\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Homework 2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Instructions:

\* To create this document, first copy and paste the full text here into a .Do document (a STATA Do-File).

\* Below each question, write the code you used to answer the question

\* Next, write your actual answer to the question by commenting out your writing (by starting the line with a \*)

\* Next, copy and paste the entire document (my writing and yours) into a Word document. This will allow me to see your code on Canvas without downloading every homework.

\* The goal is that I should be able to copy and paste your entire text into a .Do File and run the code without any errors.

\* Finally, submit file as Homework 2 on Canvas

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Topic 1: Data Management in STATA

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* 1. Import the AssetReuturns file

import excel "/Users/henryvelasquez/Documents/MBA/MBA S3/2nd 7/Machine Learning in Finance K579/HW2/AssetReturns (1).xlsx", sheet("Sheet1")firstrow case(lower) clear

frame rename default returnset

frame create recessionset

frame change recessionset

import excel "/Users/henryvelasquez/Documents/MBA/MBA S3/2nd 7/Machine Learning in Finance K579/HW2/RecessionDates.xlsx", sheet("FRED Graph") firstrow case(lower)

frame change returnset

/\* I was attempting to merge in a different way but gave up. \*/

\* 2. Estimate the Annual Equity Risk Premium in two ways and create a new variable for each called ERP1 and ERP2

\* EQUITY RISK PREMIUM = Asset Return - Risk Free Return

\* A: Annual S&P500 Return - Annual Treasury Bill

codebook annualreturnsp500

codebook annualreturntbills

/\* subtract average annual return of treasury bills from average return from sp500

.118206 - .03327 = .084936

\*/

gen ERP1 = .084936

\* B: Annual S&P500 Return - Annual Treasury Bonds

codebook annualreturntbonds

/\* subtract average annual return from t bonds from average annual return of S&P500

.118206 - .051099 = .067107

\*/

gen ERP2 = .067107

\* Which is the larger. Why?

/\*The larger return is ERP1 or Annual Treasury Bills. Treasury bills are typically short term investments that have much lower payouts. If you were to invest in longer term more volatile stocks you could get a better return\*/

\* 3. Rename your two new variables ERP\_Bills and ERP\_Bonds

rename ERP1 ERP\_Bills

rename ERP2 ERP\_Bonds

\* 4. Label the new variables Equity Risk Premium Bonds and Equity Risk Premium Bills

label variable ERP\_Bills "Equity Risk Premium Bonds"

label variable ERP\_Bonds "Equity Risk Premium Bills"

\* 5. Save your file under a new name: New\_Homework1

save New\_Homework1, replace

\* 6. Drop observations prior to 1940 and restimate the mean Equity Risk Premiums. Did they increase or decrease? Why?

drop if year<1940

/\*

sp500Return = .125966

treasuryBillReturn = .03617

treasuryBondReturn = .053074

\_\_\_\_\_\_\_\_\_\_\_\_| original | new

ERP Bills | .084936 | .089796

ERP Bonds | .067107 | .072892

The ERP's increased because you cut off the lower range of S&P500 which showed big losses in annual returns. Therefore, you increased the average annual return and gained a bigger difference between the S&P and the risk free assets.

\*/

\* 7. Sort the data by inflation. Which year has the lowest inflation?

sort inflationrate

li year inflationrate

/\* Year 1949 \*/

\* 8. Sort the data by inflation, with highest inflation first

gsort inflationrate

li year inflationrate

/\* Year 1946 \*/

\* 9. Merge the original dataset (Homework1) to a new dataset called RecessionDates

merge 1:1 year using "/Users/henryvelasquez/Documents/MBA/MBA S3/2nd 7/Machine Learning in Finance K579/HW2/New\_Homework1.dta"

\*Is this a 1:1 merge, a m:1 merge, or a 1:m merge?

\*Why aren't some matched?

