Predicting Food Insecurity with Machine Learning based on readily avaliable data

Yujun Zhou[a,*]

[a] *Department of Agriculture and Consumer Economics, University of Illinois, USA*

Abstract

We determine whether corn and soybean futures contract prices are stationary or not.

Keywords: prices, unit root, stationarity,

```
# Seed for random number generation
set.seed(42)
```

Temporal Poverty Prediction using Satellite Imagery(Chen 2017).

1. measuring and tracking areas of poor is a necessary step for targeting aid and guiding policy decisions.
2. obtaining data is time and labor intensive
3. (Blumenstock 2016)

Import the Malawi data.

```
# import the data
ihs2010<-read.csv("data/cleaned/Malawi/IHS2010.csv")
ihs2013<-read.csv("data/cleaned/Malawi/IHS2013.csv")
```

Organize the variable names and ready for analysis.

```
# levels
levels<-c("ipczone","TA","clust")

# variables
weather<-c("L12raincytot","L12day1rain","L12maxdays","floodmax")
access<-c("lag_price","lag_thinn")
asset1 <-c("roof","cells_own")
land<-c("percent_ag","elevation","nutri_reten_constrained")
distance<-c("dist_road","dist_admarc")
demo<-c("hhsize","hh_age","hh_gender","asset")

model3_variables<-c(weather,access,asset1,land,distance,demo)
model2_variables<-c(weather,access,asset1,land,distance)
model1_variables<-c(weather,access,land,distance)
```

*Corresponding Author
Email address: zhou100@illinois.edu (Yujun Zhou)

```
# goal : combine variables at different levels using pastes
# output: variables lists at different levels, TA_vars, ipczone vars, etc.

for (level in levels){
  # assign levels of variables group
  group_var_name<-paste(level,"vars",sep="_")
  assign(group_var_name,c())

  for(var in model3_variables){
    temp<-paste(level,var,sep = "_")
    new<-append(get(group_var_name),temp)
    assign(group_var_name,new)
  }
}
```

1. *Linear/tobit Results*

Create the formulas using the formula_compose function.

```
rcsi_formula<-formula_compose("RCSI",clust_vars)
logFCS_formula<-formula_compose("logFCS",clust_vars)
HDDS_formula<-formula_compose("HDDS",clust_vars)
```

```
rcsi_predictions<-linear_fit(rcsi_formula,ihs2010,ihs2013)
# lm_train_measure<-postResample(rcsi_predictions$pred_train,ihs2010$RCSI)
lm_test_measure<-postResample(rcsi_predictions$pred_test,ihs2013$RCSI)
lm_test_measure
```

```
##       RMSE   Rsquared        MAE
## 7.07641265 0.01254163 5.52598695
```

```
# scatter.smooth(rcsi_predictions$pred_test,ihs2013$RCSI)
```

try tobit instead for RCSI. The prediction value (unconditional mean) should actually be different with the assumption of non-normal / Gaussian error. The old predication function returns a latent mean.

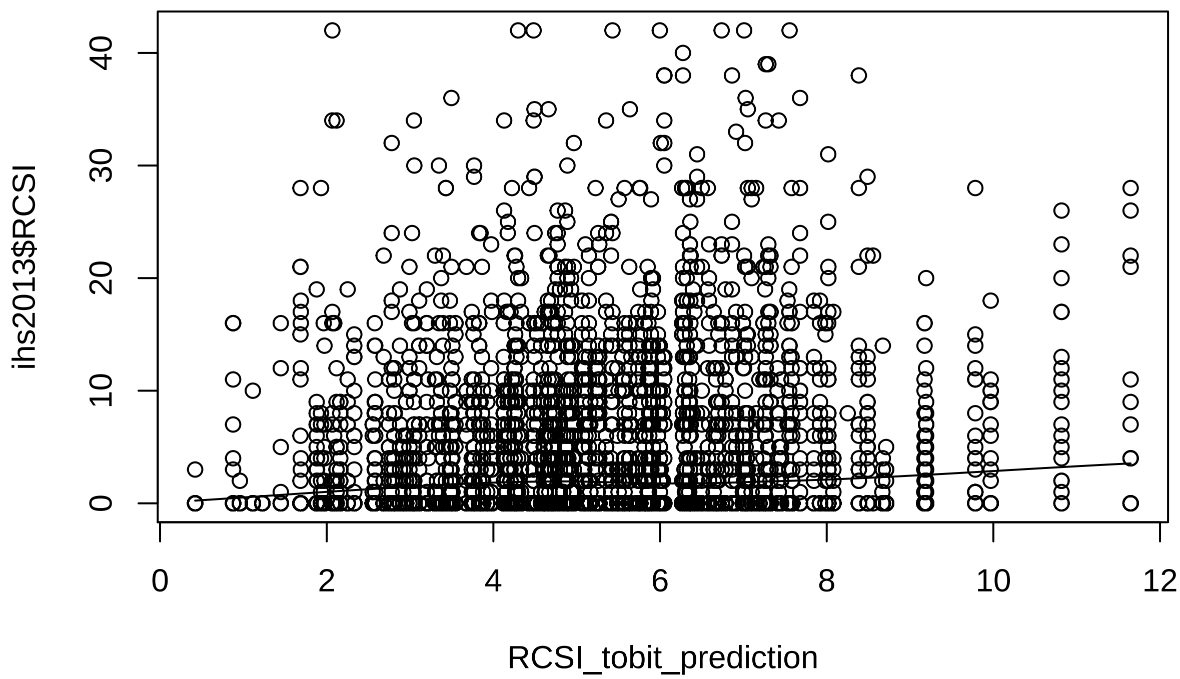
```
tobit_rcsi<-tobit(rcsi_formula,left = 0,right = Inf,data =ihs2010)
mu <- predict(tobit_rcsi,newdata= ihs2013)
sigma <- tobit_rcsi$scale
p0 <- pnorm(mu/sigma)
lambda <- function(x) dnorm(x)/pnorm(x)
ey0 <- mu + sigma * lambda(mu/sigma)
ey <- p0 * ey0
RCSI_tobit_prediction<-ey
```

There are some improvement but still slighlty

```
lm_test_measure<-postResample(RCSI_tobit_prediction,ihs2013$RCSI)
lm_test_measure
```

```
##       RMSE   Rsquared        MAE
## 7.06342825 0.01313111 5.40365933
```

```r
scatter.smooth(RCSI_tobit_prediction,ihs2013$RCSI)
```




```r
logFCS_predictions<-linear_fit(logFCS_formula,ihs2010,ihs2013)
lm_test_measure<-postResample(rcsi_predictions$pred_test,ihs2013$logFCS)
lm_test_measure
```

```
##       RMSE   Rsquared        MAE
## 2.47188996 0.09126124 2.11332725
```

```r
# scatter.smooth(logFCS_predictions$pred_test,ihs2013$logFCS)
```

```r
HDDS_predictions<-linear_fit(HDDS_formula,ihs2010,ihs2013)
lm_test_measure<-postResample(rcsi_predictions$pred_test,ihs2013$HDDS)
lm_test_measure
```

```
##       RMSE   Rsquared        MAE
## 2.38812209 0.08306447 1.84598055
```

```r
# scatter.smooth(HDDS_predictions$pred_test,ihs2013$HDDS)
```

Methods

Participants

Material

Procedure

Data analysis

We used for all our analyses.

Results

Discussion

References

Blumenstock, Joshua Evan. 2016. “Fighting Poverty with Data.” *Science* 353 (6301). American Association for the Advancement of Science: 753–54.

Chen, Derek. 2017. “Temporal Poverty Prediction Using Satellite Imagery.”