

VARX

Sri Santhosh Hari

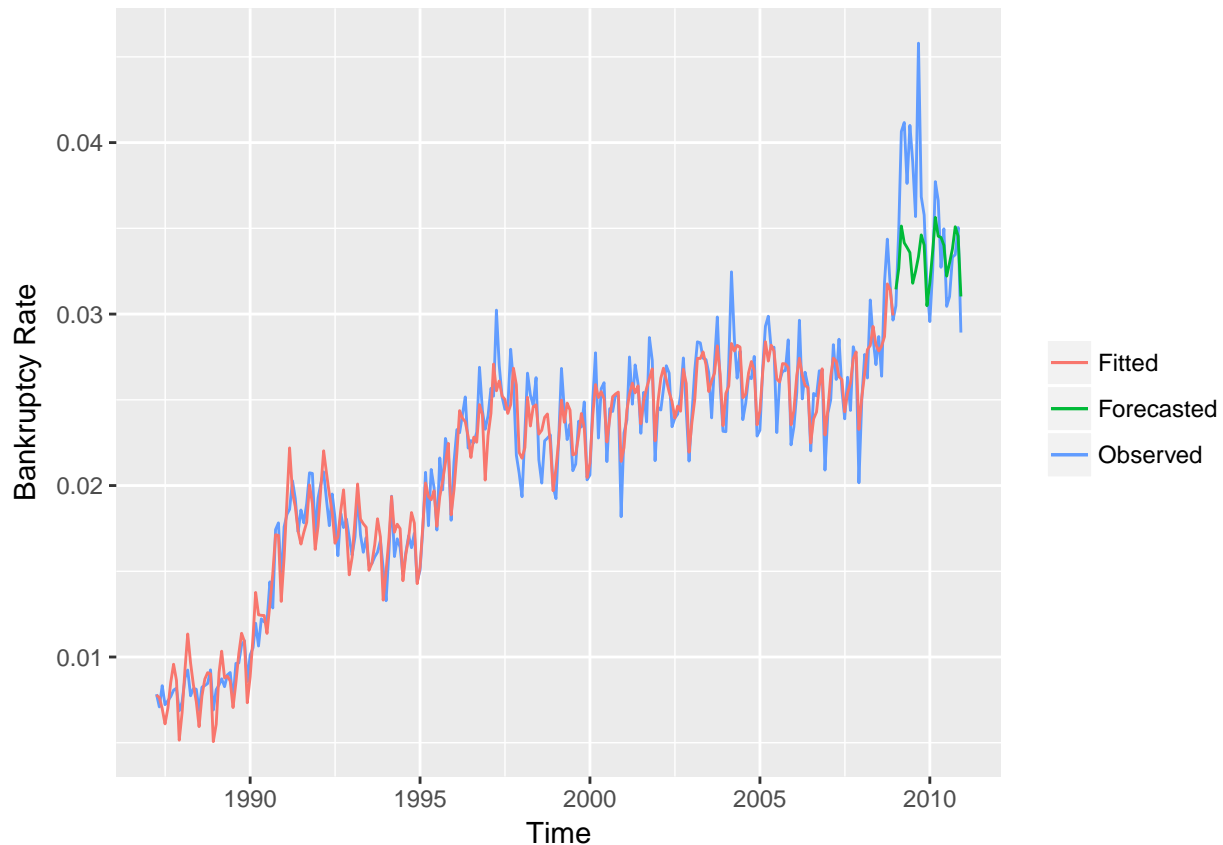
```
train_data <- read_csv("Data/train.csv")
train_data <- na.omit(train_data)
test <- read_csv("Data/test.csv")
test <- na.omit(test)
train_data$Mon <- seq.Date(as.Date("1987/1/1"), as.Date("2010/12/1"), by = "month")
test$Mon <- seq.Date(as.Date("2011/1/1"), as.Date("2012/12/1"), by = "month")
train <- train_data[1:264,]
valid <- train_data[265:288,]

VARselect(train[c("Bankruptcy_Rate", "Unemployment_Rate", "House_Price_Index")],
          lag.max = 12, season=12, exogen=train["Population"])

## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      5      3      3      5
##
## $criteria
##              1              2              3              4
## AIC(n) -1.813490e+01 -1.879141e+01 -1.903330e+01 -1.900373e+01
## HQ(n)  -1.786439e+01 -1.847019e+01 -1.866135e+01 -1.858106e+01
## SC(n)  -1.746262e+01 -1.799309e+01 -1.810893e+01 -1.795331e+01
## FPE(n)  1.331490e-08  6.908214e-09  5.426549e-09  5.592900e-09
##              5              6              7              8
## AIC(n) -1.903542e+01 -1.902085e+01 -1.897480e+01 -1.893853e+01
## HQ(n)  -1.856203e+01 -1.849674e+01 -1.839997e+01 -1.831298e+01
## SC(n)  -1.785894e+01 -1.771832e+01 -1.754622e+01 -1.738390e+01
## FPE(n)  5.422782e-09  5.507845e-09  5.774384e-09  5.996314e-09
##              9              10             11             12
## AIC(n) -1.892469e+01 -1.894178e+01 -1.894032e+01 -1.891337e+01
## HQ(n)  -1.824842e+01 -1.821479e+01 -1.816261e+01 -1.808494e+01
## SC(n)  -1.724401e+01 -1.713505e+01 -1.700754e+01 -1.685453e+01
## FPE(n)  6.090286e-09  5.999019e-09  6.021620e-09  6.202449e-09

mod_var <- VAR(train[c("Bankruptcy_Rate", "Unemployment_Rate", "House_Price_Index")],
              p=3, season=12, exogen = train["Population"])
fit_var <- data.frame(fitted(mod_var))
valid_var <- predict(mod_var, n.ahead=24, ci=0.95, dumvar = valid["Population"])
valid_pred <- data.frame(valid_var$fcst$Bankruptcy_Rate)

train_data %>%
  filter(!(Month %in% c(11987,21987,31987))) %>%
  ggplot()+
  geom_line(aes(x=Mon, y=Bankruptcy_Rate,color="Observed"))+
  geom_line(data=train[4:264,], aes(x=Mon, y=fit_var$Bankruptcy_Rate,color="Fitted"))+
  geom_line(data=valid, aes(x=Mon, y=valid_pred$fcst,color="Forecasted"))+
  labs(color='', x="Time", y="Bankruptcy Rate")
```



```
sqrt(mean((valid_pred$fcst - valid$Bankruptcy_Rate)^2))
```

```
## [1] 0.004148911
```

Retrain Model:

```
retrain_mod_var <- VAR(train_data[c("Bankruptcy_Rate", "Unemployment_Rate", "House_Price_Index")],
  p=3, season=12, exogen = train_data["Population"])
refit_var <- data.frame(fitted(retrain_mod_var))
test_pred <- data.frame(predict(retrain_mod_var, n.ahead=24, pi=0.95,
  dumvar = test["Population"])$fcst$Bankruptcy_Rate)
```

```
train_data %>%
  filter(!(Month %in% c(11987, 21987, 31987))) %>%
  ggplot()+
  geom_line(aes(x=Mon, y=Bankruptcy_Rate, color='Observed'))+
  geom_line(aes(x=Mon, y=refit_var$Bankruptcy_Rate, color='Fitted'))+
  geom_line(data=test_pred, aes(x=test$Mon, y=fcst, color='Forecast'))+
  geom_ribbon(data=test_pred, aes(x=test$Mon, ymin=lower, ymax=upper, alpha=0.5), show.legend = F)+
  labs(color='', y='Bankruptcy Rate', x='Time')
```

