Regressional Analysis on Bank Assets & Demographics (2013-2015)

CSBS Analytics and Research

Jan 11, 2017

Codes below sets the analysis environment and import data from Oracle:

```r
library(car)
library(dplyr)
library(lattice)
setwd('H:/Examiner Resource Model')
sdi <- read.csv('SDI_ERM.csv',stringsAsFactors = F)
pf <- read.csv('pf_erm.csv',stringsAsFactors = F)
```

A glimpse to these two datasets:
SDI

```
##   YEAR    REPDTE   STATE N_SDI_BANK SDI_TOTAL_ASSET
## 1 2015 31-DEC-15 Alabama        115     13900031263
## 2 2014 31-DEC-14 Alabama        118     13097044871
## 3 2013 31-DEC-13 Alabama        121     12453160580
## 4 2015 31-DEC-15  Alaska          4       154909938
## 5 2014 31-DEC-14  Alaska          4       148037750
## 6 2013 31-DEC-13  Alaska          4       132879682
```

Profile Survey

```
##   YEAR   STATE N_PF_EXAMINER N_PF_BANK PF_TOTAL_ASSET
## 1 2015 Alabama            65       115    244501000000
## 2 2014 Alabama            64       117    229776000000
## 3 2013 Alabama            60       116       217647821
## 4 2015  Alaska             4        NA              NA
## 5 2014  Alaska             3        NA              NA
## 6 2013  Alaska             3        NA              NA
```

Remove unused columns and combined the data by Year and State. (*Note: 2014 Data for some States in Profile Survey has been removed due to mistakes in data collection*).

```r
sdi <- sdi[,-2]
pf <- pf[,1:3]
rdb <- dplyr::full_join(pf,sdi,by=c("YEAR","STATE"))
```

Here is a preview for the combined data and Plot Matrix with density plots and trend lines to explore correlations among variables:

```r
scatterplotMatrix(~N_PF_EXAMINER+N_SDI_BANK+SDI_TOTAL_ASSET|YEAR,
                  data=rdb, main="Correlation Grouped in Years")
```

Correlation Grouped in Years
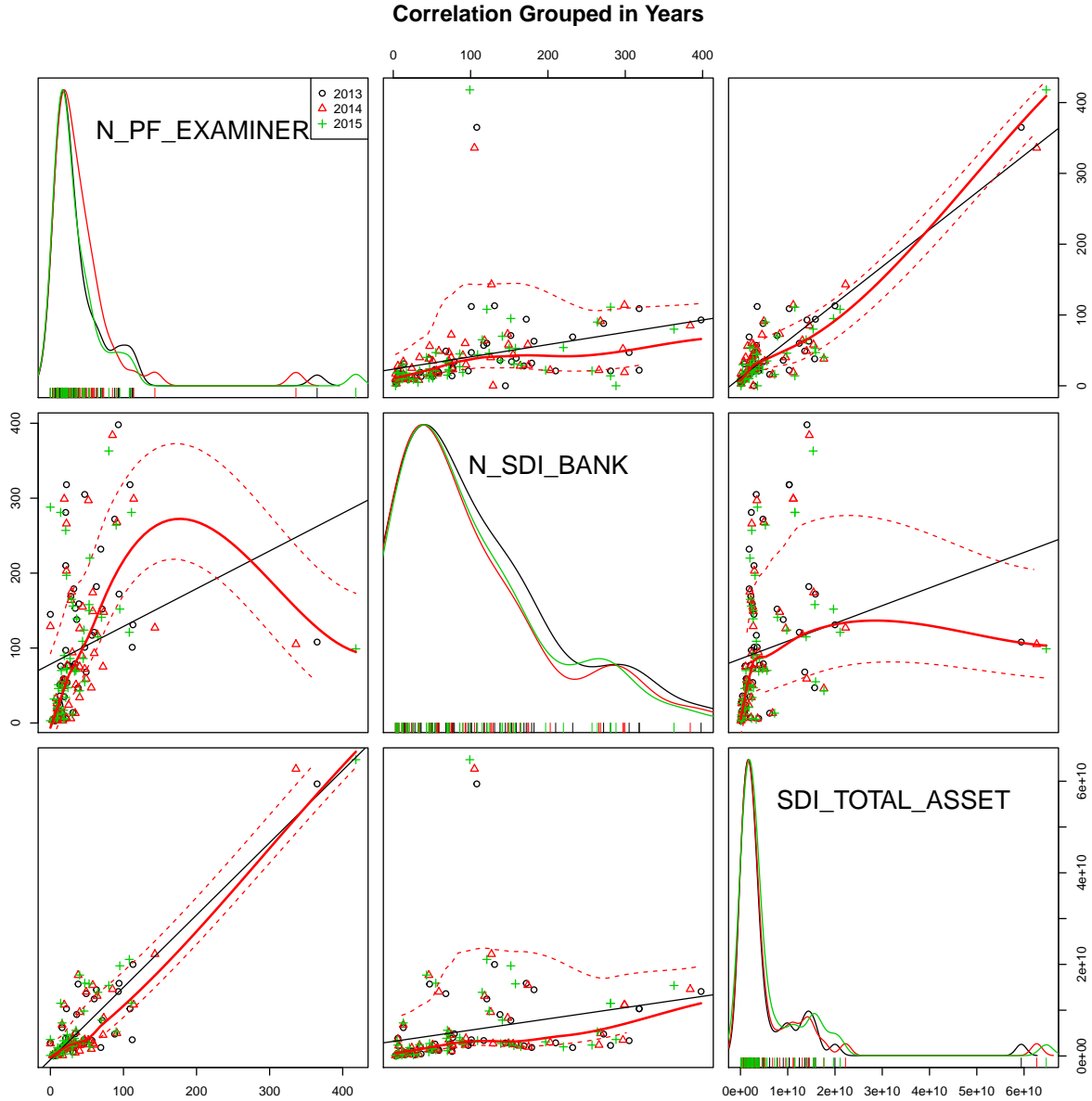
The plot matrix above indicates the correlation among *Number of Examiner*, *Count* and *Total Asset* for State Chartered Banks. They are banks classified as **SM**, **NM** and **SB**, or **OI** and **SA** supervised by FDIC. The Bank Charter Class is a classification code assigned by the FDIC based on the institution's charter type (commercial bank or savings institution), charter agent (state or federal), Federal Reserve membership status (Fed member, Fed nonmember) and its primary federal regulator (state chartered institutions are subject to both federal and state supervision).

