Times Series Analysis - HW3

Tom Davidson

14/10/2018

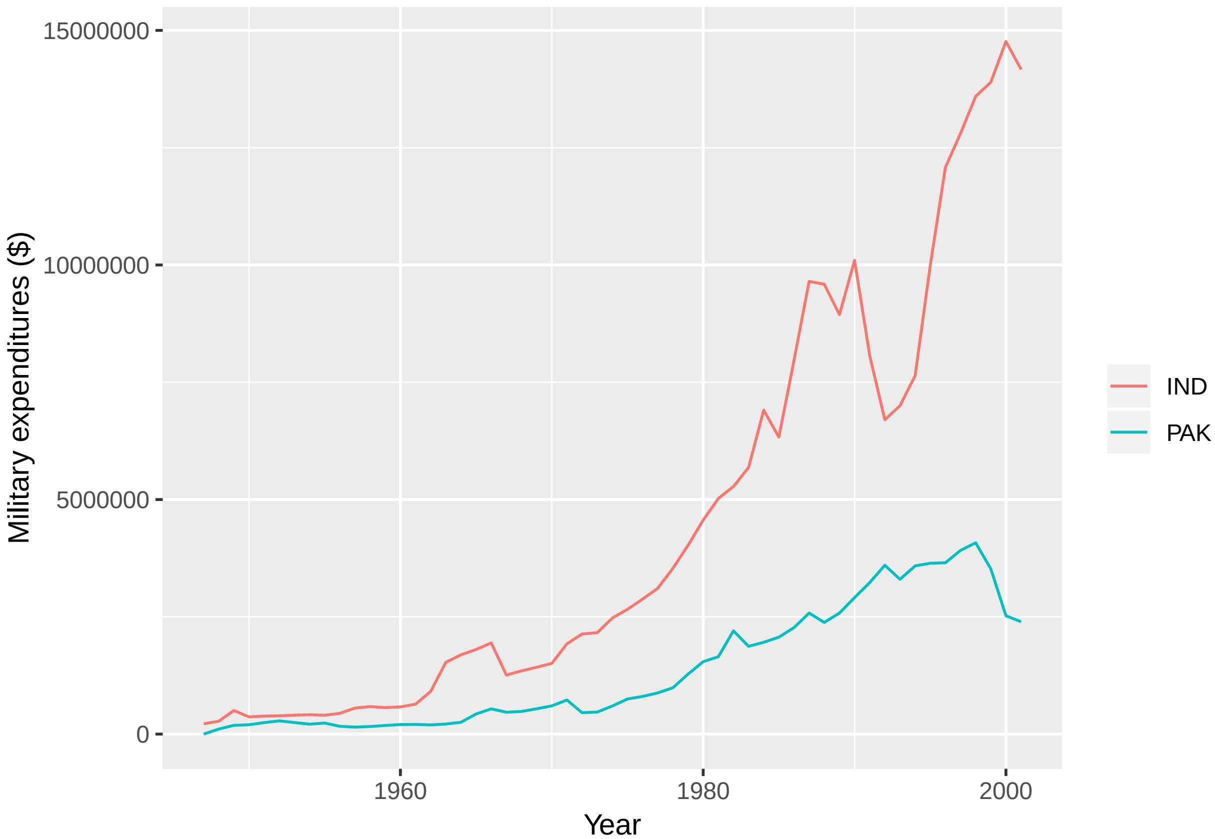
Source code can be found here: "https://github.com/t-davidson/time-series-analysis-fall2018/tree/master/HW3"

```
require("foreign")
require("tseries")
require("zoo")
require("urca")
require("knitr")
require("vars")
require("ggplot2")
require("dplyr")
require("tidyr")
```

```
data <- read.dta('data/NMC_data/NMC_5_0/NMC_5_0.dta')
data <- data[data$year <= 2001, ]
ind <- data[data$stateabb == 'IND',]
pak <- data[data$stateabb == 'PAK',]
ind_ex <- zoo(ind$milex, ind$year)
pak_ex <- zoo(pak$milex, pak$year)
```

