An Actuary's Guide to Financial Applications: Examples with EViews
By William Bourgeois

An actuary is a business professional who uses statistics to determine and analyze risks for companies. In this guide, we will discuss how an actuary would calculate and analyze various summary statistics, graphs, and other data within the program Econometric Views (EViews). EViews is a statistical package used primarily for time series, and time series related econometric analysis.

After downloading the program and opening it up, the first step to take is to create a new EViews workfile which is shown below.
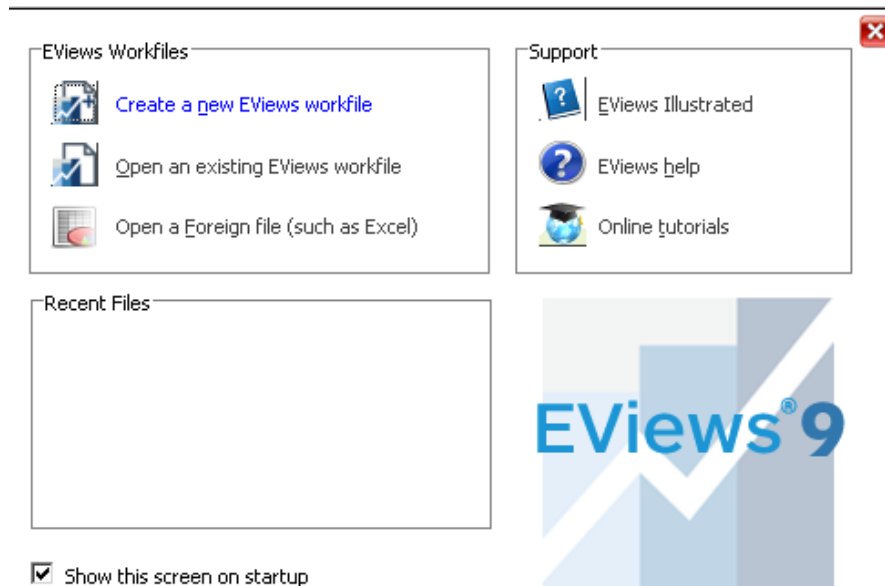
Once this is done, you now need to determine what kind of data you are trying to analyze. For our example, we will be analyzing regular monthly data starting on 5/1/1953 up until 2/1/2017.
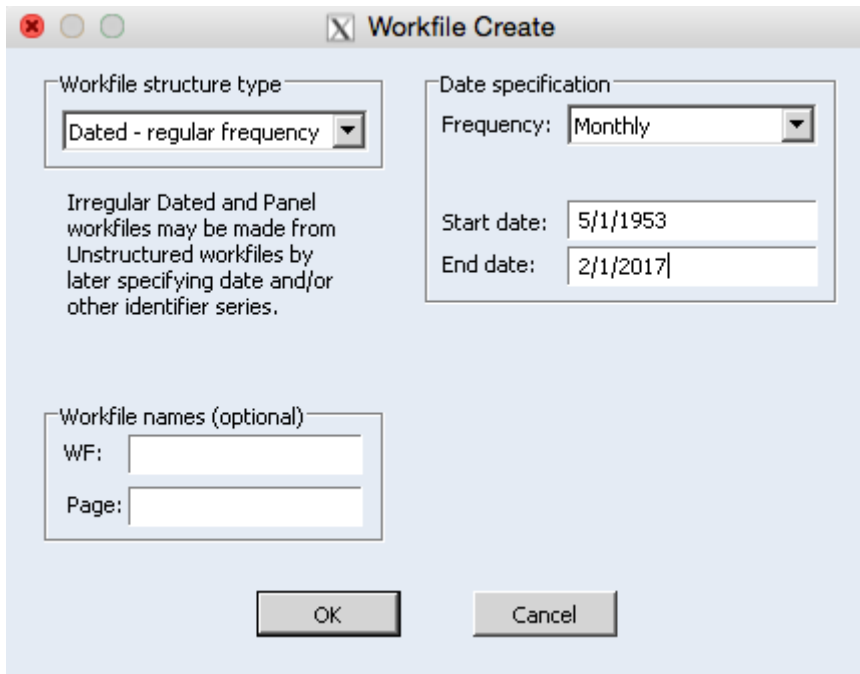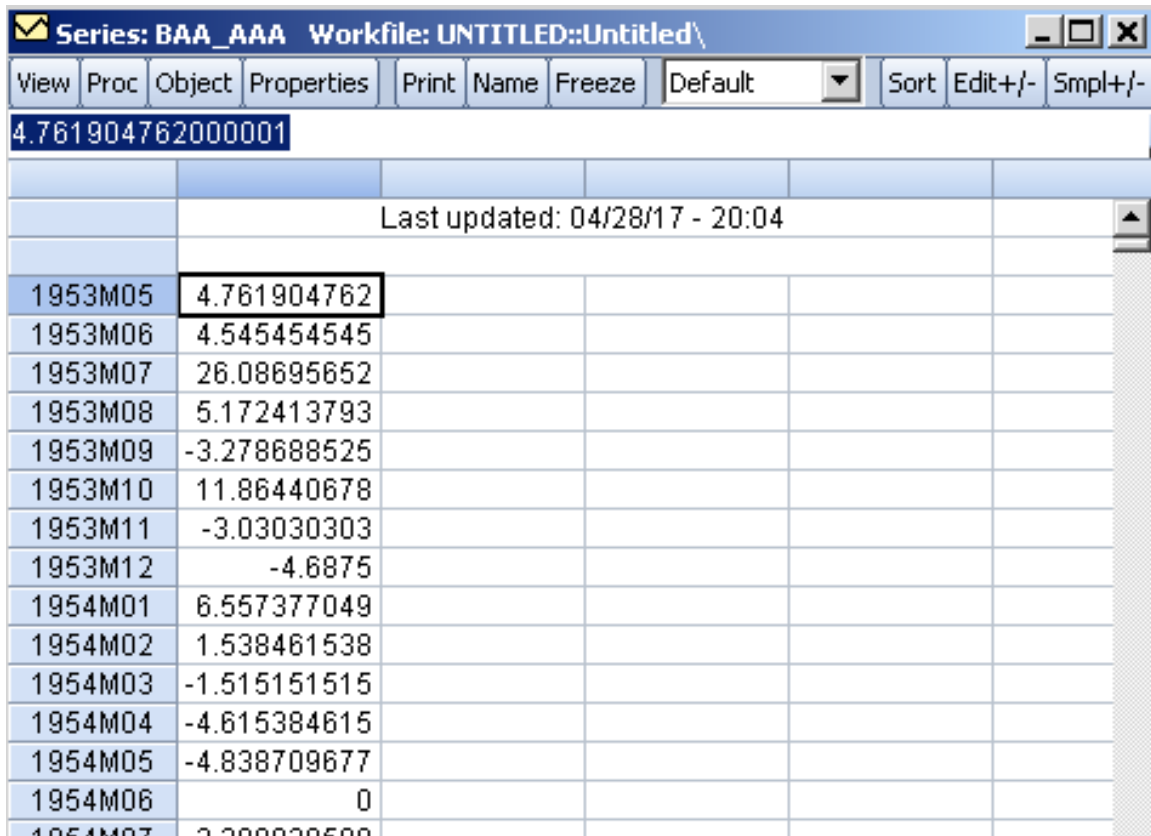
After you determine the date span of your data, you must create new objects which will be the variables that you would like to analyze. The variables in this example consist of the unemployment level (thousands of persons), consumer price index for all urban consumers and all items, 10-year treasury constant maturity rate (percent), Moody's seasoned Baa corporate bond yield percent, Moody's seasoned Aaa corporate bond yield percent, and finally the CRSP market index. Below we show how to create an object for our Baa-Aaa variable that we will be analyzing. All of the other objects can easily be created using the same method. After clicking on Object > New Object… the screenshot below shows what you must now do for the Baa – Aaa variable that we wish to analyze.
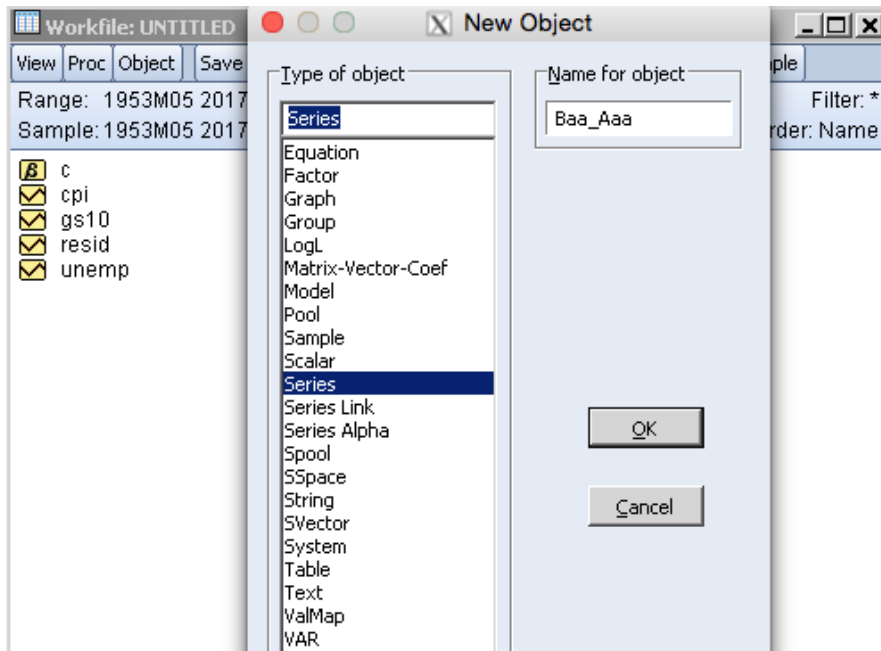
Now that we have created this new object, we must now fill it in with its
corresponding values. We have already established that the variables that we are
analyzing have data running from 5/1/1953 to 2/1/2017. We can obtain monthly
data for Aaa and Baa rated bonds online, and download it in the form of an excel
spreadsheet. We can also do the same for the other variables. Once we have the
excel spreadsheet containing the monthly data for our variables running from our
start and end dates, we can simply copy and paste this data from excel into EViews.
First you must click on the variable in your workfile that you wish to paste the
values into. Once you click on the variable, a window like the one below will appear
and you can simply click on the cell to the right of the first date and paste the
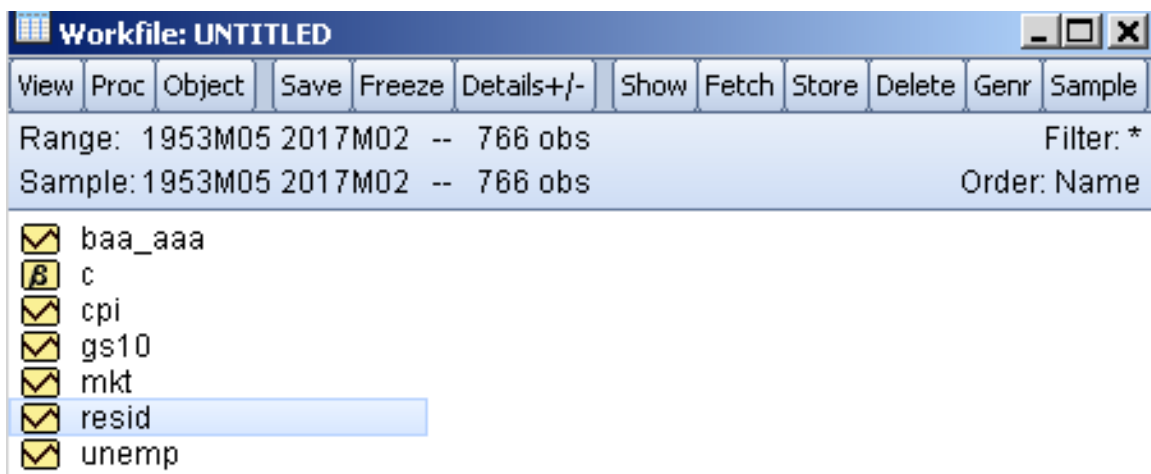
column of data.

