

# FIN3210 Week 2 Assignment

Ma Kexuan 120090651

September 28, 2023

```
[1]: import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings('ignore')
import statsmodels.api as sm
from lifelines import CoxPHFitter
```

## 0.0.1 Read the data from the disk

```
[2]: rrd_tot = pd.read_excel('FIN3210 Week 2 Renrendai loans.xlsx', sheet_name='Data_
↳Borrower')
plat_tot = pd.read_excel('FIN3210 Week 2 p2p lending platforms.xlsx',
↳sheet_name='Platform Data')
```

```
[3]: rrd_tot.drop(['IND', 'CITY', 'PURPOSE', 'MARRY', 'title', 'description', 'nickName'],
↳axis = 1, inplace = True)
rrd_tot.head()
```

```
[3]:
```

	loanId	STATUS	BIDS	DEFAULT	AMOUNT	INTEREST	MONTHS	CREDIT	HOUSE	\
0	2	CLOSED	9	0	3000	5.0	6	7	1	
1	11	CLOSED	8	0	3000	18.0	3	3	0	
2	16	CLOSED	8	0	3000	12.0	12	3	0	
3	19	CLOSED	11	0	3000	8.8	12	7	1	
4	20	CLOSED	15	0	5000	15.0	12	7	0	

  

	CAR	HOUSE_L	CAR_L	EDUCATION	WORKTIME	INCOME	AGE	opentime	\
0	1	1	1	3.0	2.0	6.0	33	2010-10-12 17:17:01	
1	0	0	0	3.0	4.0	4.0	37	2010-10-18 16:40:38	
2	0	0	0	3.0	4.0	4.0	37	2010-10-21 17:26:58	
3	1	1	1	3.0	2.0	6.0	33	2010-10-25 17:19:39	
4	1	0	0	3.0	2.0	3.0	33	2010-10-26 14:29:03	

  

	readytime
0	2010-10-15 11:00:49
1	2010-10-21 17:07:41
2	2010-10-28 14:35:07
3	2010-10-28 20:58:15
4	2010-10-28 14:34:33

```
[4]: plat_tot.drop(['fullname', 'Province', 'City', 'Background'], axis = 1, inplace =   
      ↪ True)  
plat_tot.head()
```

```
[4]: OnlineTime_YMD Bankrupt_WDZJ Collapse Benign Fraud RegCapital \  
0      20140519      20170413.0          1    0.0    0.0        500.0  
1      20151231      20170201.0          1    0.0    0.0        500.0  
2      20150504      20161201.0          1    0.0    0.0        500.0  
3      20180310      20180615.0          1    0.0    0.0        500.0  
4      20180320      20180724.0          1    0.0    1.0          5.0  
  
Capitaldeposit Obtaininvest Joinasso Autobid Transright Riskdeposit \  
0              0           0.0         1.0         0           0          0.0  
1              0           0.0         0.0         0           0          0.0  
2              0           0.0         0.0         1           1          0.0  
3              0           0.0         0.0         0           0          0.0  
4              0           0.0         0.0         0           0          0.0  
  
Thirdguarantee  
0              0.0  
1              0.0  
2              0.0  
3              0.0  
4              0.0
```

### 0.0.2 1) Present two tables for the summary statistics of the key variables in Renrendai loans.xlsx and p2p lending platforms.xlsx

Procedures: Data cleaning, preserve the relevant data.

```
[5]: rrd = rrd_tot[['BIDS', 'DEFAULT', 'AMOUNT', 'INTEREST', 'MONTHS', 'CREDIT',  
                  'HOUSE', 'CAR', 'HOUSE_L', 'CAR_L', 'EDUCATION', 'WORKTIME',  
                  'INCOME', 'AGE']]  
rrd.describe().T
```

```
[5]: count mean std min 25% 50% \  
BIDS    10000.0    24.150600    41.342608    1.0    9.0    15.0  
DEFAULT 10000.0     0.151300     0.358359     0.0     0.0     0.0  
AMOUNT  10000.0  24545.835000  38280.756524  3000.0  8000.0  14400.0  
INTEREST 10000.0    12.621900     2.273689     5.0    11.0    12.0  
MONTHS   10000.0    12.237300     8.091090     3.0     6.0    12.0  
CREDIT   10000.0     2.146300     1.530990     1.0     1.0     2.0  
HOUSE    10000.0     0.564500     0.495847     0.0     0.0     1.0  
CAR      10000.0     0.391700     0.488155     0.0     0.0     0.0  
HOUSE_L  10000.0     0.228400     0.419823     0.0     0.0     0.0  
CAR_L    10000.0     0.082200     0.274683     0.0     0.0     0.0  
EDUCATION 9996.0     2.165966     0.818108     1.0     2.0     2.0  
WORKTIME 9994.0     2.838003     0.992755     1.0     2.0     3.0
```

