Exploring the sources of loan default clustering using survival analysis with frailty

Enrique Bátiz-Zuk (Banco de México), Abdulkadir Mohamed (Leeds University), Fátima Sánchez-Cajal (Banco de México)

Status: Final Version
Division of Financial Stability
Date: Friday October 8, 2021

BANCO DE MÉXICO®

Disclaimer

Any views expressed are only the speaker's own and do not necessarily reflect those of Banco de México or its Board.

Outline

Introduction: Objective

- Explore whether loan characteristics are relevant sources of loans default clustering using survival analysis with shared frailty.

- To model the correlation between loans *'time to default'* using industry, firm and bank's systemic importance.

- Time to default correlations across industry, firm and bank's type should help serve as a source of default clustering.

Introduction: Objective

- What is the mechanism that originates loan default clusters?

  - **Industry**: Variations in loan default rates across industries stemming from internal or external shocks could be correlated and drive default clustering.

  - **Firm**: Loan default rates might be concentrated at the firm level. A firm failure would lead to multiple loan defaults and could explain loan default clusters.

  - **Bank type**: Systemically important banks (SIBs) may have incentives to grant riskier loans applying lax lending standards, and the aggregate loan default rate will be prone to correlation risk.

Introduction: Motivation

From the industry point of view

- Stackhouse (2019): Multiple challenges for financial authorities under IFRS9 / CECL:

  1. A wide range of internal bank approaches is expected to emerge.
  2. Authorities will have difficulties to supervise (validate) different approaches.
  3. Banks will face a large regulatory cost in terms of:
     - personnel,
     - knowledge,
     - data gathering,
     - methodology design.
  4. Regulatory cost intensifies across small banks as their budget is much more limited.

Introduction: Motivation

From the academic point of view

- Das et al., 2007:
  - Macro-variables are not enough to control for systematic risk in aggregate firm default rates.
  - Frailty models incorporate the effect of unobservable covariates that are correlated across firms.

- Other studies further investigate role of frailties
  - Duffie et al., 2007, Campbell et al., 2008, Duffie et al., 2009, Chava et al., 2011, Koopman et al., 2008, Koopman et al., 2011, Monfort and Renne, 2013, and Azizpour et al., 2018.

- All these studies investigate the sources of corporate default clustering at the aggregate level.

- We examine the sources of loan default clusters at the individual loan level.

Introduction: Motivation

From a statistical point of view

- Survival analysis is superior compared to other methods [probit / logit / OLS / Discriminant Analysis].

1. Estimate probability to default (PD) for 'each individual loan' for any desired risk horizon.
   - PD is available for any future time horizon whereas probit has a fixed time horizon.
   - Survival is the right tool to implement (CECL) with a simple, closed form regulatory formula.

2. Handles right-censored data
   - Survival can match the life course of the loan.
   - Probit neglects this and treats this case as missing data.

3. Correlation in 'times to default' can be incorporated with shared frailties.

Introduction: What do we do?

- We take into account the correlation of firm loans that belong to one of the following three groups:
  - Industry (i.e., six industries)
  - Firm (i.e., any firm may have more than one loan)
  - Bank type that originated the loan (i.e., D-SIBs & non-DSIBs)

- We estimate five parametric survival models (M1-M5) using accelerated failure time
  - Excluding frailty
    - M1: Microeconomic variables
    - M2: M1 + macroeconomic variables
  - Including frailty
    - M3: M2 + shared frailty by industry
    - M4: M2 + shared frailty by firm
    - M5: M2 + shared frailty by bank type that originated the loan

- Perform a one-year out-of-sample prediction horse race for the five models under analysis.

Introduction: How do we contribute?

- Propose a simple regulatory PD formula that can be used by:
  - Small banks to reduce compliance cost.
  - Regulators for top-down stress test.
  - Supervisors as benchmark to facilitate supervision.
  - Banks to include correlation risk in PD to model fat tail of portfolio credit loss distribution.

- Following Gupta et al. (2018), we fill a gap in the academic literature.
  - In the context of credit risk, there is a lack of discussion on the importance of frailty.
  - Shared frailty model at the loan level using micro data provides prediction benefits.