This process must be repeated for all of the variables, and once this is done,

we can then move on to the analytical portion of this guide. The workfile for our

example should look like the image below once all the variables have been created,

and each variable in this workfile should have pasted data values within them.
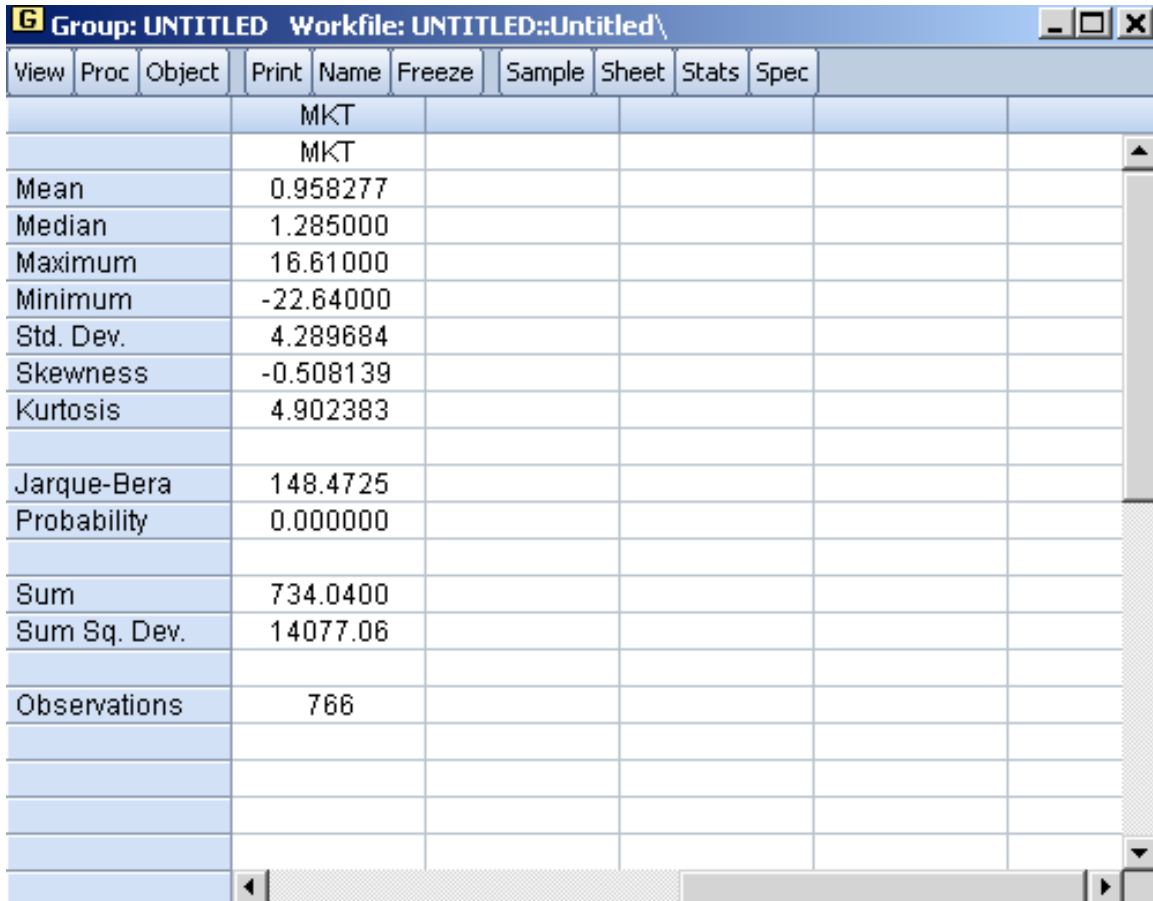
Now that our workfile contains all of our desired variables, all of which are filled out, we can begin to analyze the key characteristics of these variables. Suppose that we want to observe the descriptive statistics of one of our variables, for example, our CRSP Market Index variable. First, select the variable so that it highlights in blue by clicking on it once. After that, click on Quick > Group Statistics > Descriptive Statistics > Common Sample.
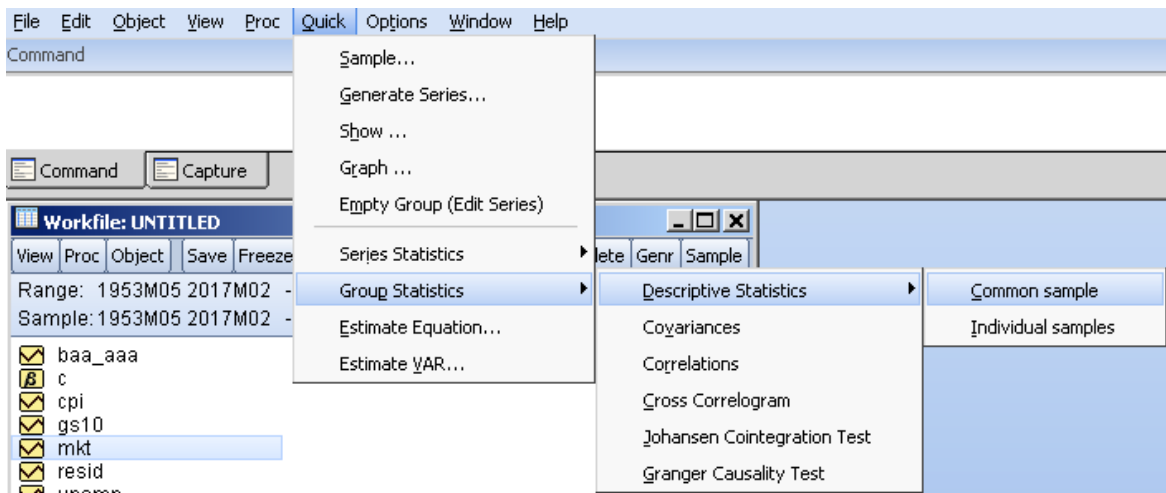
Once Common Sample is clicked, a window like the one below will appear. Make sure that the correct variable is shown in the text, then click OK.
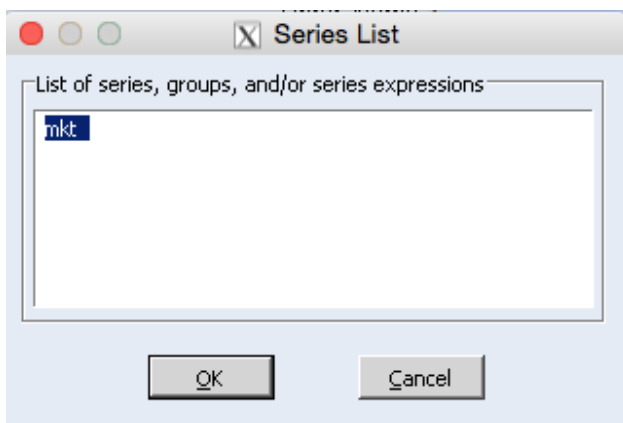
Once this is done, the descriptive statistics window will appear as such.

	MKT					
	MKT					
Mean	0.958277					
Median	1.285000					
Maximum	16.61000					
Minimum	-22.64000					
Std. Dev.	4.289684					
Skewness	-0.508139					
Kurtosis	4.902383					
Jarque-Bera	148.4725					
Probability	0.000000					
Sum	734.0400					
Sum Sq. Dev.	14077.06					
Observations	766					

The first few descriptive statistics are relatively easy to interpret. The mean is the average of all 766 observation values. The mean of 0.958277 means that, on average, the market moved up roughly 0.96% each month from 1953 to 2017. This shows that in the long run, if one were to invest in the stock market (randomly choosing stocks), he or she could expect to gain about 96 cents for every 100 dollars invested. Being keen about investing in stocks could significantly increase those long term earnings however. The median of 1.285 means that half the values of the CRSP Market Index fall below 1.285, while the other half are above that number. It is important to be able to distinguish the meanings behind the mean and median when there is significant skewness. This is due to the fact that the mean is significantly