Step 1: Plot the two series

```
data2 <- data[data$stateabb %in% c("IND", "PAK"), ]
ggplot(data = data2, aes(x = year, y = milex, group = stateabb, color = stateabb)) +
    geom_line() + xlab("Year") + ylab("Military expenditures ($)") + labs(color = "")
```



Step 2: Testing for evidence of a unit root in either series

```r
summary(ur.df(ind_ex, lags = 1))
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2599605   -24892    41779   281916  1982588
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## z.lag.1      0.0279     0.0187    1.50     0.14
## z.diff.lag   0.2481     0.1497    1.66     0.10
##
## Residual standard error: 707000 on 51 degrees of freedom
## Multiple R-squared:  0.167,  Adjusted R-squared:  0.134
```

```
## F-statistic:  5.1 on 2 and 51 DF,  p-value: 0.0096
##
##
## Value of test-statistic is: 1.496
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

```r
summary(ur.df(pak_ex, lags = 1))
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -861780   -7937   41859  127580  524048
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## z.lag.1    0.000909   0.017452    0.05    0.959
## z.diff.lag 0.256615   0.138125    1.86    0.069 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 234000 on 51 degrees of freedom
## Multiple R-squared:  0.0665, Adjusted R-squared:  0.0299
## F-statistic: 1.82 on 2 and 51 DF,  p-value: 0.173
##
##
## Value of test-statistic is: 0.0521
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

The null hypothesis in the ADF test is that there is a unit root present in the series. In both cases we are unable to reject the null hypothesis that there is a unit root. We must assume, therefore, that Indian and Pakistani defense spending are both I(1) processes.

Step 3: Estimating the cointegration equation

In this case I estimate the equation with Indian defense spending as the dependent variable.

```r
ind_coint <- summary(lm(ind_ex ~ pak_ex))
print(ind_coint)
```

```
##
## Call:
## lm(formula = ind_ex ~ pak_ex)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4740575  -477874  -185360   474453  6682592
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 225916.698 349965.469    0.65     0.52
## pak_ex           3.115      0.187   16.66   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1780000 on 53 degrees of freedom
## Multiple R-squared:  0.84,   Adjusted R-squared:  0.837
## F-statistic:  277 on 1 and 53 DF,  p-value: <2e-16
```

Step 4: Testing the residuals for stationarity

```r
ind_resid <- resid(ind_coint)
summary(ur.df(ind_resid, lags = 1))
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2885218  -242878   -28100   221573  3746112
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -0.248      0.111   -2.25    0.029 *
## z.diff.lag    0.449      0.163    2.76    0.008 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 992000 on 51 degrees of freedom
## Multiple R-squared:  0.138,  Adjusted R-squared:  0.104
## F-statistic: 4.08 on 2 and 51 DF,  p-value: 0.0227
##
##
## Value of test-statistic is: -2.245
##
```

```
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

The Augmented Dickey-Fuller test statistic exceeds the critical threshold at the 5% significance level, although not the 1% level. This means that we can reject the null hypothesis that the residuals have a unit root. Based on this finding we can conclude that this series is stationary and that the variables are cointegrated (BFHP p.161).

Step 6: Evaluating correct lag length for VAR model

```r
ind_ex.diff <- diff(ind_ex)
pak_ex.diff <- diff(pak_ex)
ind_resid.lag <- lag(ind_resid)
data <- cbind(ind_ex.diff, pak_ex.diff, ind_resid.lag)
data <- na.omit(data)
lag_info <- VARselect(cbind(data$ind_ex.diff, data$pak_ex.diff), exogen = data$ind_resid.lag,
    lag.max = 4, type = "both")
print(lag_info$selection)
```

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      3      1      1      3
```

```r
lag_info$criteria
```

```
##                1         2         3         4
## AIC(n) 5.180e+01 5.187e+01 5.179e+01 5.187e+01
## HQ(n)  5.194e+01 5.207e+01 5.205e+01 5.219e+01
## SC(n)  5.218e+01 5.240e+01 5.248e+01 5.271e+01
## FPE(n) 3.133e+22 3.374e+22 3.131e+22 3.408e+22
```

Based on all information criteria tested (although the command in R tests fewer than in Stata) it appears that 4 lags are prefered by 3/4 of the different criteria. In the BFHP analysis they only use 3 lags. I will use 4 lags in the analysis below.