Introduction: Main finding

- Including frailty by bank type is statistically superior for forecasting purposes only for bank loans to micro and small firms.
  - This is because the aggregate loan default rate is characterized by large values with clusters.
  - Including frailty leads to a loan loss provision model that captures default clustering attributed to the variability of the frailty across the two bank types.

- Including frailty is not useful for bank loans to medium-sized firms.
  - This is because the aggregate loan default rate is characterized by small values without clusters.

Outline

Literature

- Credit risk modeling can be broadly classified in:
  - Rules of thumbs and experience (Kerry 1993).
  - Structural models (Merton, 1974).:
  - Reduced-form models (Altman, 1968; Ohlson, 1980):
- Survival analysis is a reduced-form model.

Literature

- We can classify studies according to 'unit of analysis' into:
  - Firm bankruptcy
    - Gupta et al. (2018): single country (US).
    - Tsoukas (2011): multiple countries (emerging asian economies).
  - Firm loan default
    - Glennon and Nigro (2005): single country (US).
    - Dirick et al. (2017): a few banks of multiple countries (Europe).

- This paper analyses firm loan default for the banking sector of a single country

Literature

- According to survival methodology we identify:

  1. **'Classical survival models'**
     - Only two outcomes: Default & right Censored
     - Assume that every individual will eventually die.
       - Very conservative approach: risk overestimation
       - Simple and efficient to implement

  2. **'Competing risk models'**
     - Three or more outcomes: Default, Prepayment, Censored.
     - Assume that every individual will eventually die.

  3. **'Mixture cure rate models'**
     - Two or more outcomes: Default, Prepayment, Censored.
     - Assume that there is a cured fraction.
       - Less conservative and more accurate than Classical models
       - More complex and more difficult to implement.

- <u>This paper contributes to 'classical survival models'.</u>

Literature

- 'Classical survival models' have three methodologies:
  1. Parametric models
     - Two equivalent metrics:
       – Accelerated failure time (AFT)
       – Proportional hazards (PH)
  2. Semiparametric models
     - Proportional hazard (Cox (1972))
  3. Non-Parametric models
     - Survival curve (Kaplan and Meier, 1958)

- <u>This paper focus is on parametric models and uses AFT</u>
  - Simple and easy to implement closed form formula.

Outline

Data

Step 1: Definition of micro/SME loan

- There is no global standard of the definition of SMEs.
- We partially follow Mexico's Secretariat of Economy classification (3 criteria)

Enterprise size	*Sector*	*Number of employees*	*Annual sales (millions of dollars)*
Micro	All	< 11	< 0.2
Small	Agricultural and Commerce	≤ 30	≤ 5.6
	Manufacturing, Services, Construction and Communications and Transports	≤ 50	≤ 5.6
Medium	Commerce, Agricultural and Services	≤ 100	≤ 14.1
	Manufacturing, Construction and Communications and Transports	≤ 250	≤ 14.1

- Two additional criteria based on loan size:
  1. Any micro/SME loan has a maximum exposure size lower than 1 million USD.
  2. Drop micro/SMEs' loans that are lower than 10K USD.

Data

Step 2: Loan default definition

- There is no prescribed definition of loan default in IFRS 9.

- Loan that is past overdue for more than 90 consecutive days or when the bank has not received any payment during three consecutive months (i.e., Basel III).

Data

Step 3: Loan pool definition

- Loan sample has a panel data structure.
  - Each loan history is recorded and followed-up on an individual basis.

- During any study period, loans may:
  1. Originated and followed-up, such that we have loan records.

  2. We consider only those loans originated in our sample period.

Data

Step 4: Default status

- Default status
  - A number of events may happen once a loan has been originated:
    - Stay in performing state and remain outstanding or mature.
    - Default
    - Prepaid (i.e., repaid early)
    - Transferred to other financial entity
      - Securitization (i.e., pool of loans sold as a security)
      - Sold to any other bank or non-bank financial entity
    - Restructured
  - We consider only two states: default and censored

Data

Step 5: Data set characteristics

- We use monthly proprietary micro regulatory data collected by financial authorities
- Macroeconomic data that is publicly available.
- We have 1,893,927 micro/SMEs loans originated in the banking sector during January 2010 through April 2018.
- Data comprises 19,204,566 individual loan records.
- Comprehensive nature of our data allows to capture system-wide features.

