# Forecasting Economic Recovery After Financial Crises

Yelebe Desta

## Project

This project investigates how long it takes countries to recover economically after a financial crisis and whether macro-financial conditions at the onset of a crisis can be used to predict recovery time.

Using historical macroeconomic data, the goal is to forecast the number of years it takes for a country's real GDP per capita to return to its pre-crisis peak following a banking crisis. Rather than focusing on predicting whether a crisis occurs, this project shifts attention to the post-crisis recovery phase, which is central to policy design, fiscal planning, and understanding economic resilience.

The project is primarily predictive: models are trained on past crisis episodes and evaluated on unseen, more recent crises. At the same time, the structure of the final model allows for interpretation of which macro-financial variables are most strongly associated with faster or slower recoveries.

The project is located in the assignments folder under the final project folder.

## Modeling Approach

The project uses the Jordà–Schularick–Taylor (JST) Macrohistory Database, which contains annual macro-financial data for 17 advanced economies from 1870–2016. The dataset includes GDP, credit, housing prices, interest rates, inflation, money supply, and government debt.

Each crisis episode is defined using a banking crisis indicator, and recovery time is measured as the number of years it takes for real GDP per capita to return to its pre-crisis peak.

## Feature Engineering

Predictors are constructed from macroeconomic conditions at the crisis onset year, including:

- Growth rates of credit, money supply, housing prices, and CPI
- Yield curve slope (long-term minus short-term interest rates)
- Global averages of credit growth and yield curve slope

- Government debt-to-GDP level and change

These features are engineered consistently for training, validation, testing, and prediction.

## Models Evaluated

Three regression models are trained and compared:

- Ridge Regression
- Lasso Regression
- Random Forest Regressor

Hyperparameters are selected using cross-validation on the training set. Model selection is based on validation mean absolute error (MAE), and final performance is reported on a held-out test set.

- Training set: earlier historical crises
- Validation set: intermediate crisis episodes
- Test set: most recent crisis episodes

This ensures the model is evaluated in a realistic forecasting setting.

**To use the models created:**

1. **Clone the repository and make sure you have python installed on your computer**
2. **Train and save the models by running the TrainSave_JST_recovery notebook.**

This notebook:

- cleans and preprocesses the JST dataset
- constructs crisis-level observations
- trains multiple models with cross-validation

- selects the best model based on validation MAE
- saves the trained model and preprocessing metadata to model folder
- generates evaluation plots and tables in output

3. **Make a prediction using the Excel template**

Open: Raw Data/Excel_template.xlsx

This project does not predict directly from raw levels. Instead, it uses growth rates and changes, which require *at least two consecutive years of data*.

The two rows represent:

- Year $t_0 - 1$: pre-crisis year (crisisJST = 0)
- Year $t_0$: crisis onset year (crisisJST = 1)

This allows the code to compute growth rates (e.g., credit growth, inflation) in exactly the same way as during model training. Although the Excel file contains two rows, they correspond to one crisis episode, which is the single prediction target.

Editing the template

- Keep all column names unchanged
- Use one country and two consecutive years
- Set crisisJST = 1 only in the crisis year
- Use realistic year-to-year changes (moderate growth or contraction)

**4. Run the prediction**

Open and run: Code/Predict_JST_recovery.ipynb

This notebook:

- loads the trained model and saved preprocessing objects
- applies the same cleaning and feature engineering steps

- generates a predicted recovery time (in years)
- provides an interpretation (fast / moderate / slow recovery)

## Results

The final selected model is Ridge regression, which outperformed Lasso and Random Forest on validation data.

- Test MAE: ~3.1 years
- Test RMSE: ~3.7 years

This means the model predicts recovery time within approximately three years on average, which is reasonable given the complexity and noise inherent in macroeconomic recoveries.

A naive baseline model that predicts the historical mean recovery time performs worse, confirming that macro-financial conditions at crisis onset contain meaningful predictive information.

## Feature Importance

For the Ridge model, the strongest predictors of recovery time include:

- Changes in government debt-to-GDP
- Credit growth and global credit conditions
- Yield curve slope
- Housing market dynamics

These results align with economic intuition and prior literature on post-crisis recoveries.

This project demonstrates how historical macro-financial data can be used not only to study crises, but to forecast recovery trajectories in a transparent and reproducible way. The pipeline is designed to be extendable to additional countries, alternative recovery definitions, or future crisis scenarios.

## Possible Future Improvements

While this project demonstrates that macro-financial conditions at crisis onset contain predictive information about recovery time, there are several meaningful directions for future improvement and extension.

First, the recovery definition could be expanded beyond a single threshold based on real GDP per capita. Alternative definitions like partial recovery (e.g., reaching 90% of the pre-crisis peak), employment-based recovery, or multi-dimensional recovery metric could provide a more nuanced view of post-crisis dynamics. Incorporating uncertainty around recovery timing, rather than predicting a single point estimate, could also improve interpretability for policymakers..

Second, additional predictors could be incorporated. These might include fiscal policy responses, monetary policy interventions or exchange rate regimes. Including crisis-specific policy variables could help distinguish between recoveries driven by numerical fundamentals and those shaped by policy choices.

Overall, these extensions would build on the existing pipeline while preserving the project's core to use historical data to generate transparent and policy-relevant predictions about economic recovery after financial crises.