# The Application of Extreme Value Theory to Operational Loss Data

Jason Perry
Kabir Dutta

The views expressed in this paper do not necessarily reflect those of the Federal Reserve System.

# Outline

- What is operational risk, and why is it important?
- How do banks measure operational risk?
- The 2004 Loss Data Collection Exercise (LDCE) data
- Methodology
- Some preliminary results

# Operational Risk

- “The risk of loss resulting from inadequate or failed internal processes, people, and systems or from external events.”
- 7 loss types including Fraud, Business Practices, Errors, System Failures, Physical Damage
- Consequences include reputational damage and systemic impacts

# Examples of Operational Losses

- More than 100 losses exceeding $100 Million over the last decade
- Well-known losses:
  - Daiwa, Allfirst, Household Finance, Bankers Trust, Drexel, Kidder Peabody, 9/11, ...
- Some more unusual losses:
  - $21M cash stolen from Republic Bank (1991)
  - Firstar suffers $22M in overdrafts from one customer (1994)
  - A man posing as a Phillip Morris executive obtained $320M in financing from several banks (1996)

# Operational Risk Under Basel II

- Basel II takes more granular approach than Basel I, including a separate capital charge for operational risk
- In the US, operational risk will be estimated based on internal models
- Capital is 99.9th percentile of the aggregate loss distribution

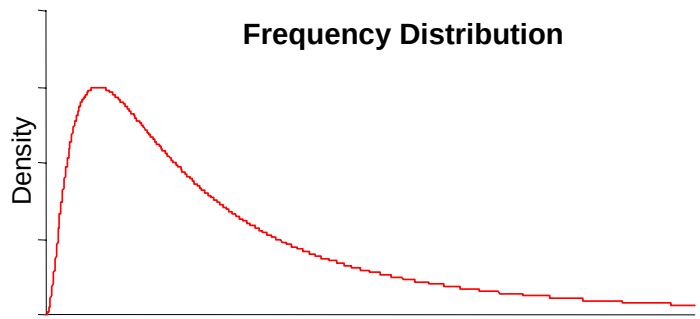
# Measuring Operational Risk under Basel II

- Basic Indicator Approach
- Standardized Approach
- Advanced Measurement Approach (AMA) has 4 components:
  - Internal Data
  - External Data
  - Scenario Analysis
  - Business Environment and Internal Control Factors

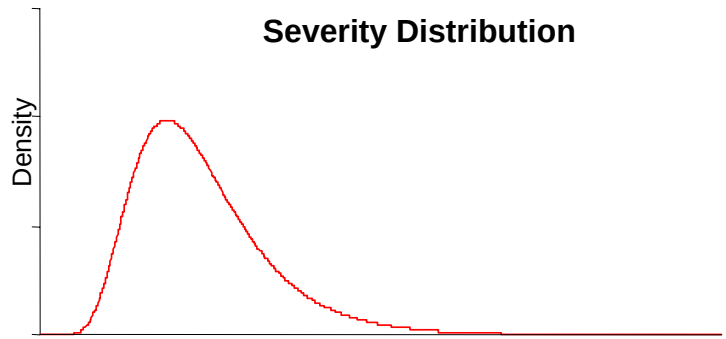
# Loss Distribution Approach

- Severity distribution
  - The distribution of loss amounts
  - e.g., Lognormal or Pareto
- Frequency distribution
  - The number of losses per year
  - e.g., Poisson distribution
- Aggregate loss distribution
  - Usually obtained via Monte Carlo simulation

