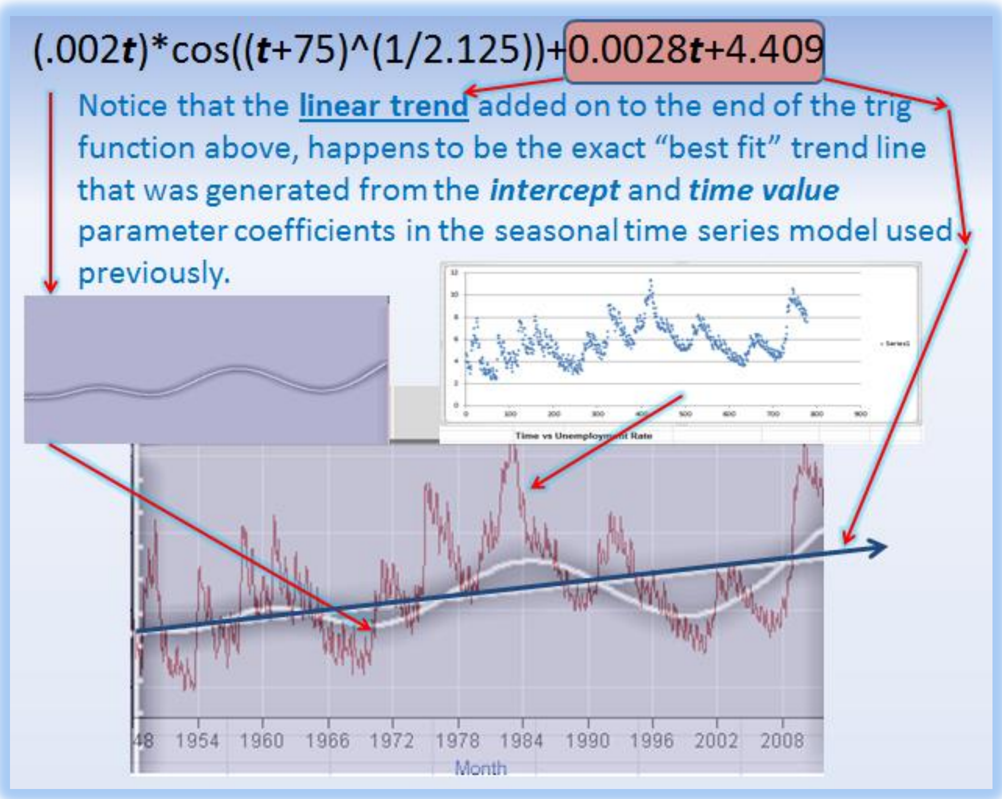
ANALYSIS OF PARAMETERS:

The time value, ***t***, is the only parameter in this type of model. I did calculate significance or correlation statistics for the parameter of this model, because these statistics were not readily accessible like they were for the other three models. Models 1, 1A, and 3 were all calculated using analysis toolpak software in Microsoft Excel, and accordingly most of the pertinent statistics were calculated automatically by the software as a part of the regression output in Excel. Model 2, however, was calculated without the use of software, and as such, its parameter statistics would have to been calculated manually. In light of the fact that Model 2 was not going to be my final model, it seemed futile to spend the time it would have taken to calculated these manually. Accordingly, these statistics on Model 2 are not available for analysis. That said, a simple visual review of the model suggests that the time value parameter is very likely significant, because it certainly does appear the unemployment rate is substantially related to time, based on Figure 10.

Additionally, if one looks closely at the function for Model 2, it is apparent that the last two terms of the function are equivalent to the non-seasonally adjusted function for Model 1B (see Figure 12 below). As such, Model 2 can actually be viewed as linear time series model for the basic trend (equivalent to the non-seasonally adjusted trend of Model 1B), with a very complex non-linear cyclical adjustment that happens to also be a function of time. Under this interpretation 0.0028t+4.409 is the linear trend and (.002t)*cos((t+75)^(1/2.125)) is the additive cyclical adjustment. Note that this is a non-constant cyclical adjustment, as opposed to a constant seasonal adjustment in Model 1B, because the size of the cyclical additive is also a function of time and thus changes (increases in this case) with time. Furthermore, the cycles themselves (measured as distance between the troughs of each consecutive cycle of the trigonometric function along the x-axis) actually increase in length as well, as time moves forward. Since these cycle lengths are non-constant and are much larger than the typical seasonal adjustment (several years vs. one month) they are referred to as "cycles" as opposed to "seasons".