INCOME	9998.0	4.309162	1.335842	1.0	3.0	4.0
AGE	10000.0	34.755500	6.682708	24.0	30.0	33.0

	75%	max
BIDS	24.0	592.0
DEFAULT	0.0	1.0
AMOUNT	26000.0	500000.0
INTEREST	13.0	24.4
MONTHS	12.0	36.0
CREDIT	3.0	7.0
HOUSE	1.0	1.0
CAR	1.0	1.0
HOUSE_L	0.0	1.0
CAR_L	0.0	1.0
EDUCATION	3.0	4.0
WORKTIME	4.0	4.0
INCOME	5.0	7.0
AGE	38.0	53.0

```
[6]: plat = plat_tot[['OnlineTime_YMD', 'Bankrupt_WDZJ', 'Collapse', 'Benign',
                    'Fraud', 'RegCapital', 'Capitaldeposit', 'Obtaininvest',
                    'Joinasso', 'Autobid', 'Transright', 'Riskdeposit', 'Thirdguarantee']]
plat_des = plat.drop(['OnlineTime_YMD', 'Bankrupt_WDZJ'], axis = 1)
plat_des.describe().T
```

```
[6]:
```

	count	mean	std	min	25%	50%	75%	\
Collapse	1000.0	0.782000	0.413094	0.0	1.0	1.0	1.0	
Benign	782.0	0.098465	0.298134	0.0	0.0	0.0	0.0	
Fraud	782.0	0.246803	0.431427	0.0	0.0	0.0	0.0	
RegCapital	1000.0	596.064330	2328.221711	2.0	100.0	300.0	500.0	
Capitaldeposit	1000.0	0.191000	0.393286	0.0	0.0	0.0	0.0	
Obtaininvest	968.0	0.026860	0.161756	0.0	0.0	0.0	0.0	
Joinasso	968.0	0.054752	0.227613	0.0	0.0	0.0	0.0	
Autobid	1000.0	0.244000	0.429708	0.0	0.0	0.0	0.0	
Transright	1000.0	0.177000	0.381860	0.0	0.0	0.0	0.0	
Riskdeposit	968.0	0.021694	0.145758	0.0	0.0	0.0	0.0	
Thirdguarantee	968.0	0.034091	0.181557	0.0	0.0	0.0	0.0	

	max
Collapse	1.0
Benign	1.0
Fraud	1.0
RegCapital	50000.0
Capitaldeposit	1.0
Obtaininvest	1.0
Joinasso	1.0

Autobid	1.0
Transright	1.0
Riskdeposit	1.0
Thirdguarantee	1.0

**0.0.3 2) Perform a logit regression and examine the relation between the default likelihood and borrower characteristics such as credit, house, car, education, work time, etc.**

```
[7]: X = rrd[['CREDIT','HOUSE','CAR','HOUSE_L',
            'CAR_L','EDUCATION','WORKTIME','INCOME','AGE']] # Choose relevant
            ↪ independent variables
y = rrd[['DEFAULT']]
X = sm.add_constant(X)
logit_model = sm.Logit(y, X, missing = 'drop').fit() # Specify the missing
            ↪ values to be dropped from the regression process
logit_model.summary()
```

Optimization terminated successfully.  
Current function value: 0.330132  
Iterations 9

```
[7]: <class 'statsmodels.iolib.summary.Summary'>
"""
```

```

                                Logit Regression Results
=====
Dep. Variable:                  DEFAULT    No. Observations:                  9990
Model:                            Logit    Df Residuals:                      9980
Method:                           MLE     Df Model:                          9
Date:                Thu, 28 Sep 2023    Pseudo R-squ.:                     0.2236
Time:                  21:02:46          Log-Likelihood:                    -3298.0
converged:                        True    LL-Null:                          -4247.9
Covariance Type:            nonrobust    LLR p-value:                       0.000
=====

```

	coef	std err	z	P> z	[0.025	0.975]
const	0.5155	0.212	2.427	0.015	0.099	0.932
CREDIT	-1.8927	0.082	-23.044	0.000	-2.054	-1.732
HOUSE	0.1438	0.073	1.968	0.049	0.001	0.287
CAR	-0.4586	0.080	-5.708	0.000	-0.616	-0.301
HOUSE_L	-0.3307	0.091	-3.633	0.000	-0.509	-0.152
CAR_L	0.1620	0.134	1.207	0.228	-0.101	0.425
EDUCATION	-0.4156	0.040	-10.426	0.000	-0.494	-0.337
WORKTIME	0.0090	0.034	0.264	0.792	-0.058	0.076
INCOME	0.1160	0.025	4.592	0.000	0.066	0.165
AGE	0.0254	0.005	4.936	0.000	0.015	0.036

```
=====
```

```
"""
```