affected by skewness, it gets "pulled" by extreme values called outliers, while the median is robust against such extreme values. Given that the mean is less than the median, we know that there are more extreme lower values than higher values due to the fact that the mean is being pulled downward.

The standard deviation is slightly harder to interpret. It tells us how tightly the data values are clustered around the mean.[1] The standard deviation of 4.29 means that about 68% of the data values fall between plus or minus 4.29 of the mean of 0.96, which is within one standard deviation. Likewise, 95% of the data values fall between two standard deviations of the mean, or between plus or minus 8.58 of the mean, and this pattern continues when we assume that the data follows

---

[1] Niles, R. (n.d.). Standard Deviation. Retrieved April 23, 2017, from http://www.robertniles.com/stats/stdev.shtml

the normal distribution which is a bell-shaped curve as shown below.
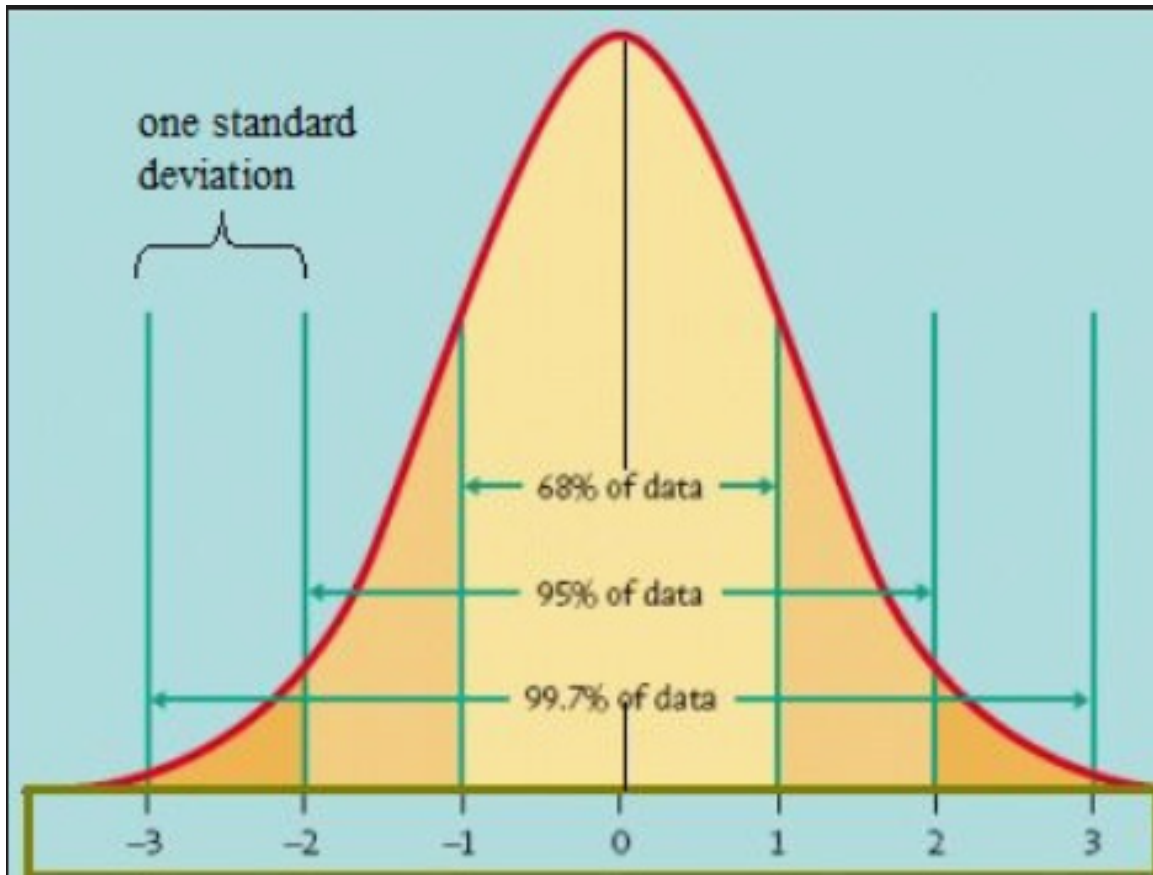
Next we have skewness, which is a measure of symmetry. In a perfect normal distribution, the skewness is zero. The negative skewness of -0.51 indicates that the left tail of our distribution is larger than the right. This skewness ties in with the mean-median relationship. We could have assumed a negative skewness by observing before that the mean was lower than the median. Kurtosis, like skewness, also involves the tails of the distribution. Kurtosis is a measure of whether the datapoints are heavy-tailed or light-tailed relative to a normal distribution. Datasets with higher kurtosis have heavier tails than datasets with lower kurtosis. In a perfect normal distribution, the kurtosis is zero. The descriptive statistics above show that we have a kurtosis of 4.90, indicating that our market dataset has heavier

