Failing Banks R Replication

Variable Definitions

Data Dictionary

November 17, 2025

Contents

1 Overview

This document provides comprehensive definitions for all 70+ variables used in the Failing Banks R replication package. Variables are organized by category for easy reference.

1.1 Naming Conventions

- **F[N]_failure**: Binary indicator for failure within N years/quarters
- **L_variable**: Lagged value (one period)
- **L[N]_variable**: Lagged value (N periods)
- **_ratio**: Expressed as proportion (0-1 scale)
- **_growth**: Log difference (approximates percentage change)

2 Identification Variables

Table 1: Bank and Time Identifiers

Variable	Type	Definition
bank_id	Integer	Unique bank identifier assigned by data source. Range: 1-99999. Consistent across time periods.
year	Integer	Calendar year from report date. Range: 1863-2024. Annual for historical era.
quarter	Integer	Calendar quarter (1-4). Only available for modern era (1959+).
quarter_number	Integer	Alias for quarter variable. Same range and definition.
report_date	Date	Date of call report submission. Date object. Range: 1863-2024.
era_group	Categorical	Era classification: "Historical" (1863-1941) or "Modern" (1959-2024).
age	Integer	Bank age in years since charter. Range: 0-150 years.

3 Failure Variables

3.1 Core Failure Indicators

Table 2: Primary Failure Variables

Variable	Type	Definition
failed_bank	Binary	Permanent indicator: equals 1 if bank ultimately failed, 0 otherwise.

(continued on next page)

(continued from previous page)

Variable	Type	Definition
days_to_failure	Integer	Days until failure (negative before failure). Range: -10000 to 0. NA if bank never fails.
months_to_failure	Integer	Months until failure. Calculated as days_to_failure / 30.4375. Range: -300 to 0.
quarters_to_failure	Integer	**KEY VARIABLE**. Quarters until failure. Calculated as days_to_failure / 91.25. Range: -300 to 0. Negative values indicate bank has not yet failed.
time_to_fail	Integer	Years until failure. Calculated as floor(days_to_failure / 365). Range: -20 to 0.
receivership_date	Date	Date receiver appointed (historical era only, 1863-1937). From OCC data.
fail_day	Date	Failure date for modern era banks (1959-2024). From FDIC data.
final_year	Integer	Last year bank appears in dataset. May not be failure year if bank merged/exited.

3.2 Horizon-Specific Failure Indicators

These variables indicate failure within specific time horizons. Created in analysis scripts (32, 35).

Table 3: Failure Horizon Indicators

Variable	Values	Definition
F1_failure	0 or 100	Fails within 1 year (4 quarters). Formula: $100 \times (quarters\_to\_failure \geq -4 and \leq -1)$
F2_failure	0 or 100	Fails within 2 years (8 quarters). Used in cross-section analysis (Script 32).
F3_failure	0 or 100	**PRIMARY OUTCOME**. Fails within 3 years (12 quarters). Formula: $100 \times (quarters\_to\_failure \geq -12 and \leq -1)$
F4_failure	0 or 100	Fails within 4 years (16 quarters). Cross-section analysis.
F5_failure	0 or 100	Fails within 5 years (20 quarters). Alternative horizon in Script 35.
F6_failure	0 or 100	Fails within 6 years (24 quarters). Cross-section only.
F1_failure_run	0 or 1	Fails within 1 year *AND* experienced bank run. Conditional indicator.
F3_failure_run	0 or 1	Fails within 3 years *AND* experienced bank run. Conditional indicator.

4 Balance Sheet Variables

Table 4: Core Balance Sheet Items

Variable	Type	Definition
assets	Continuous	Total assets in nominal dollars. Range: $100 - $1 trillion. Sum of all asset categories from call reports.
deposits	Continuous	Total deposits in nominal dollars. Range: $50 - $1 trillion. Includes demand, time, and savings deposits.
loans	Continuous	Total loans in nominal dollars. Range: $20 - $100 billion. Sum of all loan categories.
liquid	Continuous	Liquid assets in nominal dollars. Range: $10 - $100 billion. *Historical*: cash + securities + due from banks. *Modern*: cash + securities + federal funds purchased.
equity	Continuous	Total equity/book value in nominal dollars. Range: $5 - $10 billion. Formula: capital + surplus + undivided_profits.
capital	Continuous	Capital stock in nominal dollars. Par value of outstanding shares. Primary equity component for historical banks.
surplus	Continuous	Surplus/retained earnings in nominal dollars. Accumulated retained earnings over time. Historical era.
undivided_profits	Continuous	Current period undivided profits. Historical era.
log_assets	Continuous	Natural log of total assets. Range: 5-20. Used as bank size measure in regressions.

