Public Trust in Banks and Financial Institutions

A Knowledge Mining Approach

Presented By: Soha Arian, Kasra Akbari, Eman Ajmal, Joseph Martinez

Meet The Team

Soha Arian

Kasra Akbari

Eman Ajmal

Joseph Martinez

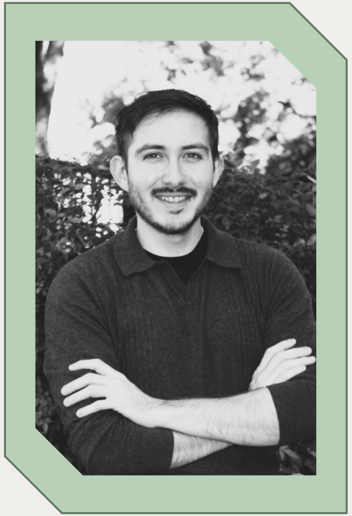


Table Of Contents

Executive Summary

This study analyzes data from the General Social Survey (1972–2024) to understand how confidence in banks varies by income, education, and age in the United States.

To ensure fair comparisons over time, income is grouped within each year. Confidence is measured as an ordered category.

The analysis uses weighted descriptive statistics and ordered logistic regression to highlight demographic differences and account for trends over time.

After completing the statistical analysis, the results are summarized by an artificial intelligence, and we evaluate whether the AI summary accurately represents the findings.

Study Objective

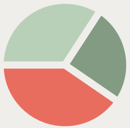
- Use the General Social Survey (1972–2024) to analyze self-reported confidence in banks
- Compare confidence levels across groups defined by income, education, and age
- Estimate how these demographic factors are associated with confidence using weighted descriptive statistics and ordered logistic regression
- Test whether the relationship between income and confidence changes over time
- Check if AI-generated summaries accurately reflect the statistical results

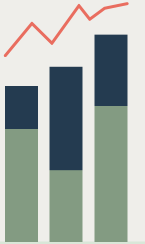
What We Are Measuring

- **Main outcome:** Confidence in banks, as reported in the General Social Survey
- **Original responses:** “A Great Deal,” “Only Some,” and “Hardly Any”
- **Recoded into three categories:** “High Confidence,” “Moderate Confidence,” and “Low Confidence”
- Treated as an ordered categorical variable for analysis

Predictor Variables

- **Income:** grouped into thirds within each year to show relative economic rank
- **Education:** measured by the highest degree completed
- **Age:** used to capture life-cycle differences at the time of the survey
- **Survey year:** included as a categorical variable to reflect historical context
- **Cohort effects** are not estimated separately because they cannot be separated from age and year in this data





What Does “Income” Represent?

- Income is divided into three groups within each survey year
- This approach measures a person’s rank compared to others in the same year, not their actual income
- As a result, “high income” in 1975 is not the same as “high income” in 2024

How Do We Solve This Weakness?

- Decide if the main focus is on relative status or actual income
- Choose one income model as the primary approach
- Compare results using income terciles and inflation-adjusted income categories
- Explain how the meaning of results changes depending on which model is used
- Discuss how interpretation shifts with each way of measuring income

Income Construction

- Income categories are not the same in every survey year
- The main measure is income terciles within each year, because relative status may affect trust in banks
- As a secondary measure, we use inflation-adjusted income to reflect absolute resources
- We compare both approaches to see if our conclusions change depending on how income is defined

Sample Selection

- Only include respondents who have valid answers for confidence, income, education, and age
- Exclude cases with missing data in the main analysis and report how many are removed
- Run a sensitivity analysis to check if missing data affects the results
- Use GSS survey weights to make the sample nationally representative
- Report the final sample size (N)

Software

- All analyses are done in R Data cleaning steps are scripted and documented
- All variable recoding is specified in detail
- Ordered logistic regression is used for modeling
- Predicted probabilities are calculated from the model results
- The code, AI prompts, and scoring rubric are included in the appendix

Descriptive Analysis

- Calculate weighted distributions of confidence levels
- Compare these distributions across income, education, and age groups
- Measure how large the differences are between demographic groups
- Identify how confidence changes over time
- Use charts and graphs to help explain the results

Regression Analysis

- The main model is ordered logistic regression, which is appropriate because confidence is an ordered categorical variable
- We test the proportional odds assumption to check if the model fits
- If this assumption does not hold, we use a multinomial model as a sensitivity check
- All results are interpreted as associations, not as evidence of cause and effect

Time Is Treated As A Control

- Survey year is included in the model to account for changes over time
- The data cover 1972 to 2024, which includes major events like banking deregulation, the 2008 financial crisis, and the COVID-19 economic shocks
- Including year as a variable does not fully capture sudden changes, generational effects, or how the impact of income may change in different periods

How Do We Solve This Weakness?