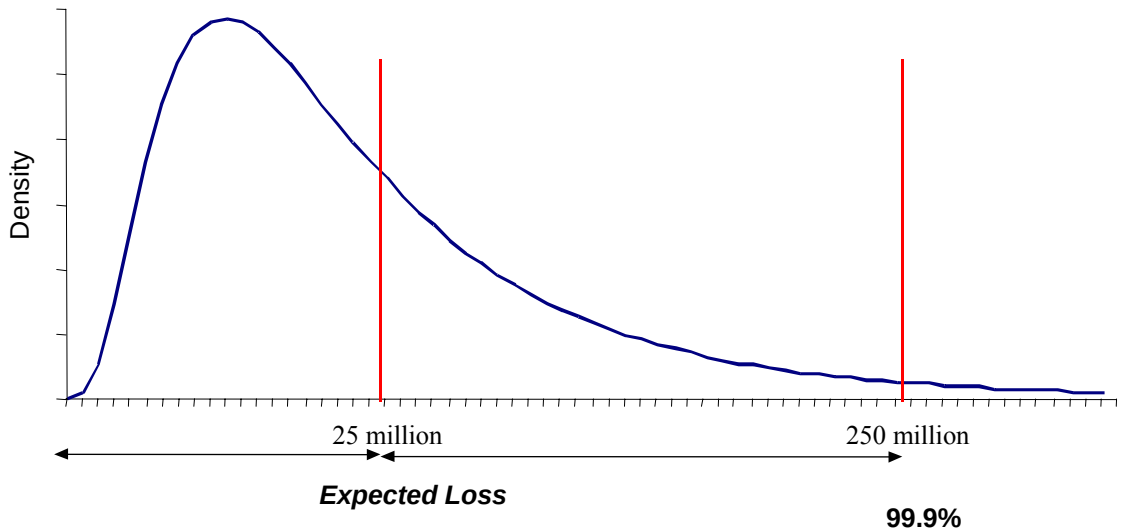
# Loss Distribution Approach




Number of Loss Events per Year




$ Value of a Loss Events




**Total Operational Loss over a 1 year time horizon.**

# The LDCE Data (2004)

- Loss Data Collection Exercise was initiated by US regulatory agencies
- Previous LDCEs carried out on an international level
- A unique dataset containing internal operational loss experience from 23 institutions with a US presence
- We selected 8 banks with more than 1,000 total losses (see TABLE)

# Data Characteristics and Challenges

- Data quality appears to vary across institutions
  - Loss thresholds
  - Rounding of loss amounts
  - Length of time series available
  - Accuracy of loss timestamps
  - Uneven loss reporting
- Heavy-tailed loss severity

# LDCE Sample Descriptive Details

**Number of Individual Losses Reported by 2004 LDCE Participants**

| Number of Losses of $10,000 or More | Number of Participants | Number of Losses of 10,000 or More | Total Number of Losses | Total Loss Amount ($M) | Comprehensiveness of Loss Data[1] | | |
|---|---|---|---|---|---|---|---|
| | | | | | Fully Comprehensive | Partially Comprehensive | No Information Provided |
| 0-250 | 6 | 640 | 134,679 | 212 | 2 | 2 | 2 |
| 250-1,000 | 5 | 2,253 | 6,125 | 283 | 2 | 1 | 2 |
| 1,000-2,500 | 8 | 13,404 | 43,814 | 8,151 | 5 | 1 | 2 |
| 2,500+ | 4 | 39,469 | 1,342,147 | 17,275 | 1 | 3 | 0 |
| Total | 23 | 55,766 | 1,526,765 | 25,920 | 10 | 7 | 6 |

# Research Questions

- Our main question: Do operational losses display similar statistical properties across different institutions?
- Does the application of classical EVT yield parameter estimates that appear reasonable and consistent?
- What do these parameter estimates imply about capital?
- The answers are of interest for multiple reasons: scientific, policy, and implementation

# Previous Research on 2002 LDCE

- de Fontnouvelle, et al. (2004)
- Moscadelli (2004)

# Simplifying Assumptions

- Frequency distributions independent
- Severity distributions independent
- Frequency independent from severity
- Frequency is distributed Poisson
- Loss severities are identically distributed

$\Rightarrow$ Losses follow a multivariate compound Poisson process

# Modeling the Severity Distribution.

We draw on extreme value theory (EVT):