5 Financial Ratios

5.1 Key Risk Measures

Table 5: Primary Financial Ratios

Variable	Definition & Interpretation
equity_ratio	**SOLVENCY MEASURE**. equity / assets. Range: 0.01-0.50 (1%-50%). Higher values indicate better capitalization and lower insolvency risk.
loan_ratio	**ASSET RISK**. loans / assets. Range: 0.20-0.90 (20%-90%). Higher values indicate more lending activity but potentially higher credit risk.
liquid_ratio	**LIQUIDITY MEASURE**. liquid / assets. Range: 0.05-0.60 (5%-60%). Higher values indicate better ability to meet short-term obligations.
surplus_ratio	Profitability indicator (historical). surplus / equity. Range: 0-2. Higher values suggest accumulated profitability.

(continued on next page)

(continued from previous page)

Variable	Definition & Interpretation
cash_ratio	Narrow liquidity measure (historical). cash_reserves / assets. Range: 0-0.50. Strictest liquidity measure.
income_ratio	Profitability measure (modern only, post-1941). net_income / assets. Range: 0-0.20 (0%-20% ROA).

6 Growth Variables

Table 6: Growth Measures

Variable	Definition & Calculation
assets_growth	Annual asset growth rate. Formula: $\ln(assets_t) - \ln(assets_{t-1})$. Range: -0.50 to +0.50 (-50% to +50%). Year-over-year growth.
deposits_growth	Annual deposit growth rate. Same formula as assets_growth but for deposits. Range: -0.50 to +0.50.
growth	**PRIMARY GROWTH MEASURE**. 3-year asset growth. Formula: $\ln(assets_t) - \ln(assets_{t-3})$. Range: -1 to +1. Used for growth quintile analysis.
growth_cat	Growth quintile (1-5). Within-year quintiles of growth variable. 1 = slowest growing (Q1), 5 = fastest growing (Q5). Created in Scripts 32 and 35.
L3_assets	Assets lagged 3 periods. Used to calculate growth variable. In dollars.

7 Lagged Variables

Lagged variables are used as predictors in regressions to avoid simultaneity bias.

Table 7: Lagged Predictors

Variable	Definition & Usage
L_equity_ratio	One-period lagged equity ratio. Used in Scripts 08, 33, 34, 51. Predicts future failure. Range: 0.01-0.50.
L_loan_ratio	One-period lagged loan ratio. Used in same scripts. Captures asset risk one period prior. Range: 0.20-0.90.

(continued on next page)

(continued from previous page)

Variable	Definition & Usage
L_liquid_ratio	One-period lagged liquid ratio. Used in same scripts. Prior period liquidity. Range: 0.05-0.60.
L_log_assets	One-period lagged log assets. Controls for prior bank size in regressions. Range: 5-20.

8 Bank Run Variables

Table 8: Deposit Run Indicators

Variable	Definition
run	Binary indicator for bank run. Equals 1 if deposit_outflow $>$ 0.10 (10% decline). Created in Script 06.
deposit_outflow	Deposit decline rate. Formula: (last_deposits - failure_deposits) / last_deposits. Range: -0.50 to +1.0. Positive = outflows (decline), negative = inflows (growth).
run_is_missing	Indicator for missing run data. Equals 1 if run information unavailable for that bank-year.

9 Receivership Variables

Variables related to bank failure outcomes and depositor recovery (historical era).

Table 9: Receivership Outcomes

Variable	Definition
last_call_deposits	Deposits at last call report before failure. In nominal dollars.
last_call_assets	Assets at last call report before failure. In nominal dollars.
deposits_at_suspension	Deposits at suspension/failure date. From receivership records.
assets_at_suspension	Assets at suspension/failure date. From receivership records.
total_assessed	Total assets assessed by receiver. Estimated liquidation value.
rho	Recovery rate. Calculated as total_assessed / deposits_at_suspension. Range: 0-1. Mean: 0.0006 (0.06%).