0.0.4 3) Perform an ols regression and examine the relation between the number of bids and borrower characteristics such as credit, house, car, education, work time, etc.

```
[8]: # The same procedure as question 2
X = rrd[['CREDIT', 'HOUSE', 'CAR', 'HOUSE_L',
         'CAR_L', 'EDUCATION', 'WORKTIME', 'INCOME', 'AGE']]
y = rrd[['BIDS']]
X = sm.add_constant(X)
ols_model = sm.OLS(y, X, missing = 'drop').fit()
ols_model.summary()
```

```
[8]: <class 'statsmodels.iolib.summary.Summary'>
```

```
"""
```

#### OLS Regression Results

```
=====
Dep. Variable:          BIDS      R-squared:                0.173
Model:                  OLS      Adj. R-squared:           0.172
Method:                 Least Squares      F-statistic:        232.1
Date:                  Thu, 28 Sep 2023      Prob (F-statistic):    0.00
Time:                  21:02:46      Log-Likelihood:       -50383.
No. Observations:      9990      AIC:                  1.008e+05
Df Residuals:          9980      BIC:                  1.009e+05
Df Model:               9
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	-50.8110	2.479	-20.497	0.000	-55.670	-45.952
CREDIT	1.8652	0.257	7.248	0.000	1.361	2.370
HOUSE	1.6099	0.926	1.738	0.082	-0.206	3.426
CAR	4.2582	0.918	4.637	0.000	2.458	6.059
HOUSE_L	-7.1289	1.030	-6.924	0.000	-9.147	-5.111
CAR_L	-7.1951	1.482	-4.854	0.000	-10.101	-4.290
EDUCATION	-2.0042	0.475	-4.218	0.000	-2.936	-1.073
WORKTIME	2.4355	0.426	5.721	0.000	1.601	3.270
INCOME	9.2260	0.308	29.918	0.000	8.622	9.831
AGE	0.8126	0.066	12.235	0.000	0.682	0.943

```
=====
Omnibus:                 11602.380      Durbin-Watson:           1.743
Prob(Omnibus):            0.000      Jarque-Bera (JB):        1282780.294
Skew:                     6.139      Prob(JB):                 0.00
Kurtosis:                 57.139      Cond. No.                 239.
=====
```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

"""

**0.0.5 4) Perform the Cox model (Proportional hazards model) and examine the relation between the platform default (survival) likelihood and platform characteristics such as RegCapital, Joinasso, etc.**

```
[9]: cox_dt = plat[['OnlineTime_YMD', 'Bankrupt_WDZJ', 'Collapse',
                  'RegCapital', 'Joinasso', 'Autobid',
                  'Capitaldeposit', 'Obtaininvest',
                  'Transright', 'Riskdeposit', 'Thirdguarantee']]
cox_dt.dropna(inplace=True) # Drop the missing values
cox_dt['OnlineTime_YMD'] = pd.to_datetime(cox_dt['OnlineTime_YMD'],
    ↪format='%Y%m%d')
cox_dt['Bankrupt_WDZJ'] = pd.to_datetime(cox_dt['Bankrupt_WDZJ'],
    ↪format='%Y%m%d') # Transfer the data into datetime variables
cox_dt['deltatime'] = (cox_dt['Bankrupt_WDZJ'] - cox_dt['OnlineTime_YMD']).dt.
    ↪days # Calculate the days difference between the 2 dates
cox_dt.drop(['OnlineTime_YMD', 'Bankrupt_WDZJ'], axis=1, inplace=True)
cph = CoxPHFitter()
cph.fit(cox_dt, duration_col = 'deltatime', event_col = 'Collapse')
cph.print_summary()
```

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%
covariate							
RegCapital	0.00	1.00	0.00	-0.00	0.00	1.00	1.00
Joinasso	-0.59	0.56	0.22	-1.03	-0.15	0.36	0.96
Autobid	-0.24	0.79	0.09	-0.41	-0.06	0.66	0.92
Capitaldeposit	-0.71	0.49	0.14	-0.99	-0.44	0.37	0.63
Obtaininvest	-0.35	0.71	0.27	-0.88	0.19	0.42	1.21
Transright	-0.37	0.69	0.11	-0.59	-0.16	0.56	0.80
Riskdeposit	-0.14	0.87	0.27	-0.67	0.38	0.51	1.46
Thirdguarantee	-0.06	0.94	0.23	-0.51	0.39	0.60	1.40

	exp(coef) upper 95%	cmp to	z	p	-log2(p)
covariate					
RegCapital	1.00	0.00	0.03	0.98	0.03
Joinasso	0.86	0.00	-2.62	0.01	6.81
Autobid	0.94	0.00	-2.61	0.01	6.80
Capitaldeposit	0.65	0.00	-5.05	0.00	21.07
Obtaininvest	1.21	0.00	-1.27	0.20	2.29
Transright	0.86	0.00	-3.37	0.00	10.40
Riskdeposit	1.46	0.00	-0.54	0.59	0.76
Thirdguarantee	1.47	0.00	-0.28	0.78	0.36