Nishal Dave: AFE Coursework Draft

The analysis is to explore the relationship between the BSE SENSEX index prices and the change in the slope of the yield curve over the ten-year period from 01/2010 to 12/2020, the yield curve slope has been computed using the spread between 15-year and 1-year bond yields. The intuition is supported by the fact that as the slope steepens this represents a widening in yields, therefore in times of uncertainty and/or low confidence bond yields should drop as investors substitute their equity for safer investments and market forces drive the prices of bonds up and vice versa for when bond prices fall. This is characterised by the following formula:

Where the annual coupon is held fixed, market forces allow the bond price to vary, in turn causing variation in the yield. Longer term bonds bear more risk to the holder compared to short term bonds and therefore under normal circumstances should have a higher yield.

My support for this selection is loosely based on the analysis by [*Chen, Roll and Ross(1986)*](https://rady.ucsd.edu/faculty/directory/valkanov/pub/classes/mfe/docs/ChenRollRoss_JB_1986.pdf), who find that the term structure of bond yields have a negative and inverse relationship with stock returns identified using the Arbitrage Pricing Theory approach. This analysis will aim to model a relation using a traditional time-series approach.

*Insert (Preliminary Analysis/insert series plots/structural break justification)*

*Insert check for spurious regression i.e. regressing level variables upon each other (looking at the Rsq and DW tstat*

*)*

*Insert (DF/ADF procedure)*

So far we have concluded that the original data are I(1), and then I(0) as a result of taking their differences. This makes the returns variables suitable for VAR modelling, however before this, it would be worthwhile to check if there exists some cointegrating relationship between the variables, before settling on a model.

Because this is a bivariate analysis, using an ECM constructed by the Engle Granger process is a sufficient enough method to identify any cointegrating relationship, as there can only be a maximum of one in this case, as it may be frivolous to use a VECM.

Before doing so, it would be advisable to use a monotonic transformation on the variables, such as a log specification. This makes interpretation easier in the context of stock returns.

Engle Granger Procedure

Step 1

Where:

In addition to this specification, we can present the long-run equilibrium by the following expression:

|  |  |
| --- | --- |
| EG Step 1: | Regression Output |
|  |  |
| Intercept | 9.64  (0.015) \*\*\* |
|  | -1.14  (0.249) \*\*\* |
| Time Trend | 0.00775  (0.0002) \*\*\* |
|  | 0.9199 |
|  | 131 |

, std. errors are reported in brackets

Running the ADF test on the residuals:

*Insert (ADF procedure)*

Step 2

Based on the ADF test, the test-stat exceeds all levels of significance, therefore, we can safely reject the null of a unit root and conclude that the residuals are stationary: , this gives evidence that there exists some cointegrating relationship between the two variables. Which takes us to the second step of the EG procedure. For this bivariate case the ECM model:

*Insert – (Discussion about lag lengths [below is not accurate])*

simplifies to the following specification:

Where is the lagged value of the long-run equilibrium and is defined as .

Therefore, the coefficient on is speed of adjustment and measures the time taken to converge to the long run equilibrium, which is just over 7 months using the following calculation

|  |  |
| --- | --- |
| EG Step 2: | Regression Output |
|  |  |
| Intercept | 0.007  (0.003) \* |
|  | -0.353  (1.114) \*\* |
|  | -0.139  (0.040) \*\* |
|  |  |
|  | 0.1065 |
|  | 131 |

As we can see, the speed of adjustment term is statistically significant at the 5% level giving confidence that changes in slope of the yield curve Granger cause returns to the BSE index. As the coefficient upon the slope of the yield curve is negative, this supports the hypothesis of an inverse relationship between the two variables

*Insert – (Diagnostics?)*

*Insert - (further discussion & IRF)*

1. #DATA SOURCES
2. #https://uk.investing.com/rates-bonds/india-1-year-bond-yield
3. #https://uk.investing.com/rates-bonds/india-15-year-bond-yield