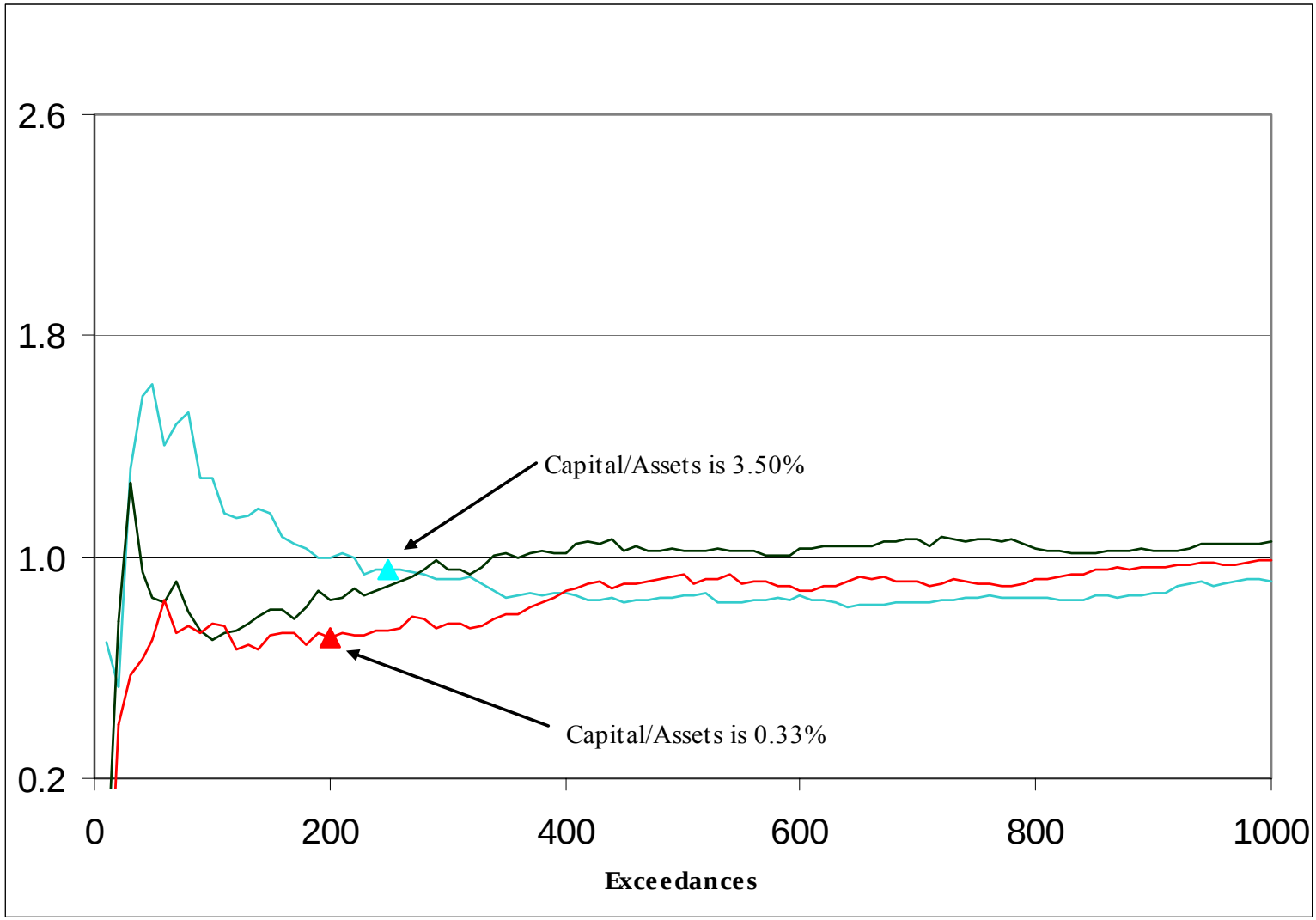
- Using the peaks over threshold (POT) approach, we model excess losses over a high threshold.
- Pickands-Balkema-de Haan Theorem: (under certain conditions) the limit distribution of scaled excesses over a high threshold is a generalized Pareto distribution (GPD):

$$\mathrm{GPD}_{\xi,b}(x) = \begin{cases} 1 - (1 + \xi x/b)^{-1/\xi} & \xi > 0 \\ 1 - \exp(-x/b) & \xi = 0 \end{cases}$$