tails and a sharper peak than the normal distribution, as depicted by the picture

below.[2]
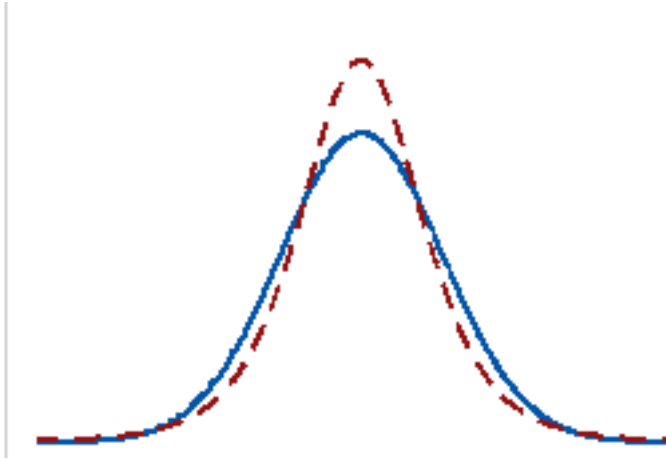
        Suppose that now we want to observe one of our variables graphically;

EViews has the capability of showing us what our data looks like through the form

of box plots, dot plots, histograms, etc. The first step is to highlight the desired

variable which you want to observe graphically. For our example, we will choose the

CRSP Market Index variable, which we named "mkt". Then you must click on Quick >

---

[2] Measures Of Skewness And Kurtosis. Retrieved April 25, 2017, from
http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm

Graph ... as shown below.
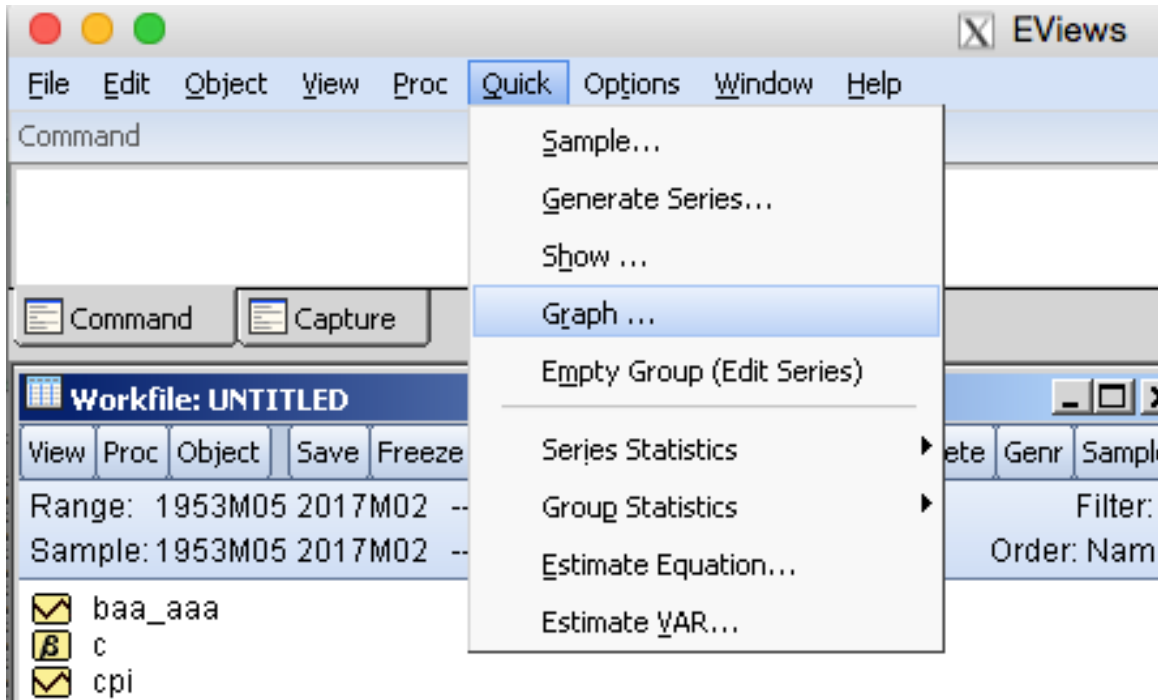
Once this is done, another window will pop up and you will need to ensure that the proper variable is selected before clicking OK.
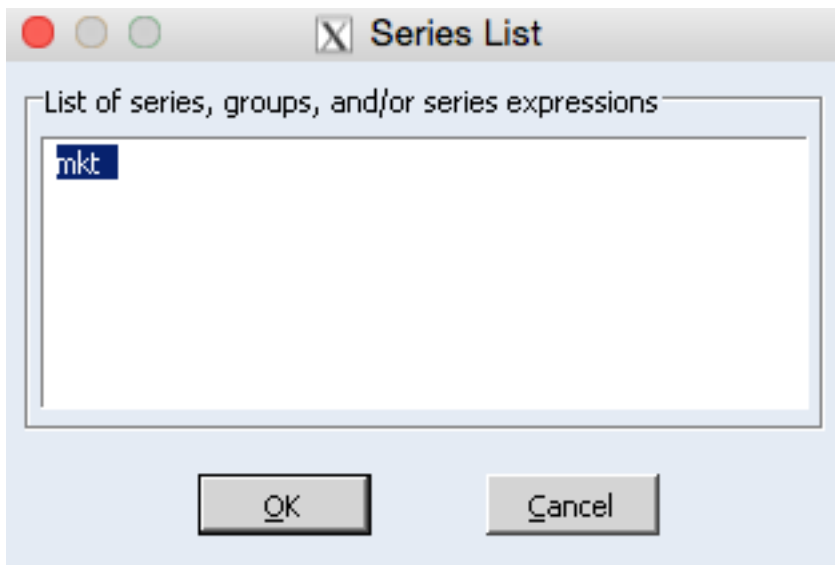
After clicking OK, the window below will appear, giving you various different