Data

Step 6: Variable selection

Macro	Micro	Shared Frailty
Yield curve proxy	Interest rate	D-SIB
Consumer confidence index $_{t-3}$	ln (# of employees)	Firm ID
Economic activity index $_{t-3}$	ln (loan size)	Firm's industry
Inflation Rate $_{t-3}$	Firm's industry	

- Micro variables control for both loan characteristics and firm characteristics.
- We do not have firm balance sheet and market information which is a typical input in the corporate literature. This is because financial authorities do not collect this information.
- Variable selection is consistent with literature and fully described in Batiz-Zuk et al.(2020).

Data

Enterprise Size	Loan Default rate	Firm Default rate
Micro	7.61%	17.77%
Small	4.23%	14.70%
Medium	1.85%	13.46%
Total	4.83%	16.36%

- Steady state loan default rates vary significantly and increase with firm size.
- Steady state loan default rate grouped by firm id are naturally higher.
  - We assume that if a firm defaults in one of its loans, then the firm defaults on all its outstanding obligations with the banking sector. This does not occur in all cases.

Data

Industry:	Agriculture	Commerce	Construction	Transport	Manufacturing	Services
Enterprise Size	**Loan Default rate**	**Loan Default rate**	**Loan Default rate**	**Loan Default rate**	**Loan Default rate**	**Loan Default rate**
Micro	8.51%	6.81%	10.05%	7.71%	7.48%	9.50%
Small	2.97%	4.20%	5.41%	3.35%	3.67%	6.11%
Medium	2.31%	1.42%	3.26%	1.31%	1.67%	2.15%
Total	4.69%	4.32%	6.05%	4.58%	3.96%	6.45%

- Loan default rates vary across six industry sectors.
- For “micro” and “medium” sized enterprises, construction has the highest default rates.
- For “small” sized enterprises, services has the largest loan default rate.

Data

	D-SIBs	Non-DSIBs
Enterprise Size	**Loan Default rate**	**Loan Default rate**
Micro	9.50%	3.17%
Small	6.11%	2.16%
Medium	2.15%	1.34%
Total	6.45%	2.25%

- Loan default rates for D-SIBs have a greater value compared to Non-DSIBs.

Outline

Methodology

- We estimate the survival model using Accelerated Failure Time Model (AFT) (Cleves et al., 2010) as:

$$\ln\left(t_{jk}\right) = \beta_0 + \beta_1 x_{1jk} + \cdots + \beta_p x_{pjk} + \ln\left(\tau_{jk}\right)$$

  - t_{jk} is time to loan default for **loan *j*** in **time span *k*** (observation) measured in days,
  - β covariate coefficients,
  - x_{djk} time-varying or time-constant p covariates ($d=1,\dots,p$),
  - $\ln(\tau_{jk})$ random quantity that follows a distribution.

- Model is estimated using maximum likelihood techniques.

- Survival function depends on the distribution assumed for τ_{jk}.
  - We compare six distributions: Exponential, Weibull, Gompertz, Log-logistic, and Log-Normal.
  - We use Akaike Information Criterion (see Akaike (1974)) to choose the best-fitting model for our data.

$$AIC = -2\log likelihood + 2p$$

  - Log-Normal provided the best fit.

Methodology

- When τ_{jk} is distributed as lognormal with parameters β_0 and σ, the conditional survival function is defined as:

$$S\left(t_{jk} \middle| x_{1jk},\ldots,x_{pjk}\right) = 1 - \Phi\left(\frac{\ln\left(t_{jk}\right) - \left(\beta_0 + \beta_1 x_{1jk} + \cdots + \beta_p x_{pjk}\right)}{\sigma}\right)$$