Lastly, since we can separate Model 2 into its linear trend portion and its additive cyclical adjustment portion, we ***can*** comment briefly on the ***t*** parameter for Model 2, based on what we know about ***t*** in Model 1B. By reviewing the results for the parameter ***t*** in Model 1B (recall that ***t*** was extremely significant, statistically speaking), we can conclude that for ***at least*** the linear trend of Model 2, the parameter ***t*** is extremely significant and also seems to make sense based on looking at the raw data scatter plot. In light of this – and the fact that the cyclical additive in Model 2 only appears (visually) to improve the fit of the model – we can conclude the time value parameter in Model 2 is probably very significant in relating unemployment rate data and time.

FIGURE 12: COMPONENT ANALYSIS OF MODEL 2



SIGNIFICANCE AND ACCURACY OF MODEL:

Although this model also lack specific statistics for the p-value of F(model) and the R square value, we can gain much insight by way of visual analysis of the model's graphic representation. To start, let's discuss the significance of the model as a whole. Since we concluded that ***t*** is very likely a significant parameter (in the previous section), we can conclude that this model is likely very significant as a whole, since ***t*** is the sole parameter in the model.

With regards to the accuracy of the model and the lack of an R square value, we can turn to graphical analysis. Recall that the R square value measure the ratio of "explained variation" compared to "total variation" in the raw data. This essentially amounts to the "goodness of fit" when translated into visual terms. As such, it is apparent from Figures 10 and 12 that the quality of fit is much better for Model 2 than it is for Model 1A or 1B. In addition to this visual observation, notice that at most of the values of ***t*** that you might randomly select on the graph, the residual (difference between the value of the actual data point and the Yhat prediction for the same ***t***) for Model 2 is less than that of Model 1A or 1B. Although this is not true for *all* of the possible values of ***t*** (the actual data has certain "sub-cycles" that Model 2 does not account for, which results in certain the Yhat for Model 1A or 1B being more accurate at certain times than the Yhat for Model 2 – quite incidentally, by the way), it is true for the vast majority of possible ***t*** values. As such, we can confidently (although not entirely) conclude that the R square value for Model 2 is significantly better (higher) than that of Model 1A or 1B. Accordingly, our model leads us to conclude that there ***is*** a relation between Time Value and unemployment rate, and that this model is ***significantly more accurate than Model 1A or 1B in quantifying that relationship***.

To fully evaluate the accuracy of this model, three key questions must be answered:

1) Does the model as a whole make logical sense?

Yes. The model makes logical sense in that there does appear to be a significant relation between Time Value and unemployment rate, and this model addresses it by non-linearly relating the two with a visually obtained "best fit" trigonometric function that passes through the data.

2) Is the model an accurate predictor of the unemployment rate in the US?

It is a ***better*** *predictor than Model 1A or 1B.* As the visuals in Figures 10 and 12 make clear, this model is a much more accurate predictor of the unemployment rate in the US than is either of the previous models. Is it a perfect predictor, or even a great predictor? Probably not. But, statistical modeling is an imperfect process and is often equal parts art and science (this is especially evident in Model 2). In light of this, I would conclude that Model 2 is a decent, if not better, predictor of unemployment rates in the US within its relevant range.

3) Does the model answer the most important questions facing the decision maker?

Not entirely. Although this model is more accurate, it still does not provide the necessary level of information needed to be of great use to the decision maker who is attempting to reduce the unemployment level in the US. This is so, because time is still the sole determinant of unemployment rates in this model. And the decision maker still has no control over time. Ergo, this is still not a very useful model for the decision make, granted it could certainly be a helpful tool in other respects since it is far more accurate.