graphs that you can use to analyze the data visually.
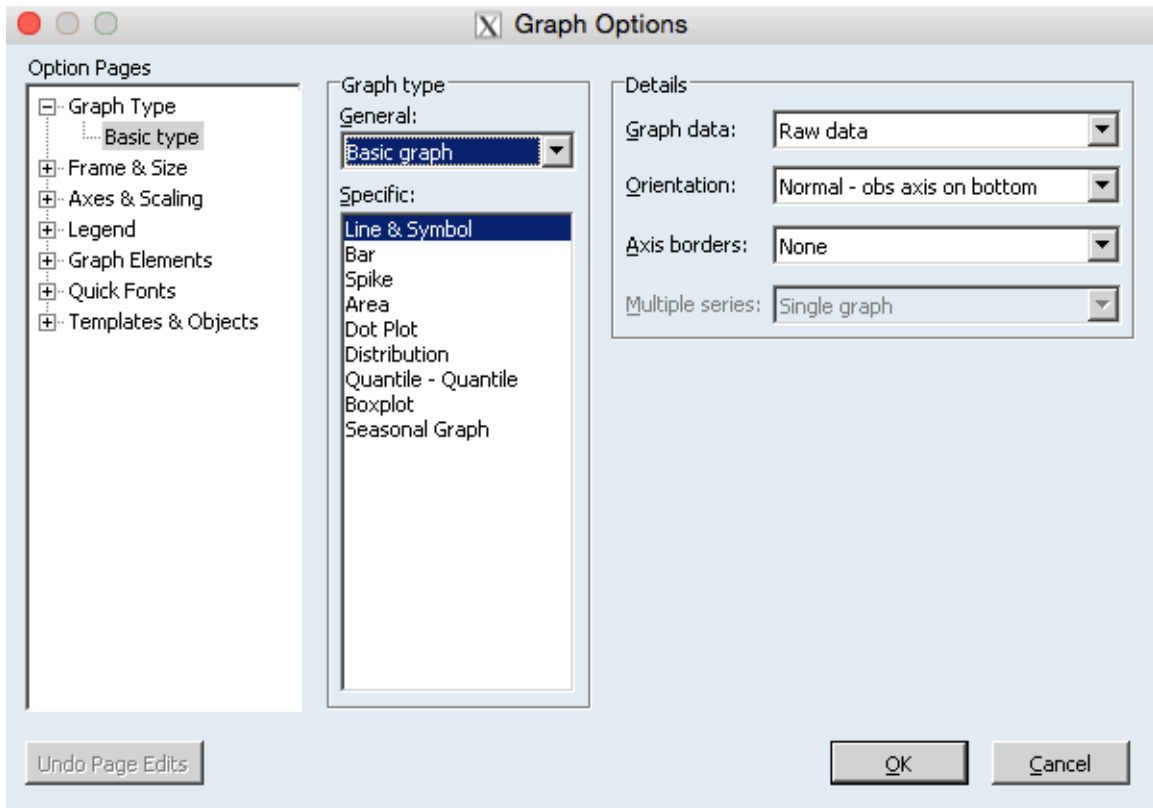
The first graph that we will analyze here is the line & symbol graph highlighted above.
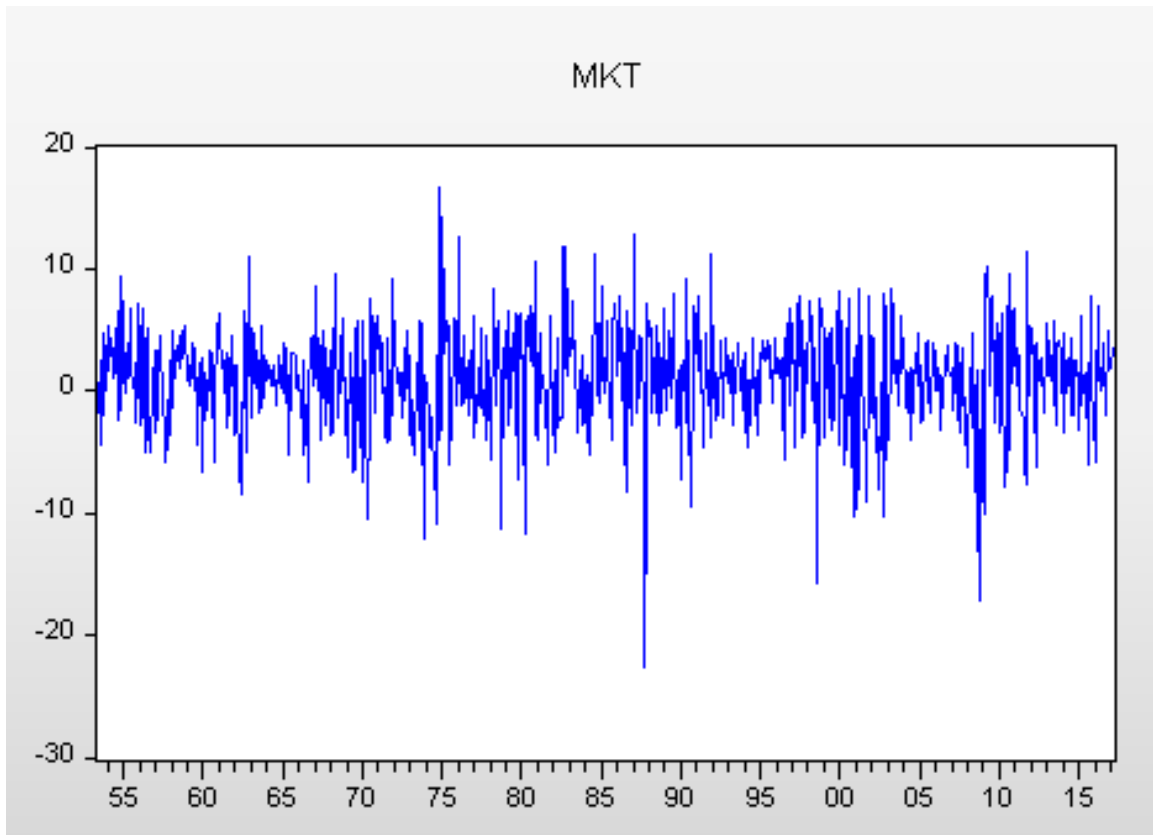
We can see that the lines are centered at about the zero axis, and they appear to be random from left to right. This is good to see, since the market is in fact very random, and the prices can fluctuate positively and negatively with no clear patterns. Had there been a clear pattern visible in our graph, or say all the lines ended up appearing above or below the zero axis, we could conclude that we did entered something wrong in EViews. We can also notice our minimum and maximum market change values as the longest blue lines below and above the zero axis, and we can easily tell that these lines match with our given minimum and maximum values above in our descriptive statistics.