  - $\Phi()$ is the CDF for standard Gaussian (normal) distribution
  - σ is a strictly positive ancillary parameter
- Shared frailty α is introduced as an unobservable multiplicative effect on the hazard.
- Gutierrez (2002. p. 24) shows that the individual survival function conditional on the frailty is

$$S(t_{jk} \mid x_{1jk},\ldots,x_{njk},\alpha) = \{S(t \mid x_{1jk},\ldots,x_{njk})\}^{\alpha}$$

- When α is distributed as gamma with mean one and variance θ the survival function becomes

$$S_\theta\left(t_{jk} \middle| x_{1jk},\ldots,x_{pjk}\right) = \left[1 - \theta\, ln\{S\left(t_{jk} \middle| x_{1jk},\ldots,x_{pjk}\right)\}\right]^{-1/\theta}$$

$$S\left(t_{ijk} \middle| x_{1ijk},\ldots,x_{nijk},\alpha_i\right) = \left\{1 - \Phi\left(\frac{\ln t_{ijk} - \left(\beta_0 + \beta_1 x_{1ijk} + \cdots + \beta_p x_{pijk}\right)}{\sigma}\right)\right\}^{\alpha_i}$$

- Madorno (2013): PD for any desired risk horizon b (measured in days) is:

$$PD\left(t_{jk} + b \middle| x_{1jk},\ldots,x_{njk}\right) = 1 - \frac{S\left(t_{jk} + b \middle| x_{1jk},\ldots,x_{njk}\right)}{S\left(t_{jk} \middle| x_{1jk},\ldots,x_{njk}\right)}$$

Outline

Results for Akaike

- Akaike Information Criterion is used to discriminate between parametric distributions in the context of non-nested models for ‘micro’, ‘small’, and ‘medium’-sized enterprises.
- We use a benchmark AFT model (i.e., M2) that incorporates micro and macro variables.

Micro firms

Distribution	AIC
Exponential	329,742
Gompertz	319,639
Weibull	308,096
Loglogistic	306,765
Lognormal	303,673

Small firms

Distribution	AIC
Exponential	259,243
Gompertz	250,202
Weibull	239,136
Loglogistic	237,789
Lognormal	234,414

Medium firms

Distribution	AIC
Exponential	62,778
Gompertz	60,894
Weibull	58,125
Loglogistic	57,871
Lognormal	56,980

Results for micro-sized firms

Variable (Expected sign)	Standard AFT		AFT + shared frailty		
	M1	M2	M3	M4	M5
Interest rate (-)	-0.0361[a]	-0.0340[a]	-0.0334[a]	-0.0198[a]	-0.0333[a]
ln(number of employees) (+)	0.0379[a]	-0.0125[a]	-0.0103[a]	0.0387[a]	-0.0113[a]
ln(loan size) (+/-)	0.0028	-0.0179[a]	-0.0144[a]	-0.0330[a]	-0.0140[a]
D-SIBs (+/-)	0.2109[a]	0.2490[a]	0.2716[a]	0.1604[a]	
Fixed effects industry (+/-)	Yes	Yes	No	Yes	Yes
Yield curve proxy (+)		-0.0763[a]	-0.0758[a]	-0.0467[a]	-0.0798[a]
Consumer confidence index$_{t-3}$ (+)		0.0210[a]	0.0229[a]	0.0260[a]	0.0220[a]
Economic activity index$_{t-3}$ (+)		0.0275[a]	0.0254[a]	0.0070[a]	0.0249[a]
Inflation$_{t-3}$ (-)		-0.0329[a]	-0.0315[a]	-0.0227[a]	-0.0355[a]
Constant	7.9400[a]	3.6763[a]	2.2950[a]	4.5718[a]	2.7220[a]
Lognormal standard deviation parameter (sigma)					
ln(sigma)	0.1526[a]	0.0575[a]	-0.4377[a]	-0.2304[a]	-0.3300[a]
Shared frailty variance parameter (theta)					
ln(theta)$_{Industry}$			0.6252[a]		
ln(theta)$_{Firm\ id}$				1.6531[a]	
ln(theta)$_{D\text{-}SIBs}$					0.2886[a]

Notes: a (b) [c] significant at 1% (5%) [10%] level (two-sided test)

