Failing Banks R Replication

Setup Instructions

Installation Guide

November 17, 2025

1 Quick Start

1.1 Stata Directory Structure (Recommended)

If you have the original Stata replication package:

```
Your_Project/
  |-- sources/      # Your existing Stata data
  |-- code/         # Extract R package here
  |-- dataclean/    # Created by scripts
  |-- output/       # Created by scripts
  +-- tempfiles/    # Created by scripts
```

Steps:

1. Extract package into `code/` folder
2. Open R: `setwd("path/to/Your_Project/code")`
3. Run: `source("RUN_ALL.R")`

Scripts auto-detect Stata structure!

1.2 R Directory Structure

Alternative structure:

```
Your_Project/
  |-- code/               # R scripts
  |-- 2_input_data/
  |   +-- Sources/        # Copy data here
  +-- 3_output_data/      # Created by scripts
```

2 Required Data

2.1 Critical Files

- `call-reports-historical.dta` (∼800 MB)

- `call-reports-modern.dta` (~1 GB)
- `FDIC/` folder with failure data
- `JST/` folder with macro data
- `GFD/` folder with yields and CPI
- `Macro/` folder with GDP data
- `occ-receiverships/` folder

2.2 Where to Get Data

1. **Original Stata package**: `qje-repkit-to-upload/sources/`
2. **FDIC website**: https://www.fdic.gov/
3. **JST Database**: https://www.macrohistory.net/database/
4. **Contact authors** for cleaned historical data

3 System Requirements

3.1 Software

- **R**: Version 4.0+ (tested on 4.4.1)
- **RStudio**: Optional but recommended

3.2 Packages (Auto-installed)

tidyverse, haven, fixest, ggplot2, pROC, readxl, lubridate, scales, broom, patchwork, caret, knitr, kableExtra, openxlsx, and 10+ more

3.3 System Resources

- **RAM**: 8GB+ recommended (4GB minimum for sample mode)
- **Disk Space**: 15GB free
- **Time**: 2-3 hours full mode, 10-15 minutes sample mode

4 Verification Steps

Before running:

4.1 For Stata Structure

`../sources/call-reports-historical.dta` exists

`../sources/call-reports-modern.dta` exists

`../sources/FDIC/` folder exists

Working directory is `code/` folder

4.2 For R Structure

`../2_input_data/Sources/` folder exists

All data files copied to `Sources/`

Working directory is `code/` folder

5 Troubleshooting

5.1 Common Issues

"Cannot find file: sources/.../[filename]"
Fix: Verify file exists and path is correct. Check spelling matches exactly.

"Cannot detect directory structure"
Fix: Ensure `code/` is next to `sources/` or `2_input_data/`
Run: `list.files("..")` in R to see parent directory contents

"Package 'xxx' not available"
Fix: Run `source("install_packages.R")` from `code/` directory

6 Outputs Generated

After successful run:

6.1 Figures (16 PDFs in output/figures/)

- `04_prob_failure_growth_all.pdf` — Key finding
- Failure time series
- Coefficient plots
- Recovery rates
- And 12 more...

6.2 Tables (4 LaTeX files in output/tables/)

- `04_assets_in_failure_assessed.tex`
- Descriptive statistics
- Recovery rates
- Rho-V analysis

6.3 Datasets (in dataclean/)

- `combined-data.dta` (2.87M observations)
- `temp_reg_data.dta` (regression dataset)
- Additional intermediate files

7 Next Steps

1. Verify data structure matches requirements
2. Set working directory to `code/` folder
3. Run `source("RUN_ALL.R")` for full mode
4. OR run `source("RUN_ALL_SAMPLE.R")` for fast test
5. Monitor console output for progress
6. Check outputs in `output/figures/` and `output/tables/`

8 Additional Documentation

- **Quick_Start_Guide.pdf** — How to run scripts
- **Variable_Definitions.pdf** — Data dictionary (70+ variables)
- **Technical_Documentation.pdf** — Detailed technical specs
- **Validation_Report.pdf** — R vs Stata comparison

For questions, see Technical_Documentation.pdf or contact authors