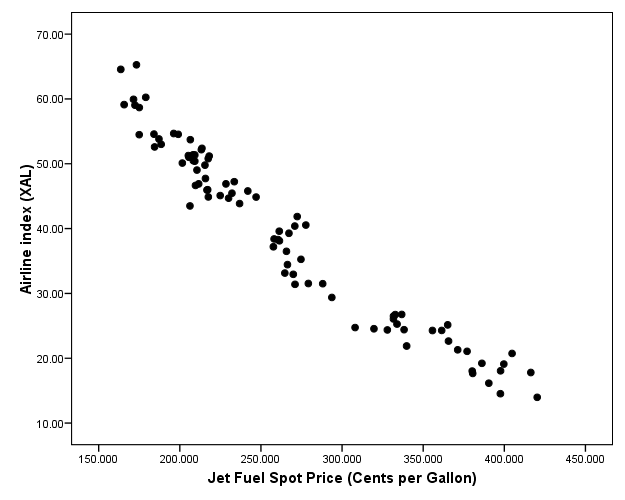
**HW4 Simple Linear Regression Analysis**

**Jet Fuel Spot prices and prices of Airline index (XAL).**

In normal times the Airline Index (XAL) stock price that is a weighted index of the stocks for large American airlines companies has strongly ties with jet fuel prices. Based on historical data, financial investors know that if jet fuel increases, the XAL index price will decrease, and viceversa. The two variables have a strong linear association as clearly displayed by the scatterplot below. So investors can use this knowledge to buy XAL index if jet fuel price goes down, and sell XAL index if jet fuel price goes up.

The data to be analyzed in this problem are XAL stock and jet fuel weekly prices from **January 5th, 2007 to August 31st, 2008**

* **The Airline Index** is designed to measure the performance of highly capitalized companies in the airline industry. The Index tracks the aggregate performance of major U.S. and overseas airlines, including Alaska Air Group Inc. (ALK); America West Holding Corporation CLB (AWA); AMR Corporation (AMR); Continental Airlines Inc. (CAL); Delta Air Lines Inc. (DAL); KLM Royal Dutch Airlines NV (KLM); Norwest Airlines Corp. (NWAC); Southwest Airlines Co. (LUV); UAL Corp. (UAL); US Airways Group Inc. (U).
* **Jet Fuel Spot Prices** are in cents per gallon.

Use the computer output attached below to find the fitted regression line to predict XAL index price from jet fuel price.

**Answer the following questions:**

1. Interpret the correlation between the two variables, and write down the regression model. [10]

Jet fuel spot price is an independent variable and Airline index is dependent variable.

So it is one dependent variable – one independent variable relationship. Therefore it is simple linear regression : = = 87.55209 - 0.18163 \*

In the Pearson Correlation Coefficients table, we can see the value r is equal to -0.97331.

So we know that it is very closer to perfect linear relationship since r is very closer to -1.

Besides, It is negative value so we know that it is a negative linear relationship.

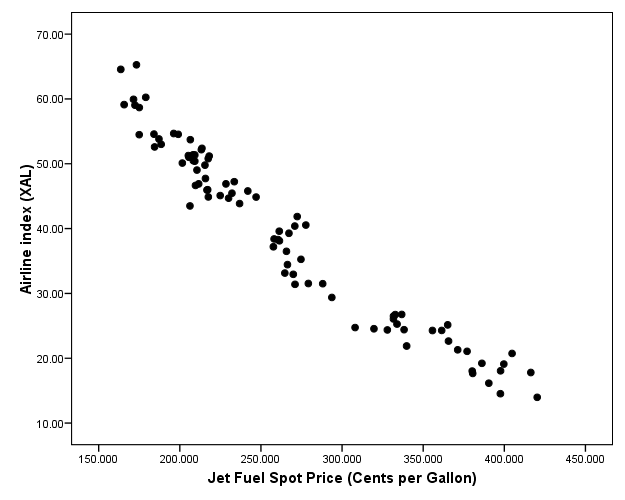
1. What’s the adjusted R2 value for the regression line? Explain it. Why we prefer to use adjusted R2 rather than the raw R2 value? [15]

Adjusted R2 value = 0.9467,

adjR2 = , n = the sample size = 86, k = the number of independent variables = 1.

Unlike the R2, the Adj-R2 value does not increase with the addition of a x-variable that does not improve the regression model and a higher adj-R2 typically indicates a better model, in terms of the training set.