Step 7: Estimating zero lag ECM

Here I aim to reproduce table 6.6 from BFHP:

```r
summary(lm(ind_ex.diff ~ ind_resid[2:55]))
```

```
##
## Call:
## lm(formula = ind_ex.diff ~ ind_resid[2:55])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2119182  -209894  -126172   222244  2216131
##
## Coefficients:
##                    Estimate  Std. Error t value Pr(>|t|)
## (Intercept)     258292.0043  96197.3477    2.69   0.0097 **
## ind_resid[2:55]      0.0762      0.0546    1.40   0.1682
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 707000 on 52 degrees of freedom
## Multiple R-squared:  0.0362, Adjusted R-squared:  0.0177
## F-statistic: 1.95 on 1 and 52 DF,  p-value: 0.168
```

```r
summary(lm(pak_ex.diff ~ ind_resid[2:55]))
```

```
##
## Call:
## lm(formula = pak_ex.diff ~ ind_resid[2:55])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -594767  -60466    1086  104985  379177
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     44361.4233 26929.1249    1.65     0.11
## ind_resid[2:55]    -0.0714     0.0153   -4.68 0.000021 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 198000 on 52 degrees of freedom
## Multiple R-squared:  0.296,  Adjusted R-squared:  0.282
## F-statistic: 21.9 on 1 and 52 DF,  p-value: 0.0000211
```

Compared to the results in Table 6.6 in BFHP, where they find that Indian defense spending appears to be tracking that of Pakistan, we see that the error correction term is only statistically significant in the second regression. This suggests instead that Pakistani defense spending is moving in equilibrium with that of India.