6. #PRELIMINARIES
7. rm(list = ls())
8. library(zoo)
9. library(xts)
10. library(tidyverse)
11. library(urca)
12. library(vars)
13. library(tsDyn)
14. options(scipen = 999)
16. setwd("/Users/nishaldave/OneDrive - University of Bristol/TB2/Applied Financial Econometrics/Coursework")
17. # setwd("C:/Users/Nish/OneDrive - University of Bristol/TB2/Applied Financial Econometrics/Coursework")
18. bse<-read.csv("S&P BSE SENSEX.csv")
19. sr<-read.csv("1yr Historical Bond Data India.csv")
20. lr<-read.csv("15yr Historical Bond Data India.csv")


24. #DATA PROCESSING/FORMATTING
25. bse$Close<-as.numeric(as.character(bse$Close))
26. bse$Date<-as.Date(bse$Date,"%d/%m/%Y")
27. sr$Date<-as.Date(sr$Date,"%B %d, %Y")
28. lr$Date<-as.Date(lr$Date,"%B %d, %Y")
30. #PROCESSING BONDS
31. data <- merge(sr, lr, join = "right")
32. data <- data.frame(sr=data$X1yr,
33. lr=data$X15yr,
34. Date = data$Date,
35. year = as.numeric(format(data$Date, format = "%Y")),
36. month = as.numeric(format(data$Date, format = "%m")),
37. day = as.numeric(format(data$Date, format = "%d")))
38. data$ym <- as.yearmon(paste(data$year, data$month), "%Y %m")
39. sr\_xts <- xts(data$sr, data$ym)
40. lr\_xts <-xts(data$lr, data$ym)
41. sr\_xts<-apply.monthly(sr\_xts,mean)
42. lr\_xts<-apply.monthly(lr\_xts,mean)
43. bonds\_xts <- merge(sr\_xts, lr\_xts, join = "right")
45. #PROCESSING INDEX VALUES
46. data2 <- data.frame(bse=bse$Close,
47. Date = bse$Date,
48. year = as.numeric(format(bse$Date, format = "%Y")),
49. month = as.numeric(format(bse$Date, format = "%m")),
50. day = as.numeric(format(bse$Date, format = "%d")))
52. data2$ym <- as.yearmon(paste(data2$year, data2$month), "%Y %m")
53. data2\_xts <- xts(data2$bse, data2$ym)
54. #TAKING A DAILY AVERAGE FOR EACH MONTH TO CONVERT INTO MONTHLY DATA FOR 10 YEARS
55. bse\_xts<-apply.monthly(data2\_xts,mean)
57. bseND <- merge.xts(bse\_xts, bonds\_xts,join = "right", fill = NA)
58. colnames(bseND)[1:3] <-c("bse","sr","lr")