The bar, spike, area, and dot plot graphs can all be interpreted very similarly to the line & symbol graph above.
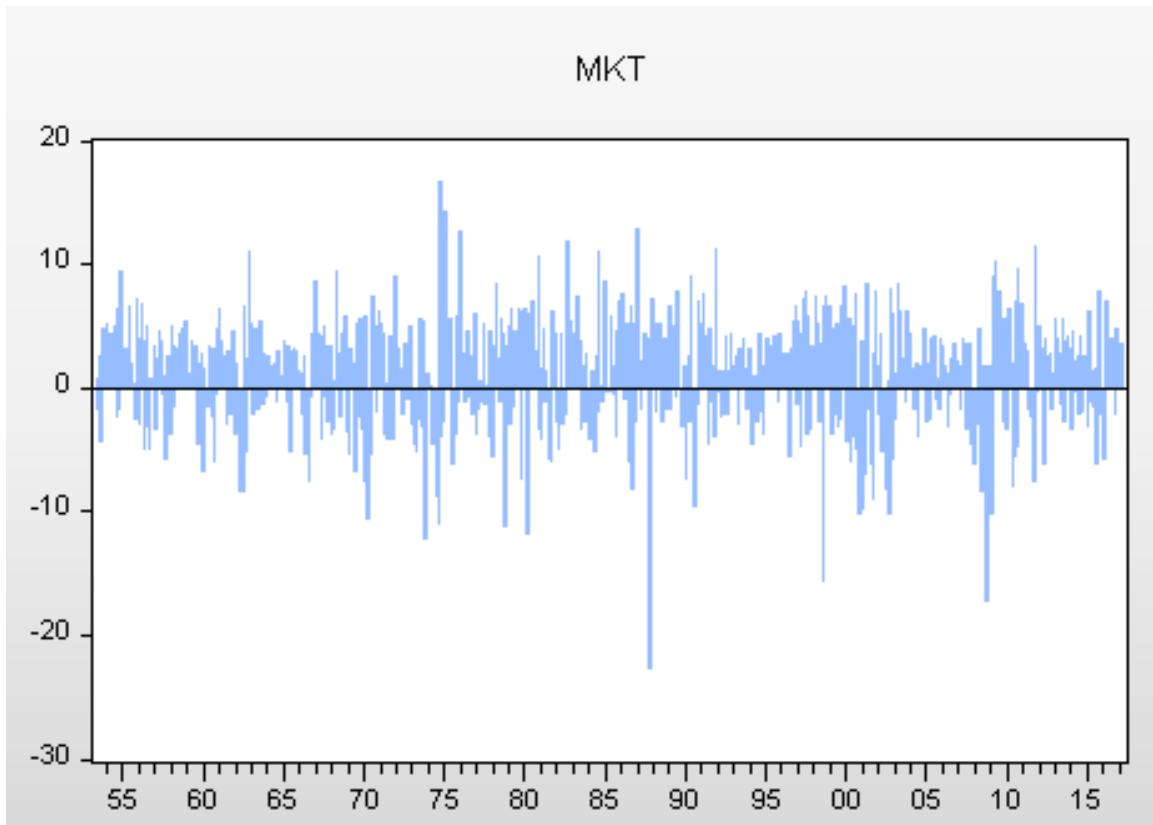
One interesting and distinct observation we can see from the bar graph above is that there is more area above the zero axis than below. This was not very clear in the line graph we previously analyzed. The spike and area graphs look extremely similar to the two graphs already shown above, so we will skip analyzing those. The dot plot looks much different from the graphs we have already seen, but we can easily gather the same information from this graph as we did with the line

and symbol graph.
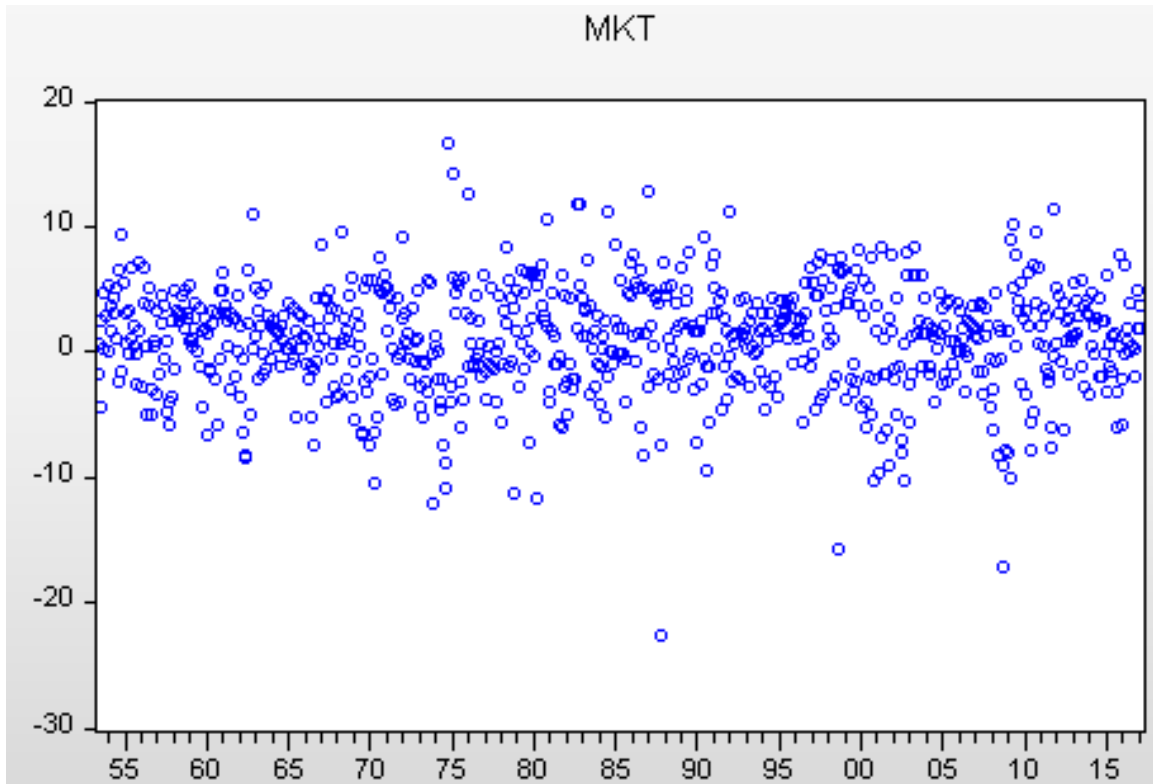
We can observe the minimum and maximum like we did earlier, as well as notice the randomness of the data in a unique way here. Other than that, there is not much more we can observe here. So if we move on to a distribution plot, we can make a few more observations that we could not do with the previous graphs.

In the distribution plot below, one of the first things we notice is that the data looks approximately normal, meaning that it closely resembles the normal distribution. However, it is clearly not perfectly normal, as we can observe some left skewness which we discussed earlier. In a perfectly normal distribution, the mean is equivalent to the median, but here we can conclude that the mean is slightly less than the median as it is being pulled by those extremely negative values. Again, we

can observe here that the mean appears positive if we look at the area under the

bars.
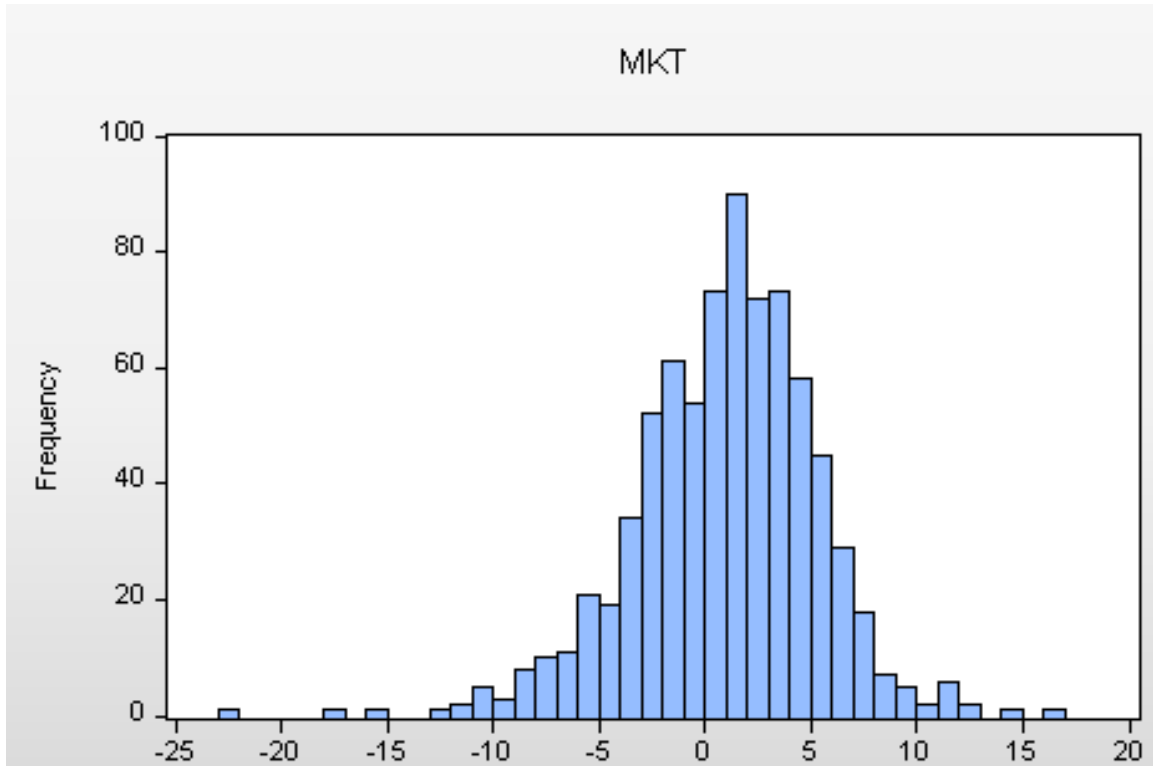
Looking at the above distribution graph, there also appear to be outliers

