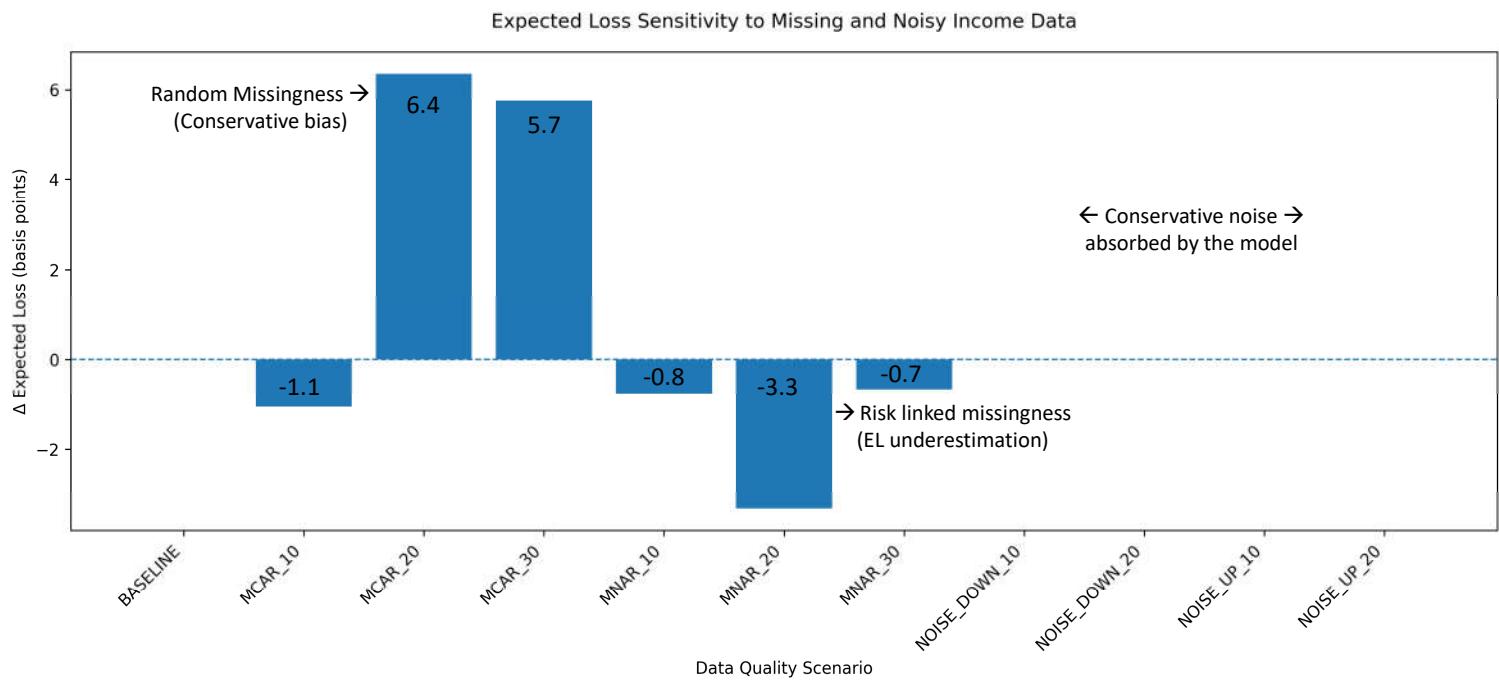
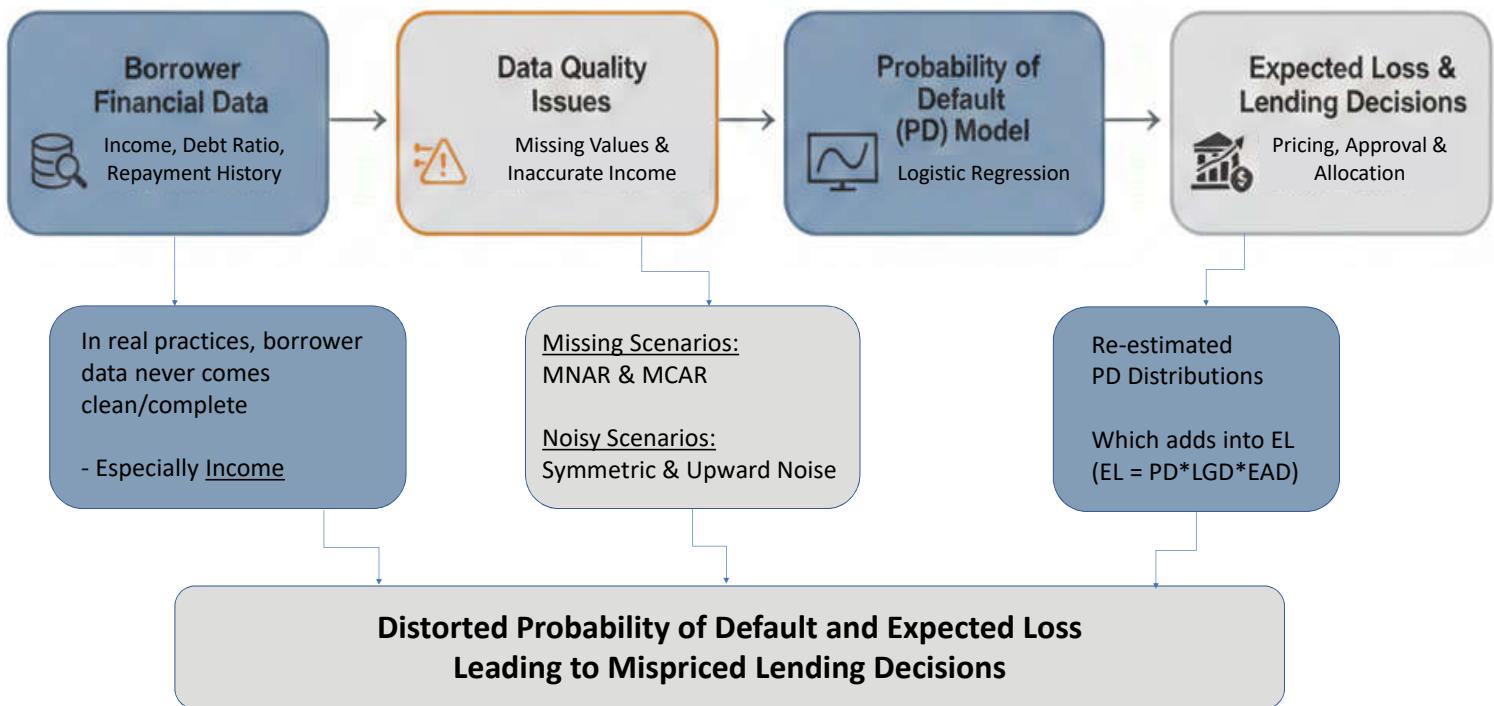


IMPACTS OF MISSING AND NOISY FINANCIAL DATA ON PROBABILITY OF DEFAULT AND EXPECTED LOSS ESTIMATION

DURGESH MISHRA & OM JADHAV

DATA QUALITY AND CREDIT RISK DEGRADATION



Methodology & Experimental Design

- Retail consumer credit risk dataset
- Baseline PD model: Logistic Regression
- Risk metric: Expected Loss = $PD * LGD * EAD$
- Data quality scenarios:
 - Missing Income: MCAR vs MNAR
 - Noisy Income: under and over reporting
- Evaluation: Δ Expected Loss (basis points) vs clean baseline

Key Findings and Implications

- Random missing income creates conservative bias in EL
- Risk-linked missingness (MNAR) causes systematic underestimation at moderate levels
- Optimistic noise > conservative noise in risk impact
- Data Assumptions directly shape risk estimates
- Missingness mechanism matters more than volume

Takeaway:

How data fails matters more than how much data is missing