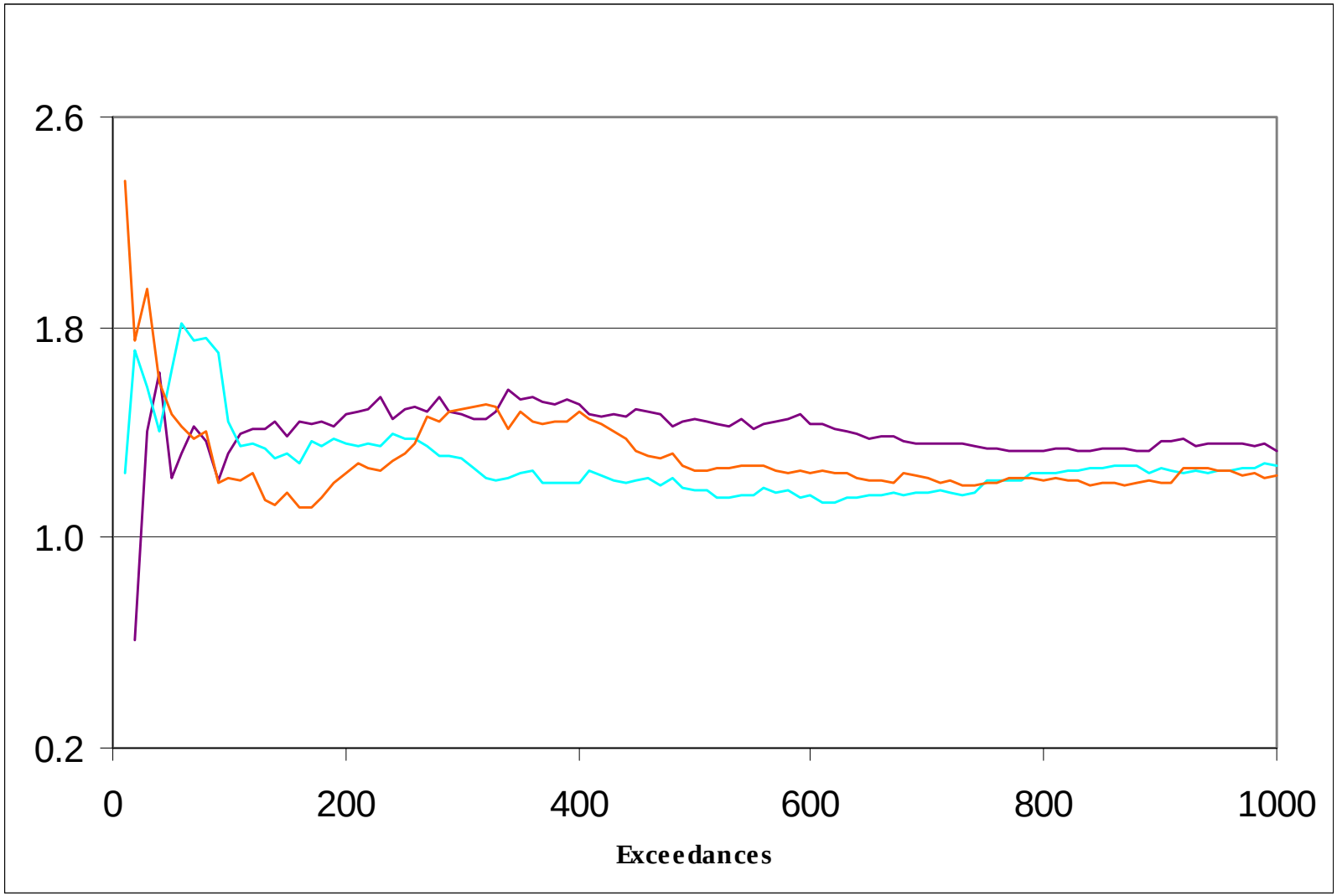
# Simulating Capital

1. Choose a threshold $\tau$
2. Fit GPD to excess losses above $\tau$ using MLE
3. Draw a frequency N from Poisson
4. Compute $N_L$ and $N_H$ from data
5. Draw $N_L$ loss severities from body (empirical sampling with replacement)
6. Draw $N_H$ losses from the fitted GPD tail
7. Sum $N_L$ and $N_H$ together to get annual loss
8. Repeat the above steps 1 million times