present in this data, which we will be able to confirm by looking at a box plot next.
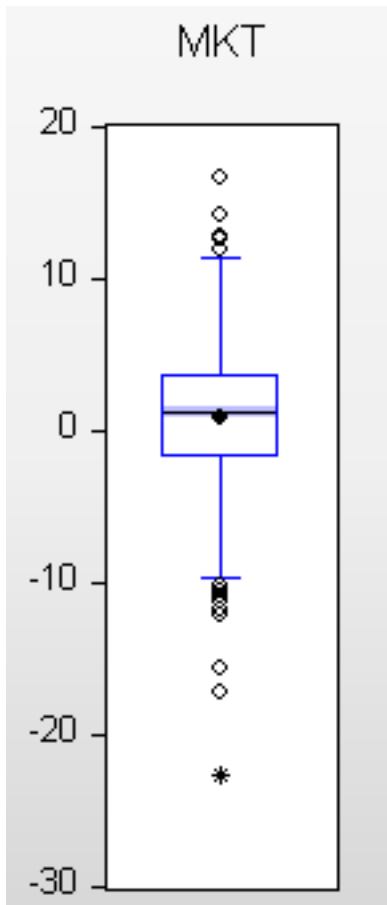
The box plot yields copious amounts of information, more than any other

graph that we have looked at thus far. It shows us the positive median which is the

center blue line, as well as the mean which is the black dot right below the median.

The box plot also divides the data up into groups; the middle 50 percent of the data

falls within the central box, 25 percent above the median and 25 percent below. This

leaves us with 25 percent of the data which appears above the box, and 25 percent

which appears below it. We noted above that there appeared to be outliers present

in this dataset; we can clearly see from this boxplot that that is the case. All of those

black dots above the upper whisker, and below the lower whisker are outliers.

There is a simple mathematical procedure used to determine outliers. The

interquartile range is what the central box is called. First you must multiply this

interquartile range by 1.5. Then this value must be added to the upper part of the

interquartile range, or the third quartile, and the value must be subtracted from the

lower part of the interquartile range, or the first quartile. This sets up the bounds

such that any data point outside this region is considered an outlier.[3] There are

actually more outliers than there appeared to be in the distribution graph. Often,

outliers in statistics are considered to be flaws in the data and are ignored, but in

instances such as this one where we are analyzing the market, outliers are perfectly

fine to have. In fact, they are practically normal in a sense since the market is always

fluctuating randomly; it is always a possibility that an extreme value appears on

occasion.

Next we will analyze the least squares regression statistics. To get EViews to

display these statistics for our datasets, click on Quick > Estimate Equation… as

---

[3] Outliers and Box Plots. Retrieved April 26, 2017, from
http://www.stat.wmich.edu/s160/book/node8.html

shown below.
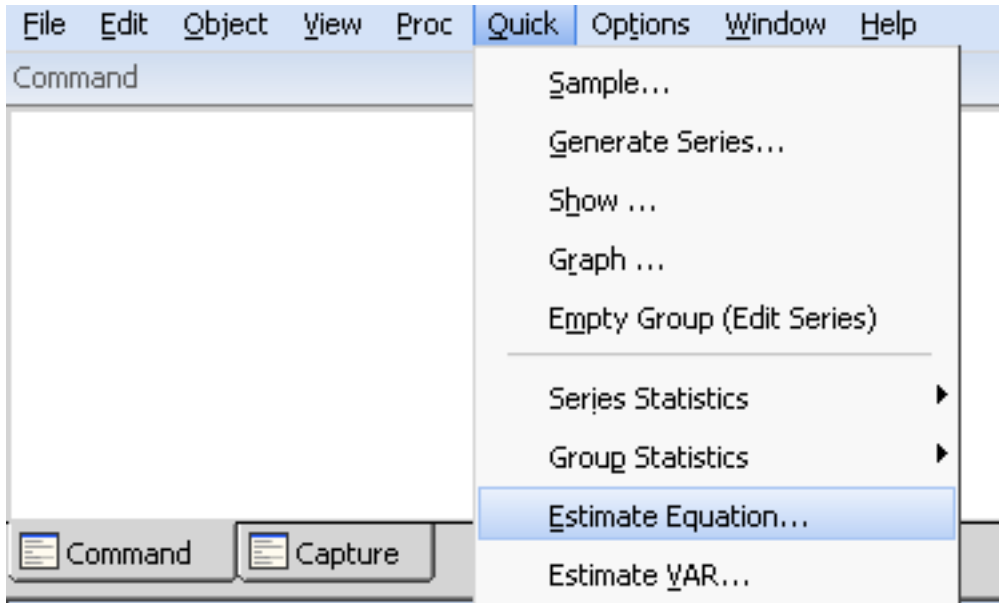
Then, a window like the one below will appear, and here we demonstrate