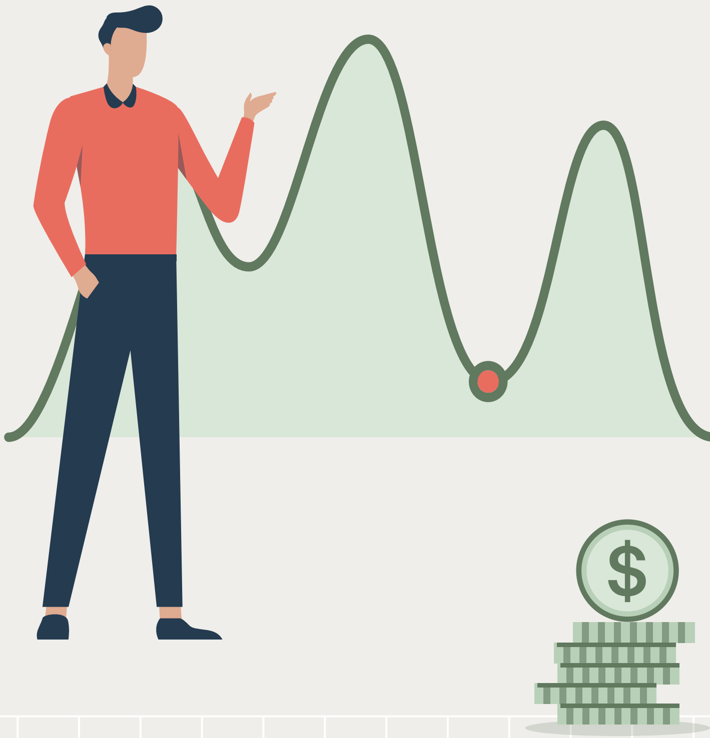
- Survey year is included as a categorical variable to allow for differences between specific periods
- We add an income by year interaction to test if the effect of income changes in different years
- We compare models before and after 2008 to see if there are major shifts in the results

Model Testing

- The main analysis includes an income by year interaction
- We also run models without this interaction for comparison
- The proportional odds assumption is tested, and if it does not hold, we use a multinomial model GSS survey weights are applied to ensure national representation
- We check if the interaction terms are stable across models
- Predicted probabilities are graphed to help interpret the results

AI Evaluation Procedure

- Export the model coefficients and predicted probabilities
- Give a standardized prompt to the AI to generate a summary
- Create a benchmark summary based only on the model results
- Two independent reviewers compare the AI summary to the benchmark
- Accuracy is measured using specific error criteria



Scoring Criteria

- Each summary is scored on four criteria: unsupported causal claims, exaggeration of effect size, omission of statistical uncertainty, and use of biased or normative language
- Each criterion is scored as 0 or 1, with clear written definitions
- We check how much the reviewers agree with each other
- The frequency and overall error rate are reported

Relevance and Limitations

- This study analyzes how different demographic groups report trust in banks
- We use national survey data and statistical models to identify patterns
- The evaluation of AI-generated summaries is transparent and systematic
- The analysis highlights which groups have lower confidence in banks
- Because the study is cross-sectional, we cannot make strong claims about cause and effect
- Confidence is self-reported, which may affect accuracy
- The long time span may hide short-term changes
- The meaning of confidence may change over the decades

References

Grable, John E., Eun Jin Kwak, and Kristy L. Archuleta. “Distrust of Banks among the Unbanked and Banked.” *International Journal of Bank Marketing*, vol. 41, no. 6, 2023, pp. 1498–1520. https://doi.org/10.1108/IJBM-10-2022-0441.

Crosignani, Matteo, et al. “Financial Inclusion in the United States: Measurement, Determinants, and Recent Developments.” *SSRN Electronic Journal*, 2025, https://doi.org/10.2139/ssrn.5534239.

General Social Survey. “Get the Data in Stata Format.” *GSS – NORC at the University of Chicago*, https://gss.norc.org/get-the-data/stata.html. Accessed 26 Feb. 2026.