LIMITATIONS & PROPOSED IMPROVEMENTS:

Based on the above section, it is apparent that this model still has some important limitations. First, it is very difficult to determine many of the statistics that inform us of the quality of this model.

Secondly, it is not of much use to hiring organization since the decision makers have no control over changing the independent variable, time. Additionally, its increasing linear base trend and the increasing nature of its additive element guarantees that this model will only be relevant within a certain domain, since the results of forecasting will eventually become infinitely large. In other words, when the Time Value is extremely large, the model makes no sense. Furthermore, this model is limited in that it does nothing to take into account the seasonality of the data. While it does include a longer term cyclic factor, it does not account for the shorter term monthly seasonality like Model 1B does.

To improve this model, I suggest (although I do not implement in this report) several changes and adjustments. First, I would manually calculate the R square value and the p-value for F(model). Next, I would incorporate a constant additive seasonal factor. In fact, one might be able to simply use the same dummy variable coefficients from Model 1B and incorporate them directly into this model. I would then manually recalculate the R square value and the p-value of F(model) statistics again to see if they have improved. Lastly, if necessary, I would tweak the additive cyclical portion of the model (the trigonometric portion) to see if "better fit" functions existed.

Predictions & Recommendations

Based on the fact that this model is based strictly on time, I will make no recommendations to the decision maker. However, I have made some predictions. The predicted unemployment rates for selected future time periods are forecasted in Figure 13 below:

Figure 13: Forecasted Unemployment Rates for Selected Future Time Periods (Model 1A, 1B, & 2)

Date	Time Value	Forecasted Unemployment Rate		
		Model 1A	Model 1B	Model 2
Dec, 2012	779	6.884	6.591	7.196
Jun, 2013	785	6.900	7.216	7.330
Dec, 2013	791	6.917	6.625	7.460
Jan, 2014	792	6.920	7.572	7.482
Dec, 2014	803	6.950	6.658	7.709
Dec, 2015	815	6.984	6.692	7.936
Dec, 2016	827	7.017	6.726	8.137
Aug, 2020	871	7.140	6.976	8.590
Nov, 2040	1354	8.484	8.161	9.912
Jan, 2500	6624	23.155	23.909	7.179

Model 3: Linear Regression

Model:

MODEL 3: LINEAR REGRESSION MODEL OF US UNEMPLOYMENT RATE

$$\begin{aligned} Y_{hat} = \quad 7.1534 \quad & - 0.2199(\textbf{\textit{GDP Growth}}) \\ & + 0.0655(\textbf{\textit{Top Marginal Individual Income Tax Rate}}) \\ & -0.1123(\textbf{\textit{Max Long-Term Capital Gains Tax Rate}}) \\ & - 0.1811(\textbf{\textit{Top Marginal Corporate Income Tax Rate}}) \\ & + 0.04169(\textbf{\textit{Spending on Major Tax Credits}}) \\ & + 0.6028(\textbf{\textit{Total Deductions as \% of Income}}) \\ & + 0.3572(\textbf{\textit{Gov't Expenditures as \% of GDP}}) \\ & - 0.0053\ (\textbf{\textit{Defense Spending}}) \\ & - 0.1631\ (\textbf{\textit{Total Annual Gov't Surplus}}) \\ & + 0.1270(\textbf{\textit{Federal Funds Effective Rate}}) \\ & - 0.2629(\textbf{\textit{Inflation Rate}}) \\ & - 0.0742(\textbf{\textit{US Population}}) \end{aligned}$$

Methodology & Approach:

The modeling technique used to create this model is ***multiple linear regression*** modeling. This modeling technique is essentially the same as a simple linear regression, except that it incorporates multiple independent variables instead of just one. This particular model has twelve independent variables (see Model 3 above).

I chose to use this modeling technique because it seemed like a logical alternative to time-series-based modeling, given that there are many factors other than time that likely go into determining the unemployment rate in the US.

