Assessing Small and Medium-sized Enterprises (SMEs) Financial Risk

Model applied: Binary logistic, probit and complementary log-log model

Stat 410 - Chi Nguyen

Why I chose this topic

I used to work at a startup, and this topic got me curious about what affects a small business’s financial health. I wanted to see if there’s a way to predict the risks that SMEs face.

Background

SMEs (Small and Medium-sized Enterprises) represent **99%** of businesses in the EU and **99.9%** in the US, which is crucial for the economy.

In 2024, SMEs contribute nearly **44% U.S GDP** and nearly two-third, roughly **63% of new jobs** growth in the U.S

Despite their importance, SMEs face **barriers** to accessing finance, primarily due to: such as high non-performing loans (NPLs), Limited capital reserves, Regulatory capital requirements, etc…

Economic factors like inflation and unemployment also impact SMEs' financial health and access to funding, affecting their growth and stability.

This analysis uses a logistic model to assess **key financial indicators** - NPLs, total assets, inflation, and unemployment - for primary predictors of financial risk in SMEs.

Dataset

The dataset

- acquired from PLOS ONE
- with SMEs from 2014–2018

I examine the risk associated with SMEs, using **Non-Performing Loans** (NPL) as a key indicator of financial health.

- >=15% as “High risk”
- <15% as “No/Low risk.

The variables:

- **Non-Performing Loans (NPL):** Indicates the percentage of overdue loans, a primary risk measure.

I set risk_indicator (to receive positive results for the output)

- 0 = High Risk (NPL >= 15%)
- 1 = No/Low Risk (NPL < 15%)

- **Unemployment Rate:** Reflects economic conditions; higher rates can reduce SME revenue and increase default risk.
- **Inflation Rate:** High inflation can raise costs, affecting SMEs' ability to service debt.
- **Total Assets** (totass_wonorm): SME financial size, with larger asset bases suggesting more stability.
(I scaled by dividing by 10,000 for analysis)

Choosing the model

	AIC	AICc	BIC
Logit	53335.67	53335.67	53370.25
Probit	53383.84	53383.84	53418.42
Cloglog	53732.94	53732.94	53767.52

From the R output, after comparing the AIC, AICC, and BIC values, the **binary logit model** has smallest values in all three criteria, and thus has a better fit

SAS code

SAS output

R code

```r
# Load the data
data = read.csv("/Users/christinenguyen/Downloads/SME - modify.csv")

# Create the binary risk variable based on `npl`
data$lowrisk = ifelse(data$npl < 15, 1, 0) # 1 = low risk (NPL < 15), 0 = high risk (NPL >= 15)
data$lowrisk = as.factor(data$lowrisk)

# Scale large numerical variables in `data`
data$totass_wonorm = data$totass_wonorm / 100000

# Fit the complementary log-log model (full model)
logit_model = glm(lowrisk ~ unemployment + inflation + totass_wonorm,
                  data = data, family = binomial(link = "logit"))

summary(logit_model)
```

R output

```
Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.905653   0.053396 -54.417   <2e-16 ***
unemployment   0.069596   0.002063  33.743   <2e-16 ***
inflation      0.014232   0.018587   0.766    0.444
totass_wonorm  0.086820   0.001922  45.163   <2e-16 ***
```

I set risk_indicator

- 0 = High Risk (NPL >= 15%)
- 1 = No/Low Risk (NPL < 15%)

Interpretation

The fitted model: $$\ln\frac{P(low\ risk)}{1-P(low\ risk)} = (-2.905653) + (0.069596).\text{unemployment} + (0.014232).\text{inflation} + (0.086820).\text{total asset}$$

At a 5% significance level,

- Unemployment and total assets are the significant predictors.
- Inflation is not the significant predictor (which make sense, since it may have fixed debt rate, gov support, etc...)

Interpretation of the estimated significant regression coefficients

Unemployment: For each 1% increase in unemployment, the estimated odds of an SME for “no/low risk" are exp(0.069596).100% = 107.2075% of those for “high risk"

Total Assets: For each additional $10,000 in total asset, the estimated odds of an SME for “no/low risk" are exp(0.086820).100% = 109.07% of those for “high risk"

Prediction

The current value in the U.S:

- Unemployment = 4.1
- Inflation = 2.4
- Mean Total Asset (computed by R) = 22.00441

```
> totass_wonorm = mean(data$totass_wonorm, na.rm = TRUE)
> print(totass_wonorm)
[1] 22.00441
```

Compute by hand:

$$P^0(\text{low risk}) = \frac{exp((-2.905653) + (0.069596).(4.1) + (0.014232).(2.4) + (0.086820).(22.00441))}{1 + exp((-2.905653) + (0.069596).(4.1) + (0.014232).(2.4) + (0.086820).(22.00441))}$$

$$= 0.33215 = 33.7215\%$$

SAS code _ Prediction

SAS output

R code _ Prediction

```
new_data = data.frame(
  unemployment = 4.1,
  inflation = 2.4,
  totass_wonorm = mean(data$totass_wonorm, na.rm = TRUE)
)

predicted_risk_logit = predict(logit_model,
                               newdata = new_data,
                               type = "response")

predicted_risk_logit
```

R output

```
> predicted_risk_logit
        1
0.3372135
```

Conclusion:

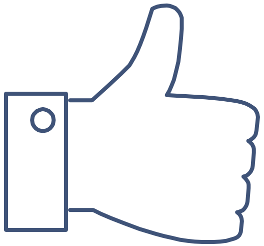
Similar results when computed by hand, R and SAS.

For an SME with:

- Total assets: $2,200,441 (scaled back by x10,000)
- Unemployment rate: 4.1%
- Inflation rate: 2.4%

The probability of having **low/no risk** (NPL < 15%) is **33.72%.**

This means there's a 33.72% chance the SME will have no or low risk, and consequently a **66.28%** chance of being classified as **high risk.**



THANKS!

DR.OLGA

& STAT 410 CLASSMATES