(continued on next page)

(continued from previous page)

Variable	Definition
depositor_loss	Depositor loss. Calculated as 1 - rho. Mean: 0.9994 (99.94%).
receivership_days	Days in receivership. From failure to final resolution.
receivership_years	Years in receivership. receivership_days / 365.25.

10 Macro Variables

Macroeconomic controls used in regressions.

Table 10: Macroeconomic Variables

Variable	Definition & Source
rgdp_pc	Real GDP per capita. From BEA (1947+) and Barro-Ursua (pre-1947). In constant dollars.
rgdp_growth	Real GDP growth rate. Log difference of rgdp_pc.
cpi	Consumer Price Index. From GFD and JST databases. Base year varies.
tbill_rate	Treasury bill rate. Short-term interest rate from GFD. Percentage points.
bond_yield	Long-term bond yield. From GFD. Percentage points.
crisis	Financial crisis indicator. From JST database. Binary: 1 = crisis year, 0 = normal.

11 Variable Creation Scripts

11.1 Script Reference

Table 11: Which Scripts Create Which Variables

Script(s)	Variables Created
04	Historical balance sheet variables, age, failed_bank, receivership_date, quarters_to_failure
05	Modern balance sheet variables, fail_day, quarters_to_failure, quarter
06	run, deposit_outflow, receivership outcomes
07	era_group, combined panel
08	Lagged variables (L_equity_ratio, L_loan_ratio, L_liquid_ratio, L_log_assets) for coefplots

(continued on next page)

(continued from previous page)

Script(s)	Variables Created
32	F1-F6_failure indicators, growth, growth_cat (cross-section analysis)
35	F1/F3/F5_failure, F1/F3_failure_run, growth, growth_cat (main regression dataset)
51	Uses lagged variables created in Script 08 for AUC analysis

12 Key Variable Notes

12.1 Most Important Variables

The following variables are central to the paper's analysis:

1. **F3_failure**: Primary outcome variable (failure within 3 years)
2. **quarters_to_failure**: Core timing variable for all failure indicators
3. **growth** & **growth_cat**: Primary finding - shrinking banks fail more
4. **equity_ratio**: Solvency measure, strong predictor of failure
5. **liquid_ratio**: Liquidity measure, important but weaker than equity
6. **loan_ratio**: Asset risk measure
7. **run**: Bank run indicator, amplifies failure risk 3-4x

12.2 Era Differences

Some variables differ by era:

- **Historical (1863-1941)**: receivership_date, surplus_ratio, cash_ratio
- **Modern (1959-2024)**: fail_day, quarter variables, income_ratio
- **Both eras**: Core balance sheet items, ratios, growth measures

12.3 Missing Data Patterns

- **run**: Only available for subset of failing banks with receivership data
- **income_ratio**: Only available post-1941 (modern reporting standards)
- **quarters_to_failure**: Only defined for failing banks (NA for survivors)
- **F[N]_failure**: Only non-zero for failing banks within horizon

Table 12: Combined Dataset Summary

Statistic	Value	Notes
Total observations	2,872,893	Bank-year/quarters
Unique banks	30,000	Historical + Modern
Failure rate (overall)	7.7%	Failed banks / all banks
Time span	1863-2024	161 years
Historical obs	340,000	1863-1941, annual
Modern obs	2,500,000	1959-2024, quarterly

13 Summary Statistics

13.1 Dataset Overview

13.2 Key Variable Ranges

Table 13: Typical Values for Core Ratios

Variable	Mean	Median	Std Dev
equity_ratio	0.12	0.10	0.06
loan_ratio	0.60	0.62	0.15
liquid_ratio	0.25	0.22	0.12
log_assets	12.5	12.3	2.1
growth (3-year)	0.15	0.12	0.35

14 Conclusion

This data dictionary provides comprehensive definitions for all variables in the Failing Banks R replication package. For questions about specific variables, see the technical documentation or examine the R scripts that create them.

14.1 Additional Resources

- **Technical_Documentation.pdf**: Detailed model specifications
- **Quick_Start_Guide.pdf**: How to run the replication
- **Source Scripts**: 1_source_code/ directory for variable creation logic