Results for small-sized firms

	Standard AFT		AFT + shared frailty		
Variable (Expected sign)	M1	M2	M3	M4	M5
Interest rate (-)	-0.0418[a]	-0.0446[a]	-0.0391[a]	-0.0084[a]	-0.0409[a]
ln(number of employees) (+)	0.0151[a]	0.0133[a]	0.0082[a]	-0.0109[a]	0.0090[a]
ln(loan size) (+/-)	-0.0308[a]	-0.0365[a]	-0.0212[a]	-0.0393[a]	-0.0213[a]
D-SIBs (+/-)	0.2605[a]	0.2668[a]	0.2825[a]	0.1857[a]	
Fixed effects industry (+/-)	Yes	Yes	No	Yes	Yes
Yield curve proxy (+)		-0.0757[a]	-0.0613[a]	-0.0424[a]	-0.0687[a]
Consumer confidence index$_{t-3}$ (+)		0.0058[a]	0.0085[a]	0.0144[a]	0.0072[a]
Economic activity index$_{t-3}$ (+)		-0.0011[c]	-0.0007	-0.0275[a]	-0.0019[a]
Inflation$_{t-3}$ (-)		-0.0203[a]	-0.0072[c]	0.0130[a]	-0.0069
Constant	8.5561[a]	8.4697[a]	6.0377[a]	9.1860[a]	6.7435[a]
Lognormal standard deviation parameter (sigma)					
ln(sigma)	0.0757[a]	0.0684[a]	-0.5158[a]	-0.2246[a]	-0.4136[a]
Shared frailty variance parameter (theta)					
ln(theta)$_{Industry}$			0.8747[a]		
ln(theta)$_{Firm\ id}$				2.0240[a]	
ln(theta)$_{D\text{-}SIBs}$					0.6211[a]

Notes: a (b) [c] significant at 1% (5%) [10%] level (two-sided test)

Results for medium-sized firms

Variable (Expected sign)	Standard AFT		AFT + shared frailty		
	M1	M2	M3	M4	M5
Interest rate (-)	-0.0564[a]	-0.0595[a]	-0.0485[a]	0.0026	-0.0515[a]
ln(number of employees) (+)	-0.0080	-0.0069	-0.0083	-0.0275[b]	-0.0023
ln(loan size) (+/-)	0.0049	-0.0009	0.0216[a]	0.0246[a]	0.0195[a]
D-SIBs (+/-)	0.1176[a]	0.1289[a]	0.1155[a]	0.0267	
Fixed effects industry (+/-)	Yes	Yes	No	Yes	Yes
Yield curve proxy (+)		-0.0215[c]	-0.0024	-0.0168	-0.0014
Consumer confidence index$_{t-3}$ (+)		0.0047[a]	0.0073[a]	0.0094[a]	0.0067[a]
Economic activity index$_{t-3}$ (+)		0.0014	0.0021[c]	-0.0427[a]	0.0016
Inflation$_{t-3}$ (-)		0.0358[a]	0.0546[a]	0.0862[a]	0.0535[a]
Constant	8.6013[a]	8.0296[a]	5.2253[a]	10.1571[a]	5.4420[a]
Lognormal standard deviation parameter (sigma)					
ln(sigma)	0.1603[a]	0.1575[a]	-0.4600[a]	-0.1359[a]	-0.4294[a]
Shared frailty variance parameter (theta)					
ln(theta)$_{Industry}$			0.9998[a]		
ln(theta)$_{Firm\ id}$				2.3314[a]	
ln(theta)$_{D\text{-}SIBs}$					0.9007[a]

Notes: a (b) [c] significant at 1% (5%) [10%] level (two-sided test)

Summary of Results: Micro Variables

- Regarding Micro Variables:
  - Interest rate has a negative impact on the time to default across all firm sizes.
  - Impact of number of employees varies per firm size:
    - a negative influence on “micro”,
    - a positive influence on small firms and
    - not statistically significant influence on medium enterprises.
  - Bank type (i.e., D-SIBs, non-DSIBs) indicator is significant and positive.