Analysis & Results:

COMPUTATIONAL ACCURACY:

The computational accuracy of this model is very precise. Note that this does not refer to the accuracy of the model as a predictor of unemployment rate (that is discussed in the "Significance and Accuracy of Model" section). This model was calculated using a regression analysis toolpak in Microsoft Excel. As such, the calculations are very accurate – thus the computational accuracy of the model is very high.

ANALYSIS OF PARAMETERS:

All twelve of the parameters used in this model are significant in a statistical sense. Some are extremely significant, others are very significant, and a few are just barely significant (based on their corresponding p-values…see Figure 14 below). The fact that all twelve independent variables are significant indicates that this is a very strong and significant model.

By looking at the coefficients of each parameter in our model, we can obtain a logical interpretation of what each parameter in our model really means. Since unemployment is an inverse concept by nature (meaning higher unemployment is bad and lower unemployment is good), these parameter coefficients must be interpreted with this in mind. For instance, higher GDP Growth is "good" for the economy, thus we would expect its coefficient to be negative so that as GDP Growth increases, ***un***employment decreases; a "good" for a "good". In other words, these parameter coefficients should be interpreted as follows: a negative coefficient indicates that an increase in the parameter value causes a decrease in unemployment, while a positive coefficient indicates that an increase in parameter value causes an increase in unemployment. In light of this, we can see that the two parameters that seem to have the strongest positive affect on unemployment (i.e. reducing it) are: Inflation Rate and GDP Growth. So, as both of these parameter input values increase, unemployment decreases significantly. The two parameters that seem to have the strongest negative affect on unemployment (i.e. increasing it) are: Total [tax] Deductions as a % of Income and Gov't Expenditures as a % of GDP. Thus, as both of these parameter input values increase, unemployment also increases significantly.

These twelve parameter coefficients can be interpreted in many different ways, based on the interpreter's pre-existing frame of reference, previous experiences, and knowledge of economics and politics. As such, I will leave interpretation of the rest of the parameters (both of their p-values and their coefficients) up to the reader, based on Figure 14 below:

FIGURE 14: COEFFICIENTS, P-VALUES, AND P-VALUE INTERPRETATIONS FOR MODEL 3 PARAMETERS

	Run 7			
X#	***Parameter Names***	***Coefficients***	***P-value***	***Interpretation***
	Intercept	7.153388248	0.094411686	*significant*
1	**GDP Growth**	-0.219916088	0.000729734	*extremely significant*
2	**Top Marginal Income Tax Rate for Individuals**	0.065473975	0.007559184	*very significant*
3	**Max Long-Term Capital Gains Tax Rate**	-0.112303534	0.001535107	*very significant*
4	**Top Marginal Corporate Income Tax Rate**	-0.181142113	0.004710828	*very significant*
5	**Spending on Major Tax Credits***	0.041686849	0.00910521	*very significant*
6	**Total Deductions as % of Income**	0.60277657	1.87817E-08	*extremely significant*
7	**Gov't Expenditures as % of GDP**	0.357185572	0.004281115	*very significant*
8	**Defense Spending**	-0.005328174	0.008544622	*very significant*
9	**Total Annual Gov't Surplus as % of GDP**	-0.163105562	0.083207808	*significant*
10	**Federal Funds Effective Rate**	0.126964046	0.079638483	*significant*
11	**Inflation Rate**	-0.262912229	0.00233198	*very significant*
12	**US Population**	-0.074183847	0.001219095	very significant

SIGNIFICANCE AND ACCURACY OF MODEL:

As a whole, this model is extremely significant in a statistical sense. We can conclude this based on the p-value of F(Model), 1.26467×10^{-14}, shown in Figure 15 below. The F-test in statistics tests whether the dependent variable is significantly related to any of the independent variables. In this case, it tests whether the unemployment rate (dependent variable) is significantly related to any of our twelve independent variables. Since each of our independent variables is statistically significant on its own (as discussed in the previous section), it makes sense that the model as a whole would also be statistically significant – and the p-value of F(model) demonstrates that it is. As such, we can assume that our model relates at least one of our independent variables to unemployment rate with only a 1.26467×10^{-11} % chance of making a false assumption. Hence, we can conclude that our model is extremely significant.