/\* This is a 1:1 merge. Some are not matched because there is data missing for earlier years in the recession data \*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Topic 2: Statsistical Analysis in STATA

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* 10. Estimate the correlation between recessions and AnnualReturns on Corporate Bonds

correlate recession

/\*

| recess~n a~ebonds

-------------+------------------

recession | 1.0000

annua~ebonds | -0.0878 1.0000

Correlation coefficient -0.0878

\*/

\* 11. Run a linear regression with the y-Variable as Recession and the x-variables are: Annual Returns on S&P500, inflation, and Annual Returns on Real Estate

reg recession annualreturnsp500 inflationrate annualreturnrealestate

/\*

Source | SS df MS Number of obs = 94

-------------+---------------------------------- F(3, 90) = 5.50

Model | 1.95353137 3 .651177124 Prob > F = 0.0016

Residual | 10.6528516 90 .118365018 R-squared = 0.1550

-------------+---------------------------------- Adj R-squared = 0.1268

Total | 12.606383 93 .135552505 Root MSE = .34404

-------------------------------------------------------------------------------------

recession | Coefficient Std. err. t P>|t| [95% conf. interval]

--------------------+----------------------------------------------------------------

annualreturnsp500 | -.5184188 .1860427 -2.79 0.006 -.8880251 -.1488126

inflationrate | -1.41147 1.080596 -1.31 0.195 -3.558263 .735322

annualreturnreale~e | -.8275749 .6921328 -1.20 0.235 -2.202618 .5474678

\_cons | .3003492 .0503616 5.96 0.000 .2002972 .4004013

-------------------------------------------------------------------------------------

\*/

\* 12. Estimate the predicted values of the regression

predict pred

li year pred if pred !=.

/\*

+------------------+

| year pred |

|------------------|

74. | 1928 .0772721 |

75. | 1929 .3522395 |

76. | 1930 .5564958 |

77. | 1931 .7266204 |

78. | 1932 .5767457 |

|------------------|

79. | 1933 .0620469 |

80. | 1934 .2609816 |

81. | 1935 -.0650168 |

82. | 1936 .087652 |

83. | 1937 .4220045 |

|------------------|

84. | 1938 .194995 |

85. | 1939 .3168103 |

86. | 1940 .3182504 |

87. | 1941 .2957431 |

88. | 1942 .0459543 |

|------------------|

89. | 1943 .0338966 |

90. | 1944 .0320184 |

91. | 1945 -.0145948 |

92. | 1946 -.1112932 |

93. | 1947 -.0273249 |

|------------------|

94. | 1948 .2115484 |

95. | 1949 .2339512 |

96. | 1950 .0268005 |

97. | 1951 .0428312 |

98. | 1952 .1591741 |

|------------------|

99. | 1953 .2006994 |

100. | 1954 .0306995 |

101. | 1955 .1261223 |

102. | 1956 .212045 |

103. | 1957 .2911332 |

|------------------|

104. | 1958 .0433927 |

105. | 1959 .2124992 |

106. | 1960 .2730183 |

107. | 1961 .1446754 |

108. | 1962 .3246011 |

|------------------|

109. | 1963 .1422765 |

110. | 1964 .1911062 |

111. | 1965 .1952273 |

112. | 1966 .2931023 |

113. | 1967 .1148571 |

|------------------|

114. | 1968 .1435079 |

115. | 1969 .1977083 |

116. | 1970 .135248 |

117. | 1971 .1453858 |

118. | 1972 .130301 |

|------------------|

119. | 1973 .2232929 |

120. | 1974 .1771075 |

121. | 1975 -.0454486 |

122. | 1976 .0405169 |

123. | 1977 .1207266 |

|------------------|

124. | 1978 .0091908 |

125. | 1979 -.0969551 |

126. | 1980 -.1021535 |

127. | 1981 .1566055 |

128. | 1982 .1357944 |

|------------------|

129. | 1983 .0917299 |

130. | 1984 .1739829 |

131. | 1985 .0229395 |

132. | 1986 .1094375 |

133. | 1987 .1424881 |

|------------------|

134. | 1988 .0925476 |

135. | 1989 .0352698 |

136. | 1990 .2356823 |

137. | 1991 .1017644 |

138. | 1992 .2138009 |

|------------------|

139. | 1993 .1919718 |

140. | 1994 .2349959 |

141. | 1995 .0567497 |

142. | 1996 .1158837 |

143. | 1997 .0714891 |

|------------------|

144. | 1998 .0773261 |

145. | 1999 .0906664 |

146. | 2000 .2225147 |

147. | 2001 .2847048 |

148. | 2002 .3015367 |

|------------------|

149. | 2003 .0455222 |

150. | 2004 .0857759 |

151. | 2005 .115232 |

152. | 2006 .1692557 |

153. | 2007 .259041 |

|------------------|

154. | 2008 .58787 |

155. | 2009 .159341 |

156. | 2010 .2364436 |

157. | 2011 .2797928 |

158. | 2012 .1401171 |

|------------------|

159. | 2013 .0237895 |

160. | 2014 .1822082 |

161. | 2015 .2397747 |

162. | 2016 .1661697 |

163. | 2017 .1071445 |

|------------------|

164. | 2018 .2578301 |

165. | 2019 .0756905 |

166. | 2020 .1020802 |

167. | 2021 -.0826466 |

+------------------+

\*/

\* 13. Estimate the residuals of the regression

predict residuals

li residuals if residuals !=.