62. rm(bse\_xts,bse,data2,data2\_xts,lr\_xts,sr\_xts,sr,lr,data,bonds\_xts)
64. #THIS FUNCTION TURNS THE SHORT RUN AND LONG RUN YIELDS INTO A SPREAD WHICH CHARACTERISES THE SLOPE OF THE YIELD CURVE.
65. bseND$term<-bseND$lr-bseND$sr
67. #SOME BASIC PLOTS
68. plot(index(bseND),bseND$bse,type="l")
69. plot(index(bseND),bseND$sr,type="l")
70. plot(index(bseND),bseND$lr,type="l")
71. plot(index(bseND),bseND$term,type="l")
73. #STEP 1
74. #RUNNING UNIT ROOT TESTS ON THE TREND
75. adfbse<-ur.df(bseND$bse,type="trend",selectlags="AIC")
76. adfsr<-ur.df(bseND$sr,type="trend",selectlags="AIC")
77. adflr<-ur.df(bseND$lr,type="trend",selectlags="AIC")
78. adfterm<-ur.df(bseND$term,type="trend",selectlags="AIC")
80. #ADF for bse
81. summary(adfbse)
82. #ADF for sr
83. summary(adfsr)
84. #ADF for lr
85. summary(adflr)
86. #ADF for term
87. summary(adfterm)
88. #Results show that all are non-stationary at the 5% level except term structure, which is stationary at the 5% level.
90. #STEP 2
91. #RUNNING A REGRESSION ON THE TREND
92. Dbse <- na.omit(diff(bseND$bse, differences=1))
93. Dsr <- na.omit(diff(bseND$sr, differences=1))
94. Dlr <- na.omit(diff(bseND$lr, differences=1))
95. Dterm <- na.omit(diff(bseND$term, differences=1))
97. Lbse <- na.omit(stats::lag(bseND$bse, k=1))
98. Lsr <- na.omit(stats::lag(bseND$sr, k=1))
99. Llr <- na.omit(stats::lag(bseND$lr, k=1))
100. Lterm <- na.omit(stats::lag(bseND$term, k=1))
102. LDbse <- na.omit(stats::lag(Dbse, k=1))
103. LDsr <- na.omit(stats::lag(Dsr, k=1))
104. LDlr <- na.omit(stats::lag(Dlr, k=1))
105. LDterm <- na.omit(stats::lag(Dterm, k=1))
107. ttdata<-cbind(Dbse,Dsr,Dlr,Dterm,Lbse,Lsr,Llr,Lterm,LDbse,LDsr,LDlr,LDterm)
108. ttdata<-na.omit(ttdata)
110. adfbse<-lm(ttdata$bse ~ c(1:length(ttdata$bse))+ttdata$bse.2)
111. adfsr<-lm(ttdata$sr ~ c(1:length(ttdata$sr))+ttdata$sr.2)
112. adflr<-lm(ttdata$lr ~ c(1:length(ttdata$lr))+ttdata$lr.2)
113. adfterm<-lm(ttdata$term ~ c(1:length(ttdata$term))+ttdata$term.2)
115. #ADF for bse
116. summary(adfbse)
117. #ADF for sr
118. summary(adfsr)
119. #ADF for lr
120. summary(adflr)
121. #ADF for term
122. summary(adfterm)
123. #The results show that the time trend cannot be rejected at the 1% level for the variables - this is necessary as financial time series data
124. #has been empirically proven to be DS and therefore needs strong evidence otherwise
126. #STEP 3
127. adfbse <- ur.df(bseND$bse, type="drift", selectlags="AIC")
128. adfsr <- ur.df(bseND$sr, type="drift", selectlags="AIC")
129. adflr <- ur.df(bseND$lr, type="drift", selectlags="AIC")
130. adfterm <- ur.df(bseND$term, type="drift", selectlags="AIC")
131. #ADF for bse
132. summary(adfbse)
133. #ADF for sr
134. summary(adfsr)
135. #ADF for lr
136. summary(adflr)
137. #ADF for term
138. summary(adfterm)
139. #Non-stationary for all variables at 5% except for term variable.
141. #STEP 4
142. adfbse <- ur.df(bseND$bse, type="none")
143. adfsr <- ur.df(bseND$sr, type="none")
144. adflr <- ur.df(bseND$lr, type="none")
145. adfterm <- ur.df(bseND$term, type="none")
146. #ADF for bse
147. summary(adfbse)
148. #ADF for sr
149. summary(adfsr)
150. #ADF for lr
151. summary(adflr)
152. #ADF for term
153. summary(adfterm)
154. #It is not possible to reject any of the variables at the 1% level in this situation, therefore it is concluded that, both relevant variables
155. #bse and term are difference stationary. Individually, the short run and long run yields are difference stationary, which would support
156. #the spread to being difference stationary as well, even though it showed less evidence of such compared to the two individually.