Summary of Results: Macro Variables

- Regarding Macro Variables:
  - For all firm sizes, macro variables have a significant impact on the survival rates of SMEs.
  - Yield curve proxy is negative for micro/small enterprises, and not statistically significant for medium firms.
  - Changes in 3-month lagged consumer confidence have positive & significant impact irrespective of firm size.
  - Changes in economic activity have a mixed impact:
    - positive and significant for micro firms,
    - negative and significant on small firms and
    - not significant (i.e. M2, M3, and M5) for medium firms.
  - Changes of inflation rate are mixed.
    - Negative positive for micro and small.
    - Positive for medium sized firms.

Summary of Results: Ancilliary Parameters

- Regarding Ancillary Parameters:
  - Coefficient for ln(σ) is significant for all specifications across the three firm types.
    - Sign of ln(σ) is positive (negative) when frailty is excluded (included).
  - Coefficient of variance parameter ln(theta) is positive and significant for all specifications.
    - Shared frailty coefficient captures the effect of default clustering attributed to the variability of the frailty across industry, firm and bank type.
    - Size of variance parameter ln(theta) is extremely large for M4. The variance parameter, is:
      - exp 2.3314 =**10.29** (medium),
      - exp 2.0240 = **7.57** (small),
      - exp 1.65 = **5.22** (micro).
      - This has an impact on the value of all the estimated coefficients independently of firm size.
    - Lando and Nielsen (2010) have argued to aggregate to account for parent-subsidiary relationships. However, we don't have the information.

Results: Goodness of fit

Micro firms	Standard AFT		AFT + shared frailty		
	M1	M2	M3	M4	M5
Chi^2 statistic	2,839.7[a]	9,305.1[a]	9,064.8[a]	5,295.1[a]	8,951.5[a]
LogLikelihood	-155,054	-151,821	-151,501	-134,439	-151,725
AIC	310,130	303,672	303,024	268,910	303,480

Small firms	Standard AFT		AFT + shared frailty		
	M1	M2	M3	M4	M5
Chi^2 statistic	2,770.6[a]	3,219.7[a]	2,939.0[a]	3,706.4[a]	2,197.7[a]
LogLikelihood	-117,416	-117,192	-116,765	-92,277	-117,100
AIC	234,854	234,414	233,552	184,586	234,230

Medium firms	Standard AFT		AFT + shared frailty		
	M1	M2	M3	M4	M5
Chi^2 statistic	1,082.2[a]	1,149.3[a]	738.7[a]	1,186.1[a]	914.2[a]
LogLikelihood	-28,508.47	-28,474.93	-28,377.74	-20,137.52	-28,390.85
AIC	57,039	56,980	56,777	40,307	56,812

Summary of Results: Goodness of fit

- We report the Chi-squared statistic based on Wald test to assess if the model is significant.
- We report the log-likelihood and Akaike Information Criterion (AIC) of each model to assess the fit.
- Model that provides the best fit (M4) according to log-L and AIC criteria is not the best in terms of forecasting.
- These models do not have an R-squared or an adjusted R-squared measure.
- There is a pseudo R-squared but its use is not standard in the literature as this measure is not bounded.

Outline

Model Validation

Validation can be explained in three steps.

1. Hold-in sample:
   - Estimate survival AFT model using period from January 2010 to December 2015.
   - Idea is to use 80% of the sample size.

2. Hold-out or test sample:
   - Use the period from January 2016 to May 2017.
   - Idea is to use remaining 20% of the sample size.

3. Estimate the one year PD as proposed by Madorno et al. (2013) for the hold-out sample.

4. Compute Area Under Receiver Operating Characteristic Curve (AUROC) for the hold-out-sample.
   - Size of our hold-out sample is large: 174k (medium); 349k (small); 230k (micro)
   - Most of previous studies (see Gupta et al., 2018, p.461) have ROC curves that are steps rather than concave estimates of AUROC.