Step 8: Running a VAR model to predict differenced Indian and Pakistani defense spending

```r
var.model <- VAR(y = cbind(data$ind_ex.diff, data$pak_ex.diff), p = 4, exogen = data$ind_resid.lag)
summary(var.model)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: data.ind_ex.diff, data.pak_ex.diff
## Deterministic variables: const
## Sample size: 50
## Log Likelihood: -1417.489
## Roots of the characteristic polynomial:
## 1.32 0.716 0.716 0.669 0.669 0.539 0.539 0.455
## Call:
## VAR(y = cbind(data$ind_ex.diff, data$pak_ex.diff), p = 4, exogen = data$ind_resid.lag)
##
##
## Estimation results for equation data.ind_ex.diff:
## =================================================
## data.ind_ex.diff = data.ind_ex.diff.l1 + data.pak_ex.diff.l1 + data.ind_ex.diff.l2 + data.pak_ex.dif
##
```

```
##                        Estimate  Std. Error t value Pr(>|t|)
## data.ind_ex.diff.l1      0.4497      0.1320    3.41  0.00151 **
## data.pak_ex.diff.l1     -1.4953      0.5528   -2.70  0.00999 **
## data.ind_ex.diff.l2      0.2316      0.1495    1.55  0.12912
## data.pak_ex.diff.l2     -1.0498      0.6164   -1.70  0.09630 .
## data.ind_ex.diff.l3      0.4328      0.1480    2.92  0.00567 **
## data.pak_ex.diff.l3     -0.5710      0.5936   -0.96  0.34190
## data.ind_ex.diff.l4     -0.0112      0.1642   -0.07  0.94588
## data.pak_ex.diff.l4      0.3454      0.5644    0.61  0.54400
## const               76978.6225 126132.1820    0.61  0.54511
## exogen                  -0.4856      0.1301   -3.73  0.00059 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 609000 on 40 degrees of freedom
## Multiple R-Squared: 0.444,   Adjusted R-squared: 0.319
## F-statistic: 3.55 on 9 and 40 DF,  p-value: 0.00258
##
##
## Estimation results for equation data.pak_ex.diff:
## ====================================================
## data.pak_ex.diff = data.ind_ex.diff.l1 + data.pak_ex.diff.l1 + data.ind_ex.diff.l2 + data.pak_ex.dif
##
##                       Estimate Std. Error t value Pr(>|t|)
## data.ind_ex.diff.l1    -0.0171     0.0535   -0.32    0.750
## data.pak_ex.diff.l1     0.4243     0.2238    1.90    0.065 .
## data.ind_ex.diff.l2    -0.0104     0.0605   -0.17    0.865
## data.pak_ex.diff.l2     0.1731     0.2496    0.69    0.492
## data.ind_ex.diff.l3    -0.0749     0.0599   -1.25    0.219
## data.pak_ex.diff.l3     0.1064     0.2403    0.44    0.660
## data.ind_ex.diff.l4    -0.1191     0.0665   -1.79    0.081 .
## data.pak_ex.diff.l4     0.1026     0.2285    0.45    0.656
## const               60305.3122 51070.4497    1.18    0.245
## exogen                  0.0685     0.0527    1.30    0.201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 247000 on 40 degrees of freedom
## Multiple R-Squared: 0.157,   Adjusted R-squared: -0.0329
## F-statistic: 0.826 on 9 and 40 DF,  p-value: 0.596
##
##
##
## Covariance matrix of residuals:
##                  data.ind_ex.diff data.pak_ex.diff
## data.ind_ex.diff     371126147447      -5362574624
## data.pak_ex.diff      -5362574624      60842787223
##
## Correlation matrix of residuals:
##                  data.ind_ex.diff data.pak_ex.diff
## data.ind_ex.diff           1.0000          -0.0357
## data.pak_ex.diff          -0.0357           1.0000
```

In this VAR model predicting the first-difference in Indian and Pakistani defense spending we see that the error correction terms are statistically significant in both cases, althougn their signs are opposite. For India, the EC coefficient of 0.22 is much smaller than that observed in BFHP Table 6.7, suggesting that India still increases spending in response to change in Pakistani spending, but to a lesser degree than in earlier periods. Whereas the BFHP coefficient for the EC in the Pakistan equation was not significant, with a value of -0.15, here we do observe a significant coefficient of -0.11. This suggests that Pakistan decreases its defense spending in response to changes in Indian defense spending.

Step 8: Granger-causality testin

```r
causality(var.model, cause = "data.ind_ex.diff")$Granger
```

```
##
##  Granger causality H0: data.ind_ex.diff do not Granger-cause
##  data.pak_ex.diff
##
## data:  VAR object var.model
## F-Test = 1.2, df1 = 4, df2 = 80, p-value = 0.3
```

```r
causality(var.model, cause = "data.pak_ex.diff")$Granger
```

```
##
##  Granger causality H0: data.pak_ex.diff do not Granger-cause
##  data.ind_ex.diff
##
## data:  VAR object var.model
## F-Test = 2.6, df1 = 4, df2 = 80, p-value = 0.04
```

The results of a Granger causality test show that changes in Indian defense spending Granger-cause changes in Pakistani defense spending ($p = 0.04$) but that the reverse is not true, at least at conventional levels of statistical significance. In the BFHP results we see that both series Granger-cause each other. This suggests that when we consider recent years that Pakistan is responding to India's changes but that India is not paying as much attention to Pakistan. Since Pakistan is the smaller of the two nations this makes sense; it has to keep up with changes in the rate of Indian defense spending but cannot keep up in absolute terms. India is large enough that it can just increase defense spending without too much regard for changes in Pakistan's spending, since it already has a far larger military.