Financial Risk Management

Finance 530Q, Spring 2019

Prepared by David Hsieh

Assignment 3: Time-varying covariates and competing risks

Purpose and Overview:

This assignment adds time-varying explanatory variables (or "covariates") to the FNMA mortgages. Since we have time-varying covariates, we must use the long format for our data to estimate both the Cox PH model and the multiperiod logit model. In addition, we will deal with competing risks – default and prepayments.

Part of the objective of this assignment is to merge data from multiple sources. The other main objective is to estimate competing risks models.

Directions:

Step 1: Download external Data

Download the following three R scripts to retrieve data from the web:

Rates.R	# this file retrieves mortgage rates from Freddie Mac
Ue_msa.R	# this file retrieves unemployment rates by MSA
Freddie_hpi.R	# this file retrieves house price indices by MSA

MSA stands for "Metropolitan Statistical Area".

Run these three scripts files to create three data sets:
        rates
        ue_msa
        hpi_msa

Step 2: Create origination data from randomly sampled mortgages

Download TeamAssignment3_cdata_Q1.rda
The Data_C data table contains about 99,000 randomly sampled loans from the 2000 Q1 acquisition file.

Perform the following tasks on Data_C:
- convert the MSA variable from character to integer.
- remove loans whose MSA is not in ue_msa$MSA

- remove loans whose MSA is not in hpi_msa$MSA
- remove loans with no FICO score (CSCORE_B)
- remove loans with no original value (ORIG_VAL)
- remove loans with no origination amount (ORIG_AMT)
- remove loans with no origination interest rate (ORIG_RT)
This should leave you with 77,159 loans

Set the ORIG_DTE variable to be integer in the form YYYYMM.

Merge Data_C and rates[,c(1,2)] using by.x = "ORIG_DTE", by.y = "yearmon", all.x=TRUE

Merge Data_C and hpi_msa using by.x = c("ORIG_DTE","MSA")), by.y = c("yearmon","MSA"), all.x=TRUE
Then rename the "hpi" variable to be "hpi0"

Create the spread variable as ORIG_RT – rate

Keep only the following variables:
"LOAN_ID","OLTV","CSCORE_B","spread","ORIG_VAL","hpi0","MSA",
"ORIG_RT","NUM_BO","PURPOSE","PROP_TYP","OCC_STAT","DTI","FTHB_FLG"

Step 3: Create the loan-month data

Load TeamAssignment3_pdata_Q1.rda
It contains the Data_P data set of the loans in the performance file of 2000 Q1.

Perform the following tasks on Data_P:

- rename the variable "Monthly.Rpt.Prd" as "yearmon"
- convert the "yearmon" variable from character to integer format (YYYYMM)
- keep only the loans with LOAD_ID in Data_C$LOAN_ID
There should be 2,661,798 rows in Data_P

- merge Data_P and Data_C, using by="LOAN_ID", all.x=TRUE
Call the merged data set **data1**
- order data1 first by LOAN_ID, then by yearmon.
- create the status variable using the code:
data1$status <- ifelse(data1$Zero.Bal.Code %in% c("02","03","09","15"),"default",
                ifelse(data1$Zero.Bal.Code %in% c("01"),"prepaid","censored"))

Step 4: create the "cvr" variable.

Merge the "rates" data table into the "data1" data table, using on="yearmon".
The cvr variable is ORIG_RT / rate

Step 5: create the ue variable

Merge the "ue_msa" data table into the "data1" data table, using
on = c("yearmon","MSA")

Step 6: create the pneq variable

Merge the "hpi_msa" data table into the "data1" data table, using
on=c("yearmon", "MSA")

Then create the pneq variable as follows:
Calculate the "val" variable as ORIG_VAL * hpi / hpi0
Calculate the "pneq" variable as pnorm(log(LAST_UPB/val)/(100*spi))

Step 7: Create "start" and "end" as:
        start = Loan.Age
        end = Loan.Age + 1

Estimation

Use coxph to estimate a default model, and a prepayment model.

You must include "CSCORE_B" and "pneq" as one of the covariates. In addition, you can
include any of the following variables as covariates"
:"OLTV","spread","ue","cvr","NUM_BO","PURPOSE","PROP_TYP","OCC_STAT","DTI",
"FTHB_FLG"

Discuss your choice of covariates.

Prediction: Consider the following hypothetical mortgage over 60 months:

CSCORE_B = 720, and pneq = 0. Set all other covariates to be the median of your sample.

Use your estimated default model predict the cumulative default probabilities for each of the 60
months of this mortgage.

Use your estimated prepayment model to predict the cumulative prepayment probabilities for
each of the 60 months of this mortgage.

Submission:

Your submission consists of two files uploaded to Canvas.

File 1: PDF file with 3 powerpoint slides of your findings.
Slide 1: Give the estimates of your default and prepayment models. Explain your choice of covariates.
Slide 2: Graph the cumulative default probabilities in step 8.
Slide 3: Graph the cumulative prepayment probabilities in step 8.

File 2: One R script with all your calculations for the entire assignment.

Each team will present their findings in class.

Notes on DATA.TABLE

The R package, data.table, uses some very fast algorithms to merge data. Here are examples of "right join" and "left join":

```
> DT = data.table(x=rep(c("b","a","c"),each=3), y=c(1,3,6), v=1:9)
> DT
   x y v
1: b 1 1
2: b 3 2
3: b 6 3
4: a 1 4
5: a 3 5
6: a 6 6
7: c 1 7
8: c 3 8
9: c 6 9

> X = data.table(x=c("c","b"), v=8:7, foo=c(4,2))
> X
   x v foo
1: c 8   4
2: b 7   2

> DT[X, on="x"]                    # right join
   x y v i.v foo
1: c 1 7   8   4
2: c 3 8   8   4
3: c 6 9   8   4
4: b 1 1   7   2
5: b 3 2   7   2
6: b 6 3   7   2

> X[DT, on="x"]                    # left join
   x  v foo y i.v
1: b  7   2 1   1
2: b  7   2 3   2
3: b  7   2 6   3
4: a NA  NA 1   4
5: a NA  NA 3   5
6: a NA  NA 6   6
7: c  8   4 1   7
8: c  8   4 3   8
9: c  8   4 6   9
```