* **N** = commercial bank, national (federal) charter and Fed member, supervised by the Office of the Comptroller of the Currency (OCC)

* **SM** = commercial or savings bank, state charter and Fed member, supervised by the Federal Reserve (FRB)

* **NM** = commercial bank, state charter and Fed nonmember, supervised by the FDIC or OCC

* **SB** = savings banks, state charter, supervised by the FDIC

* **SA** = As of July 21, 2011, FDIC supervised state chartered thrifts and OCC supervised federally chartered thrifts. Prior to that date, state or federally chartered savings associations supervised by the Office of Thrift Supervision (OTS).

* **OI** = insured U.S. branch of a foreign chartered institution (IBA)

Constructing Linear Model using simple linear regression model defined as

$$y_i = \alpha + \beta_i x_i + \epsilon_i$$

```
lm1 <- lm(N_PF_EXAMINER~SDI_TOTAL_ASSET+N_SDI_BANK+YEAR,rdb)
summary(lm1)
```

```
##
## Call:
## lm(formula = N_PF_EXAMINER ~ SDI_TOTAL_ASSET + N_SDI_BANK + YEAR,
##     data = rdb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.080  -5.670  -0.409   7.070  79.714
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.561e+03  4.398e+03   1.037   0.3014
## SDI_TOTAL_ASSET  5.133e-09  1.954e-10  26.274   <2e-16 ***
## N_SDI_BANK       4.674e-02  1.992e-02   2.346   0.0202 *
## YEAR            -2.261e+00  2.184e+00  -1.035   0.3022
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.45 on 153 degrees of freedom
## Multiple R-squared:  0.8345, Adjusted R-squared:  0.8312
## F-statistic: 257.1 on 3 and 153 DF,  p-value: < 2.2e-16
```

The model summary above indicates **YEAR** is not significant in the model and can be safely removed. The number of asterisks represents the degree of significance where *** indicates strong correlation and . indicates weaker correlation. Among all these variables, **Bank Total Asset** (SDI_TOTAL_ASSET) is the most significant one.

```
lm2 <- lm(N_PF_EXAMINER~SDI_TOTAL_ASSET+N_SDI_BANK,rdb)
summary(lm2)
```

```
##
## Call:
## lm(formula = N_PF_EXAMINER ~ SDI_TOTAL_ASSET + N_SDI_BANK, data = rdb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -65.486  -6.339  -0.199   7.924  81.926
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.108e+00  2.692e+00   2.640  0.00914 **
## SDI_TOTAL_ASSET  5.125e-09  1.953e-10  26.247  < 2e-16 ***
## N_SDI_BANK       4.764e-02  1.990e-02   2.394  0.01789 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.45 on 154 degrees of freedom
```

```
## Multiple R-squared:  0.8333, Adjusted R-squared:  0.8311
## F-statistic: 384.9 on 2 and 154 DF,  p-value: < 2.2e-16
```

The second linear model excludes YEAR from the model. The significance of SDI_TOTAL_ASSET remains. Thus, the Number of Examiner for a State Department can be approximated with the following linear equation:

$$Number\ of\ Examiner = 5.125^{-9} * Total\ Asset + 4.764^{-2} * Number\ of\ Bank + 7.108$$

FYI, The plot below shows how the linear model is fit.

```
par(mfrow=c(2,2))
plot(lm2)
```