1. Use the regression line to predict the index price: [10]
   1. Predict the XAL price on Sept. 7th, knowing that the fuel price was 224.885 cents.
   2. Predict the XAL price for Oct. 5th, knowing that the fuel price was 301.928 cents.



­

1. XAL price on Sept. 7th is approximately equal to $48.
2. XAL price on Oct. 5th is approximately equal to $32.

**Computer output**

The CORR Procedure

2 Variables: fuel xal

Simple Statistics

Variable N Mean Std Dev Sum Minimum Maximum

fuel 86 267.24903 72.53431 22983 163.60200 420.25000

xal 86 39.01081 13.53587 3355 13.98000 64.56000

Simple Statistics

Variable Label

fuel Weekly jet fuel spot price

xal Airline index weekly price

Pearson Correlation Coefficients, N = 86

Prob > |r| under H0: Rho=0

fuel xal

fuel 1.00000 -0.97331

Weekly jet fuel spot price <.0001

xal -0.97331 1.00000

Airline index weekly price <.0001

The REG Procedure

Model: MODEL1

Dependent Variable: xal Airline index weekly price

Number of Observations Read 86

Number of Observations Used 86

Analysis of Variance

Sum of Mean

Source DF Squares Square F Value Pr > F

Model 1 14754 14754 1511.05 <.0001

Error 84 820.15556 9.76376

Corrected Total 85 15574

Root MSE 3.12470 R-Square 0.9473

Dependent Mean 39.01081 Adj R-Sq 0.9467

Coeff Var 8.00983

Parameter Estimates

Parameter Standard

Variable Label DF Estimate Error t Value Pr > |t|

Intercept Intercept 1 87.55209 1.29340 67.69 <.0001

fuel Weekly jet fuel spot price 1 -0.18163 0.00467 -38.87 <.0001

1. Here are the observed XAL weekly prices on Sept. 7th is $45.1, and on Oct. 5th was $46.9. What are the prediction errors for the two observations? If you were an investor, which of the two predictions would you trust more? Explain your answer. [10]

Error of prediction = y – y hat.= observed Y – predicted Y

Sept. 7th’s error of prediction = 45.1 - 48 = -2.9

Oct. 5th’s error of prediction = 46.9 - 32 = 14.9

the accuracy of the prediction depends on how much spread out the observations are around the line. Therefore, if I were an investor, I would like to trust Sept. 7th more.

1. As we all know so well, financial markets were in turmoil last year, caused by a severe crisis of the banking industry. The graphs below shows the data for XAL stock price and fuel price until March 2010. Examine what the graphs show. Does the association between the two variables change after August 2008? [15]

*Data collected from* ***Jan. 2007 to March 2010****.*

The relationship between the two variables was inversely proportional until August 2008.

But now we can see that the relationship is proportional after September 2008.

And it looks like error of prediction increased and Given that the slope is gentle, we can now see that the effect of independent variable is less than before subprime situation.

***Data from January 2007 to August 2008 Data from Sept 2008 to March 2010***

*Dual axis Plot that shows the changes in XAL index price and jet fuel price over time*

1. If you were an investor, would you still use information about jet fuel price to make a buy or sell decision for the XAL index? [10]

Yes. In the subprime situation, sentiment on market has been applied too much.

But contrary to the time before that situation, now we can use the proportional relationship between the two variables to the investment.

1. What are the steps in residual analysis. And further state how you can identify the problems in each step, and how to solve these issues accordingly. [30]

Steps in Residual Analysis :

1. Validate the constant variance

Plot residual vs predicted values: To check constant variance for the residuals.

So if the pattern of the spread in the residuals as the predicted values increase is constant, It is acceptable. But pattern of the spread increases or decrease then increase, It is unacceptable.

So we may need to apply a transformation on y, such as log transformation, and then re-fit the regression model.

1. Validate the linearity relationship

Plot residual vs each x-variables: To check linearity assumptions for Y and the X-variable.

If points are randomly scattered around the zero line, it is good case of linear relationship.

But if points are having a certain pattern, it is not a linear relationship.

So like the first step, we may need to apply a transformation on y, such as log transformation.

1. Validate normal distribution of residuals

So there are two ways to check whether a variable (including residual) follows normal distribution or not.

Solution-1 is QQPlot and Solution-2 is Normality test.

In the QQplot, if points are not follow the linear line, it means the variable does not follows normal distribution.

In the Normality test, For example, in Shapiro-Wilk Normality Test, if p-value > 0.05, we say it follows normal distribution at 95% confidence level.

1. Identify potential outliers

We often use the standardized residuals to identify outliers.

Possible outliers are observations with standardized/studentized residuals |