By looking at the model's R square value (see Figure 15 below), we can get a sense how well our model explains the variations in the raw unemployment data. Looking at Figure 15, we see that R square for this model is 0.867338, which means that our model accounts for 86.73% of the total variation in the raw data. This is a fairly good R square value, especially in light of how many different independent variables we have used in this regression. Thus, we can conclude that this model does relate our twelve independent variables and unemployment rate fairly accurately.

To properly evaluate the accuracy of this model, three key questions must be answered:

1) **Does the model as a whole make logical sense?**

   Yes. The model makes logical sense in that most (albeit not all) of the parameter coefficients are positively or negatively correlated to the unemployment rate output of our model in a similar direction and magnitude as we would expect them to be based on common knowledge and real life experiences.

2) **Is the model an accurate predictor of the unemployment rate in the US?**

   Yes. Although this is not a perfect predictor and could undoubtedly use improvements, it is definitely a fairly accurate predictor of unemployment rates in the US. The R square value and the p-value of F(model) strongly support this claim.

3) **Does the model answer the most important questions facing the decision maker?**

   Yes. Because this multiple regression models unemployment rate as a function of twelve different and pertinent independent variables (as opposed to just a function of time), it gives the decision maker the necessary information needed to formulate specific public policy aimed at improving the unemployment rate in the US. As mentioned in the previous section, how the information is interpreted is largely up to the interpreter – but this model ***does*** give the decision maker the necessary information and statistics to properly interpret the data and come to his/her own conclusions based on the model (Figure 14 should be one of the primary components of Model 3 that they analyze as they interpret the model and come to their own conclusions).

FIGURE 15: R SQUARE VALUE AND P-VALUE OF F(MODEL) FOR MODEL 3

Statistic:	Value:	Interpretation:
R Square	0.867338	Good level of correlation; improved significantly
F(Model)	1.26467E-14	Very high level of significance; declined slightly though

LIMITATIONS & PROPOSED IMPROVEMENTS:

Based on the above section, it is apparent that this is a strong and fairly accurate model. That said, it assuredly still has some limitations. First, it is still a linear model, and is thus inherently limited since many functions of science and society are not easily related by simple linearity. As such, this model would very likely benefit by making some of its parameters quadratic or exponential in terms of their input values, or by adding some interaction effects. Secondly, its R square, although fairly good, could certainly use improvement (to get at least into the low to mid .90s). Lastly, the population parameter has a stronger effect on the model than is probably necessary or accurate.

To improve this model, I suggest (although do not implement in this report) several changes and adjustments. First, I would incorporate several interaction effects into the model. This could be a particularly good improvement, especially since many of the independent variables seem to be inherently related on their own. For instance, ***GDP Growth and Gov't Expenditures as a % of GDP*** are certainly related on their own – in fact by its very definition ***Gov't Expenditures as a % of GDP*** is related to ***GDP Growth***. Next, I would convert all dollar amounts to current-value Purchasing Parity Power (PPP) dollar amounts and convert all % parameters to percentages based on ratios using current-value PPP dollars, where applicable. This would ensure that all of the parameter data is in the same "units" before being regressed upon. Additionally, I would replace ***Population*** with ***Population Growth Rate***. I think this might give less weight to the population factor (since the parameter input values would be much smaller…% number rather than a population-type number), which could improve the model. If that doesn't improve the model, then I would try removing the population variable all together and then rerunning the model. In general, I would use this iterative trial and error process of tweaking, rerunning, and then reassessing the model over and over until I had an even better model – and one that I was completely comfortable with and confident in.