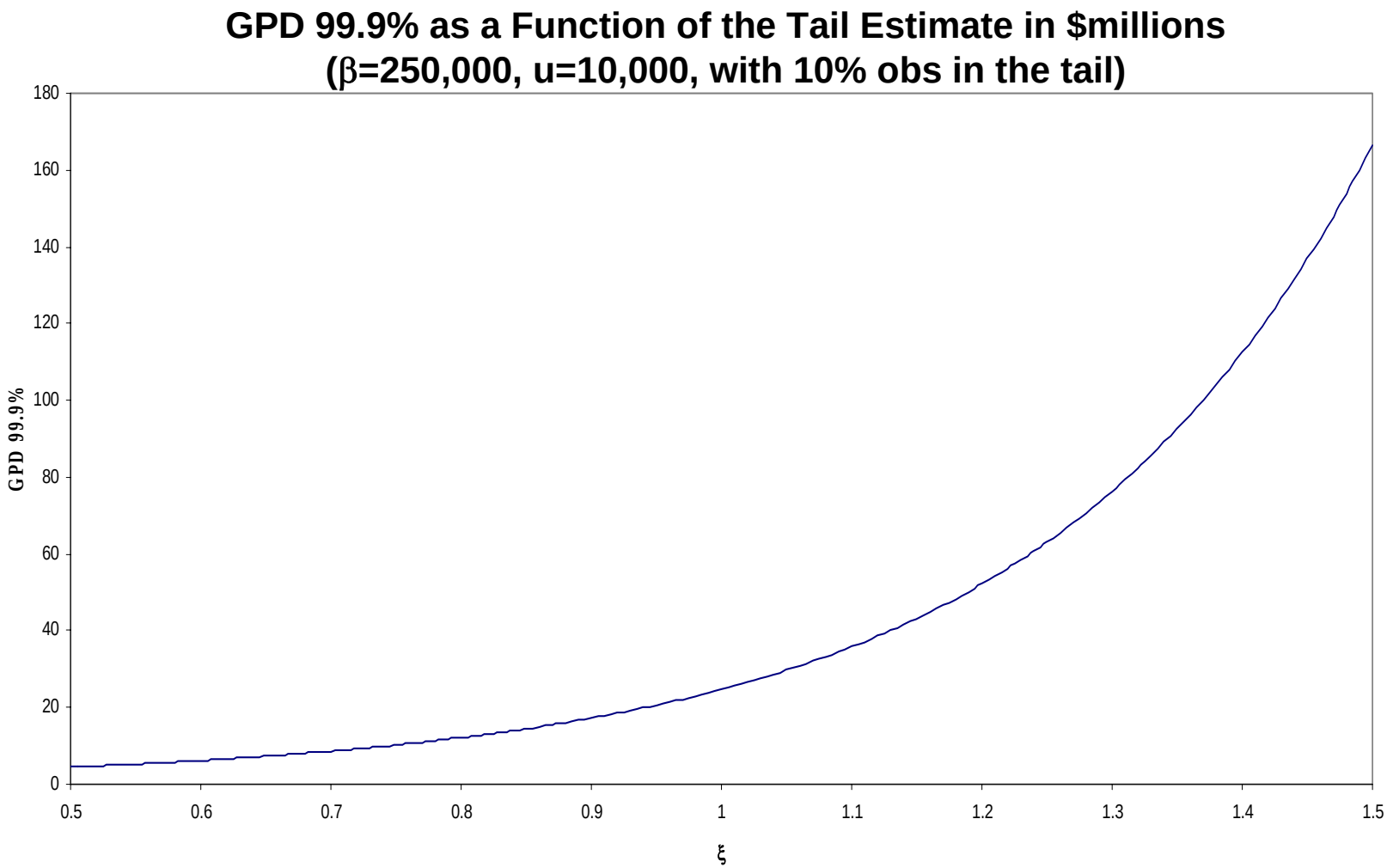
# GPD Tail Estimates By Exceedances for Three Institutions



# Tail Parameter Estimates Exceed 1 for Three Institutions



# GPD 99.9% Versus the Tail Estimate



# Key Observations

- The tail estimates for institutions are very different (although always implying a heavy-tailed loss severity distribution).
- Some banks have tail estimates above 1.
- These results hold at the business line level as well.

# Potential Explanations

- Loss severity distributions for some banks are not in the maximum domain of attraction of the GEV
- The GPD poorly approximates the tail in finite samples for these distributions
- Losses are not i.i.d.
- Problem of data heterogeneity