some of the language that EViews uses when we must enter in the regression

equation.
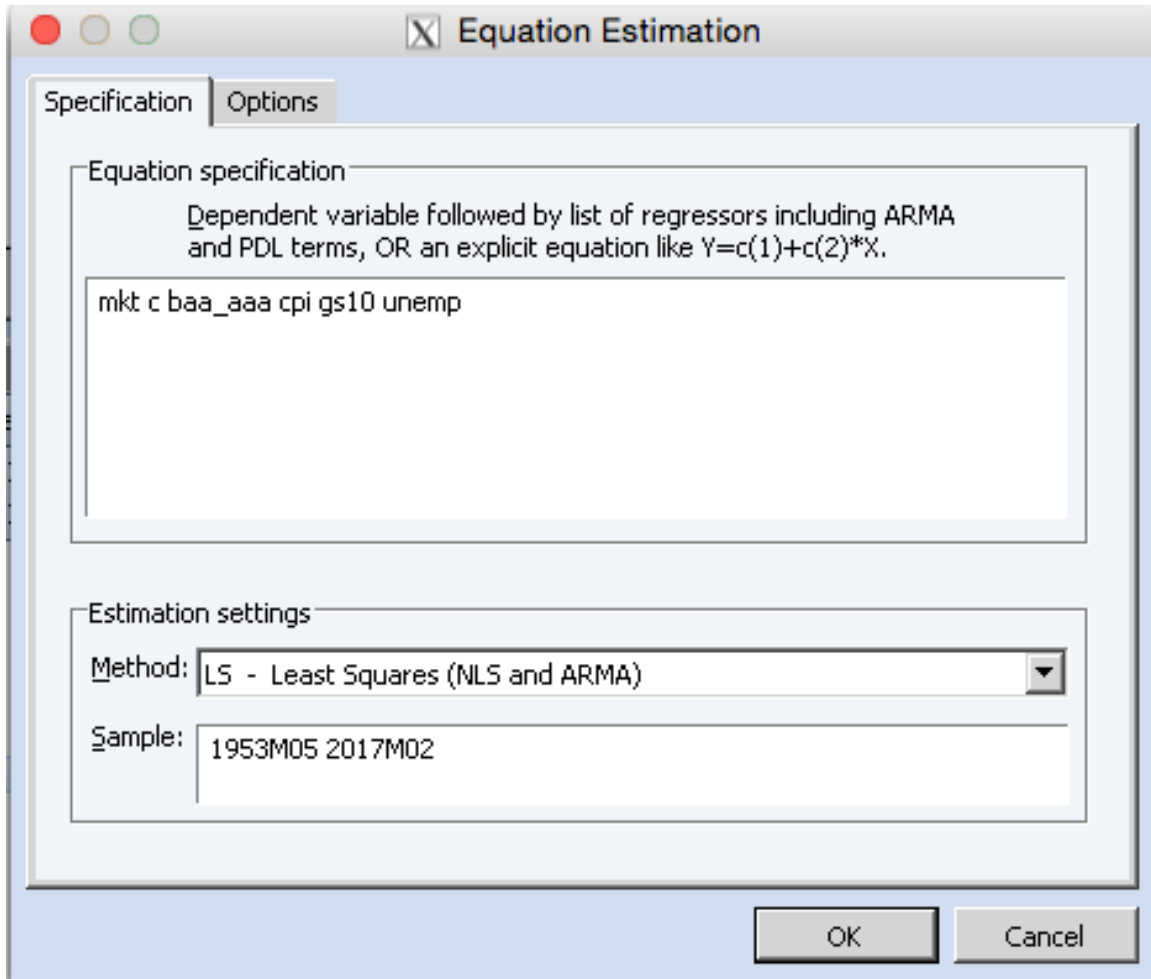
All of our variables are accounted for, and we make sure that the method used is LS – Least Squares (NLS and ARMA), and then select OK. Finally, we can observe a window that displays the least squares regression statistics that we are ready to analyze.

The equation that we base our multifactor model off of is:

$$Mkt = C + B_{baa\_aaa}baa\_aaa + B_{cpi}cpi + B_{gs10}gs10 + B_{unemp}unemp + e$$

In the above equation, the Bs are the loadings on each of the respective variables, C is the constant term, and e is the idiosyncratic error term. Notice how EViews

automatically takes into account the error term, and assumes the addition of the variables and the loadings of the variables, since when we inputted the variables, the only thing that separated the names of each of the variables was a space.

This multifactor model equation is important because it puts all of our variables that we wish to analyze together in a regression model. It helps us to analyze the relationship of our overall market variable with our other variables. Essentially, it is an attempt to explain market returns through our variables of Baa and Aaa corporate bond yield percents, 10-year treasury constant maturity rates, consumer price index, and unemployment rates. This is an interesting approach because although we can try to explain market returns via regression analysis, the efficient market hypothesis states that it will certainly not be able to explain returns in the market. The efficient market hypothesis is a financial theory that states that it is impossible to predict and outperform the market because stocks always trade at their fair value on stock exchanges.[4] The predictability of the stock market has been an ongoing debate since the creation of the stock market, and will probably continue on forever. Our multifactor regression model is just one of countless different methods that people have tried employing to attempt to predict the market, and this number is increasing all the time. There are some methods used in an attempt to predict the market that are better than others, but we still do not know of the best method to this date, and we may never find out, since the stock market is such a complex and dynamic entity.

---

[4] Efficient Market Hypothesis - EMH. (2016, September 28). Retrieved May 31, 2017, from http://www.investopedia.com/terms/e/efficientmarkethypothesis.asp

Displayed below is the output from our least squares regression model using the above equation.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.291034	0.211805	6.095387	0.0000
BAA_AAA	-0.033282	0.018400	-1.808809	0.0709
CPI	-1.073232	0.503362	-2.132130	0.0333
GS10	-0.037829	0.034569	-1.094310	0.2742
UNEMP	-0.016356	0.045323	-0.360881	0.7183

R-squared	0.012673	Mean dependent var	0.958277
Adjusted R-squared	0.007483	S.D. dependent var	4.289684
S.E. of regression	4.273603	Akaike info criterion	5.749298
Sum squared resid	13898.67	Schwarz criterion	5.779593
Log likelihood	-2196.981	Hannan-Quinn criter.	5.760960
F-statistic	2.441997	Durbin-Watson stat	1.875043
Prob(F-statistic)	0.045445		

A great thing about this table is that we can compare it to an ANOVA table in Excel so that we can make sure that we entered in our equation properly in EViews. After ensuring that the table we have displayed above is correct, we can now begin to analyze the relatively large amount of data that we have.

First we can analyze the coefficient column in the table. The coefficient is the change in response per unit change in the predictor.[5] For example, the overall market changes -0.016 units per unit change in the unemployment rate. Also, each coefficient provides an estimate of the sum of the risk premium associated with the corresponding variable, if any risk premium exists for that variable at all.[6] Looking

[5] Brooks, C. (2014). Introductory Econometrics for Finance (Second ed.). Cambridge, United Kingdom: Cambridge University Press. Page 29.
[6] Chen, N., Roll, R., & Ross, S. A. (n.d.). Economic Forces and the Stock Market. Retrieved May 30, 2017, from

at our table above we can see that all of our variables have negative risk premium, given their negative coefficients.

The standard errors are the standard errors of the regression coefficients, and can be used for testing hypotheses and for creating confidence intervals.[7] The smaller the standard error is, the larger the t-statistic becomes. Smaller standard errors do not require the estimated and hypothesized values to be far away from each other in order for the null hypothesis to be rejected. The null hypothesis and alternative hypothesis are statements that are tested in the hypothesis-testing framework. With hypothesis testing, a significance level known as alpha is chosen. The most common significance level is 5%, which means that 5% of the total distribution will be in the rejection region, or the region where we reject the null hypothesis and where there is explanatory power.[8] For example, in the table above, we can see that there is only one of our variables that has a probability number under 0.05, CPI. This means that CPI is the only variable that we have which has explanatory power, and the other variables do not have explanatory power at the 0.05 significance level.

With an R-square value of only 0.0127, the least squares regression model used explains very little of the variability of our response variables around their means.

http://rady.ucsd.edu/faculty/directory/valkanov/pub/classes/mfe/docs/ChenRollRoss_J B_1986.pdf

[7] How to Read the Output From Simple Linear Regression Analyses. Retrieved on May 15, 2017 from http://www.jerrydallal.com/lhsp/slrout.htm.

[8] Brooks. Page 56.

The F test will tell us if our group of variables is jointly significant. We can see that with our probability value of 0.045 for our f statistic, our variables are in fact jointly significant, and altogether have explanatory power.

The Durbin Watson (DW) statistic tests for a relationship between an error and its immediately previous value – it is a test for first order autocorrelation.[9] The DW statistic that we have of 1.875 indicates that there is some positive autocorrelation in the residuals of our variables.[10]

We have analyzed only a small portion of the statistical analyses outputs and financial applications that EViews is capable of producing, but it is a great start for any student interested in the actuarial, mathematical, or financial fields to learn more about an econometric program and how useful that program can be.

---

[9] Brooks. Page 144.
[10] Brooks. Page 146.