PREDICTIONS & RECOMMENDATIONS

Based on the above discussion of this model I would make the following recommendations to the decision maker to improve the unemployment rate in the US:

- Focus on Economic Growth (GDP Growth)
- Lower the top marginal individual income tax rate and income tax rates across the board
- Reduce government spending on tax credits
- Limit the amount and type of tax deductions available

- MUST reduce annual gov't deficits
- If you have to raise taxes, raise capital gains rates and/or corporate income rates (I don't personally agree with this, but that's what the model says!)
- Keep an eye on interest rates and inflation; increasing federal funds rates (i.e. interest rates) and increasing inflation might be good for unemployment rates, but it is bad for the longer term forecast

Lastly, I have included two tables below showing possible predictions based on this model. The first (Figure 16) shows approximately the current-day, real-life values of the parameters as inputs, and the resulting unemployment rate forecast as the output, based on this model. As you can see, the model predicts an unemployment rate of 8.81%. The current actual unemployment rate in the US is approximately 8.0%, which is well within the margin of error, especially given that the input values were only a "best guess" approximation of what the actual current-day values are. In light of this, I feel very confident about the quality and accuracy of Model 3.

The second table (Figure 17) shows a predicted unemployment rate based on a set of possible input values for each parameter. These input values in Figure 17 are roughly the levels that I would anticipate over the next four years, based on the Obama administration's track record over the past four years and other pertinent economic, political, social, and international current events. As such, I predict the unemployment rate in the US to rise to 8.5% within the next calendar year, and then to remain at or above 8.5% for the three years following that.

FIGURE 16: MODELED UNEMPLOYMENT RATE BASED ON CURRENT-VALUE PARAMETER INPUTS (MODEL 3)

MODEL:		
Variable:	**Variable Name:**	**Input Value:**
X1	GDP Growth	1.8
X2	Top Marginal Income Tax Rate for Individuals	35
X3	Max Long-Term Capital Gains Tax Rate	15
X4	Top Marginal Corporate Income Tax Rate	35
X5	Spending on Major Tax Credits*	125
X6	Total Deductions as % of Income	25
X7	Gov't Expenditures as % of GDP	39
X8	Defense Spending	650
X9	Total Annual Gov't Surplus as % of GDP	-5
X10	Federal Funds Effective Rate	0.1
X11	Inflation Rate	3
X12	US Population	310
Yhat	**Forecasted UNEMPLOYMENT RATE**	**8.814290711**

FIGURE 17: PREDICTED UNEMPLOYMENT RATE BASED ON ANTICIPATED FUTURE-VALUE PARAMETER INPUTS (MODEL 3)

MODEL:		
Variable:	**Variable Name:**	**Input Value:**
X1	GDP Growth	1.5
X2	Top Marginal Income Tax Rate for Individuals	45
X3	Max Long-Term Capital Gains Tax Rate	35
X4	Top Marginal Corporate Income Tax Rate	35
X5	Spending on Major Tax Credits*	143
X6	Total Deductions as % of Income	28
X7	Gov't Expenditures as % of GDP	40
X8	Defense Spending	600
X9	Total Annual Gov't Surplus as % of GDP	-7
X10	Federal Funds Effective Rate	0.1
X11	Inflation Rate	5
X12	US Population	315
Yhat	**Forecasted UNEMPLOYMENT RATE**	**9.900689328**

Conclusions

Based on the above analyses of each of the four models, I have come to the conclusion that Model 3 is by far the best model. It is extremely significant, has the best R square value by a long shot, and is the most useful to the decision maker since it uses applicable independent variables as parameters in the model. Most of Model 3's parameters are things that the decision maker can control – or at least influence – unlike time, which he/she cannot control. Although Model 3 is a quality model and a good predictor of unemployment rate, it still has its limitations and biases just like any other statistical model. As such, I would recommend that the decision maker keep these limitations and biases in mind as he/she uses Model 3 and my accompanying predictions and recommendations to come to his/her decisions.

References

http://en.wikipedia.org/wiki/Income_tax_in_the_United_States

http://www.taxpolicycenter.org/index.cfm

http://inflationdata.com/inflation/Inflation_Rate/HistoricalInflation.aspx

ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt

http://www.federalreserve.gov/releases/h15/data.htm

http://www.usgovernmentspending.com/download_multi_year_1910_2017USm_13s1li001mcn_30t

http://www.bea.gov/national/index.htm#gdp

http://en.wikipedia.org/wiki/Historical_GDP_of_the_People's_Republic_of_China

http://stateimpact.npr.org/idaho/2011/11/16/calculating-idaho-unemployment-insurance-benefits/