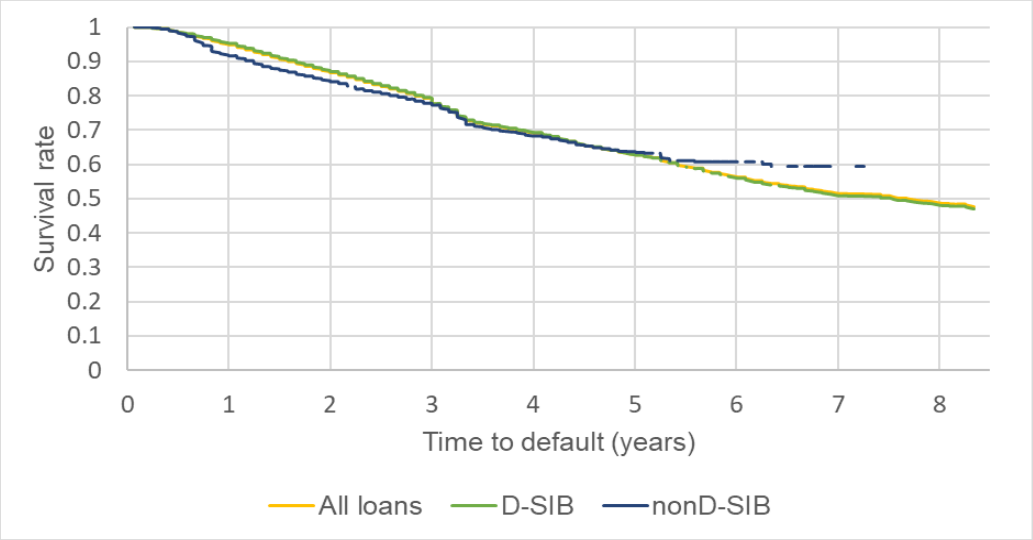
Model Validation

Model M1-M5	Survival AFT M1	Survival AFT M2	Survival AFT + shared frailty M3 Industry	Survival AFT + shared frailty M4 Firm	Survival AFT + shared frailty M5 Bank Type
Micro firms					
AUROC-H	0.5365	0.5545	0.5543	0.5649	0.6208
Small firms					
AUROC-H	0.5126	0.5188	0.5019	0.5485	0.5807
Medium firms					
AUROC-H	0.7495	0.7980	0.7889	0.6089	0.7936

- Size of AUROC-H differs depending on the firm size and specification.
  - **Micro firms**: M5 has the best significant prediction performance. Size of AUROC is regarded as a fair.
  - **Small firms**: M5 has the best significant prediction performance. Size of AUROC is regarded as poor.
  - **Medium firms**: M2, M3 and M5 have the best prediction performance. Size of AUROC is regarded as very good.
    - Difference in AUROC between M2, M3 and M5 is <u>not statistically significant</u> at 5%.
- AUROC performance of medium firms improves because aggregate loan default rate has lower values and variability.
  - Low values of AUROC might be due to lack of more financial information such as balance sheet data.

Model Validation

- A closer inspection of the evolution of the Kaplan Meier loan survival function by bank type for “micro” loans shows that, in the long run, the number of defaults is relatively high compared to non-DSIBs.



Outline

I. Introduction
II. Background Literature
III. Data
IV. Methodology
V. Results
VI. Model Validation
VII. Conclusions
VIII. References

Conclusions

- Bank's systemic importance is a key source of default clustering for loans granted to micro and small enterprises. This evidence is important for developing policy tools to mitigate systemic risk.

- More evidence at the international level is desirable.

- Two policy recommendations
  1. **Financial authorities should use a PD regulatory formula as a benchmark.** This issue is relevant for:
     - Banks to recognize loss in a simple, timely and accurate way.
     - Investors to compare loan loss provisions based on a similar benchmark to promote transparency.
     - Supervisors to verify banks compliance.
     - Regulators for stress testing purposes.

  2. **PD regulatory formula should include a shared frailty by bank type**. Especially, for loans characterized by large average value of aggregate loan default rate (e.g., micro and small).
     - Including frailty promotes a loan loss provision model that captures default clustering
     - Clustering is attributed to the variability of the frailty across the two bank types.