/\*

residuals |

|-----------|

74. | .0772721 |

75. | .3522395 |

76. | .5564958 |

77. | .7266204 |

78. | .5767457 |

|-----------|

79. | .0620469 |

80. | .2609816 |

81. | -.0650168 |

82. | .087652 |

83. | .4220045 |

|-----------|

84. | .194995 |

85. | .3168103 |

86. | .3182504 |

87. | .2957431 |

88. | .0459543 |

|-----------|

89. | .0338966 |

90. | .0320184 |

91. | -.0145948 |

92. | -.1112932 |

93. | -.0273249 |

|-----------|

94. | .2115484 |

95. | .2339512 |

96. | .0268005 |

97. | .0428312 |

98. | .1591741 |

|-----------|

99. | .2006994 |

100. | .0306995 |

101. | .1261223 |

102. | .212045 |

103. | .2911332 |

|-----------|

104. | .0433927 |

105. | .2124992 |

106. | .2730183 |

107. | .1446754 |

108. | .3246011 |

|-----------|

109. | .1422765 |

110. | .1911062 |

111. | .1952273 |

112. | .2931023 |

113. | .1148571 |

|-----------|

114. | .1435079 |

115. | .1977083 |

116. | .135248 |

117. | .1453858 |

118. | .130301 |

|-----------|

119. | .2232929 |

120. | .1771075 |

121. | -.0454486 |

122. | .0405169 |

123. | .1207266 |

|-----------|

124. | .0091908 |

125. | -.0969551 |

126. | -.1021535 |

127. | .1566055 |

128. | .1357944 |

|-----------|

129. | .0917299 |

130. | .1739829 |

131. | .0229395 |

132. | .1094375 |

133. | .1424881 |

|-----------|

134. | .0925476 |

135. | .0352698 |

136. | .2356823 |

137. | .1017644 |

138. | .2138009 |

|-----------|

139. | .1919718 |

140. | .2349959 |

141. | .0567497 |

142. | .1158837 |

143. | .0714891 |

|-----------|

144. | .0773261 |

145. | .0906664 |

146. | .2225147 |

147. | .2847048 |

148. | .3015367 |

|-----------|

149. | .0455222 |

150. | .0857759 |

151. | .115232 |

152. | .1692557 |

153. | .259041 |

|-----------|

154. | .58787 |

155. | .159341 |

156. | .2364436 |

157. | .2797928 |

158. | .1401171 |

|-----------|

159. | .0237895 |

160. | .1822082 |

161. | .2397747 |

162. | .1661697 |

163. | .1071445 |

|-----------|

164. | .2578301 |

165. | .0756905 |

166. | .1020802 |

167. | -.0826466 |

+-----------+

\*/

\* 14. Estimate the Mean-Squared Error

reg recession annualreturnsp500 inflationrate annualreturnrealestate

/\* RMSE = .34404 \*/

/\* HERE IS THE OTHER WAY FROM CLASS\*/

sum(residuals)

/\*std dev of residuals = .03384476\*/

\* 15. Run a Logistic Regression where the y-variable is Recession and the x-variable is Annual Return on Corporate Bonds and Inflation

logit recession annualreturncorporatebonds inflationrate

/\*

Iteration 0: log likelihood = -41.262576

Iteration 1: log likelihood = -38.298118

Iteration 2: log likelihood = -38.141588

Iteration 3: log likelihood = -38.141098

Iteration 4: log likelihood = -38.141098

Logistic regression Number of obs = 94

LR chi2(2) = 6.24

Prob > chi2 = 0.0441

Log likelihood = -38.141098 Pseudo R2 = 0.0756

-------------------------------------------------------------------------------------

recession | Coefficient Std. err. z P>|z| [95% conf. interval]

--------------------+----------------------------------------------------------------

annualreturncorpo~s | -3.325744 3.999197 -0.83 0.406 -11.16403 4.512539

inflationrate | -17.8441 8.247193 -2.16 0.030 -34.0083 -1.679898

\_cons | -1.024286 .4196204 -2.44 0.015 -1.846727 -.2018453

-------------------------------------------------------------------------------------

\*/