159. adfbse <- ur.df(ttdata$bse, type="drift", selectlags="AIC")
160. adfterm <- ur.df(ttdata$term, type="drift", selectlags="AIC")
161. #ADF for First difference bse
162. summary(adfbse)
163. #ADF for First different term
164. summary(adfterm)
165. #Based on the conclusions seen in this test - we can reject the null for a unit root and conclude difference stationarity with an
166. #integrated order of 1.
168. #CALCULATING THE RETURNS FOR THE VARIABLES
169. #Take logs for percentage form
170. bseND$rbse <- diff(log(bseND$bse))
171. bseND$rsr <- log(bseND$sr)
172. bseND$rlr <- log(bseND$lr)
173. bseND$rterm <- diff(bseND$rlr-bseND$rsr)
174. rbseND <- na.omit(cbind(bseND$rbse,bseND$rsr,bseND$rlr,bseND$rterm))
175. vardata <- cbind(bseND$rbse, bseND$rsr, bseND$rlr, bseND$rterm)
176. vardata <- na.omit(vardata)
178. plot(bseND$rbse, col="red")
179. lines(bseND$rterm, col="blue")
181. rm(adfbse,adflr,adfsr,adfterm,Dbse,Dlr,Dsr,Dterm,Lbse,LDbse,LDlr,LDsr,LDterm,Llr,Lsr,Lterm,ttdata)
183. #So far we have concluded that the original data are I(1), and then I(0) as a result of taking their differences. This makes the returns
184. #variables suitable for VAR modeling, however before this, it would be worthwhile to check if there exists some cointegrating relationship
185. #between the the variables, before settling on a model. Do the series' share a common stochastic drift?
187. #Lag selection
188. eg1var <- na.omit(cbind(log(bseND$bse),bseND$rterm))
189. VARselect(eg1var, lag.max=24, type = "both")
190. #Using the level variables the number of lags that minimizes the AIC using the VARselect method is 2, using a maximum lag of 24 months
191. #and including both a trend and drift which are evident in the series' plots.
193. #Engle Granger
194. #Before we consider any further cointegration analysis - we should consider doing a 2-step EG procedure, this will identify any possible
195. #cointegrating relation
196. eg1var <- na.omit(cbind(log(bseND$bse),bseND$rlr-bseND$rsr))
197. colnames(eg1var)[2] <- "rterm"
198. ?rename
199. #Engle Granger Step 1
200. Step1 <- lm(eg1var$bse ~ eg1var$rterm + c(1:length(eg1var$bse)))
201. summary(Step1)
202. #Are the residuals stationary or not?
203. predres <- Step1$residuals
204. ADFres <- ur.df(predres, type="none", selectlags="AIC")
205. summary(ADFres)
206. #using T^{0.25}=132^{0.25}=3 lags approx
207. ADFres <- ur.df(predres, type="none", lags=3)
208. summary(ADFres)
209. #Reject the null at all levels - conclude that the residuals do not contain a unit root and are stationary, and also implies there may
210. #be some cointegrating relationship between the BSE SENSEX and the slope of the yield curve.
212. #Engle Granger Step 2
213. Step2 <- lm(rbseND$rbse ~ rbseND$rterm + stats::lag(rbseND$rterm) + stats::lag(rbseND$rbse)+
214. stats::lag(rbseND$rterm, k=2) + stats::lag(rbseND$rbse, k=2)+ na.omit(stats::lag(predres)))
215. summary (Step2)
216. AIC(Step2)
217. #Drop the second lag of term
218. Step2 <- lm(rbseND$rbse ~ rbseND$rterm + stats::lag(rbseND$rterm) + stats::lag(rbseND$rbse)+
219. stats::lag(rbseND$rbse, k=2)+ na.omit(stats::lag(predres)))
220. summary (Step2)
221. AIC(Step2)
222. #Drop the second lag of bse
223. Step2 <- lm(rbseND$rbse ~ rbseND$rterm + stats::lag(rbseND$rterm) + stats::lag(rbseND$rbse)+
224. na.omit(stats::lag(predres)))
225. summary (Step2)
226. AIC(Step2)
227. #Drop the first lag of rterm
228. Step2 <- lm(rbseND$rbse ~ rbseND$rterm + stats::lag(rbseND$rbse)+
229. + na.omit(stats::lag(predres)))
230. summary (Step2)
231. AIC(Step2)
232. #Drop the first lag of bse
233. Step2 <- lm(rbseND$rbse ~ rbseND$rterm + na.omit(stats::lag(predres)))
234. summary (Step2)
235. AIC(Step2)
236. #All variables are now significant - and the ecm term has interpretation