Outline

References

- **Akaike, H.** (1974). “A new look at the statistical model identification”. IEEE Transaction on Automatic Control 19: 716–723.
- **Altman, E.** (1968). “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy.” *Journal of Finance* 23(4): 589–609.
- **Basel Committee on Banking Supervision** (BCBS), 2017. Basel III: Finalising post crisis reforms.
- **Batiz-Zuk, E., López-Gallo, F., Mmuhamed, A., Sánchez-Cajal, F.** (2020) “Determinants of loan survival rates for small and medium-sized enterprises: Evidence from an emerging economy” International Journal of Finance and Economics: 1-15.
- **Bluhm, C., Overbeck, L., and C. Wagner** (2003). “An Introduction to Credit Risk Modelling”, Chapman & Hall, London.
- **Campbell, J., Hilscher, J. and Szilagyi, J**. (2008) “In search of distress risk”. The Journal of Finance 63(6), 2899-2939.
- **Carling, K., Jacobson, T., Lindé, J. and Roszbach, K.** (2007) “Corporate credit risk modelling and the macroeconomy”. Journal of Banking and Finance 31, 845-868.
- **Cleves, M., Gutierrez, R., Gould, W. and Marchenko, Y.** (2010) “An introduction to Survival Analysis Using Stata”. Stata Press, Texas.
- **Cox, D.** (1972) “Regression models and life-tables”. Journal of the Royal Statistical Society. Series B (Methodological) 34, 187-220.

References

- **Covas, F. and Nelson, W.** (2018) "Current expected credit loss: Lessons from 2007-2009". Banking Policy Institute Working Paper.
- **Das, S., Duffie, D., Kapadia, N. and Saita, L.** (2007) "Common failings: How corporate defaults are correlated". The Journal of Finance 62(1), 93-117.
- **Dirick, L., Claeskens, G. and Baesens, B.** (2017) "Time to default in credit scoring using survival analysis: A benchmark study". Journal of the Operational Research Society 68, 652-665.
- **Duffie, D., Eckner, A., Horel, G. and Saita, L.** (2009) "Frailty correlated default". The Journal of Finance 64(5), 2089-2123.
- **Duffie, D., Saita, L. and Wang, K.** (2007) "Multi-period corporate default prediction with stochastic covariates". Journal of Financial Economics 83, 635-665.
- **Glennon, D. and Nigro, P.** (2005)"Measuring the default risk of small business loan: A survival analysis approach". Journal of Money Credit and Banking. 37:923–947.
- **Gupta, J. Gregoriou, A., Ebrahimi, T.** (2018) "Empirical comparison of hazard models in predicting SMEs failure. Quantitative Fianance 18(3), 437-466.
- **Gupta, J., Gregoriou. A. and Healy, J.** (2015) "Forecasting bankruptcy for SMEs using hazard function: To what extent does size matter?" Review of Quantitative Finance and Accounting, 45,845–869.

References

- **Holmes P., Hunt A., Stone I. (2010)** "An analysis of new firm survival using a hazard function." Applied Economics 42:185–195.
- **Kaplan, E. and Meier, P.** (1958) "Nonparametric estimation from incomplete observations". Journal of the American Statistical Association 53 (282), 457-481.
- **Lando, D. and Nielsen, M.** (2010). "Correlation in corporate defaults: Contagion or conditional independence?" Journal of Financial Intermediation 19(3), 355-372.
- **Shumway, T.** (2001) "Forecasting bankruptcy more accurately: A simple hazard model" *Journal of Business* 74(1): 101–124.
- **Stackhouse, J.** (2019a). "Managing CECL change: How banks can prepare for the new accounting standard". Federal Reserve Bank of St. Louis. Available at:https://www.stlouisfed.org/on-the-economy/2019/september/cecl-how-bankscan-manage-change-prepare; accessed on February 21, 2020.
- **Stackhouse, J.,** (2019b). "The CECL model: Accounting changes coming for banks. Federal Reserve Bank of St. Louis." Available at: https://www.stlouisfed.org/onthe-economy/2019/august/cecl-model-accounting-changes-coming-banks; accessed on February 21, 2020.
- **Zmijewski, M.** (1984). "Methodological issues related to the estimation of financial distress prediction models." *Journal of Accounting Research* 22(3): 59–82.

BANCO DE MÉXICO®

www.banxico.org.mx

Based on publicly available information