KU LEUVEN

Testing and improving the robustness of amortized Bayesian inference for cognitive models

Yufei Wu, Stefan T. Radev, Francis Tuerlinckx

March 27th 2025

KU LEUVEN

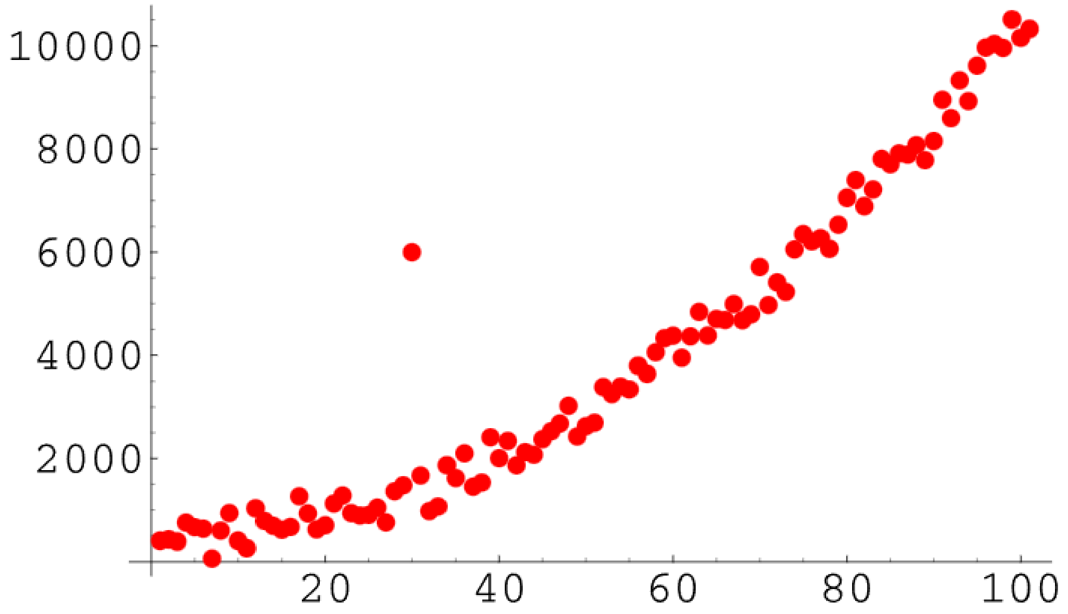
Testing and improving the robustness of amortized Bayesian inference for cognitive models

Yufei Wu, Stefan T. Radev, Francis Tuerlinckx

March 27th 2025

The presence of outliers

- Robustness refers to the resilience of an estimator to outliers.



Cognitive models

- Mathematical expressions of cognitive processes with interpretable parameters
- Some models are sensitive to outliers due to the nature of its assumptions
- E.g., the drift diffusion model (DDM)

Research question

- How to study the influence of outliers of complex stochastic cognitive models?
- How to robustify the inference of complex stochastic cognitive models?

1

Testing and improving the robustness of amortized Bayesian inference for Cognitive Models

Origin of robust statistics

ROBUST ESTIMATION OF A LOCATION PARAMETER[1]

BY PETER J. HUBER[2]

University of California, Berkeley

Which is more robust?

$$\boldsymbol{x} \sim N(\mu, 1^2)$$

Original Data	Contaminated Data
1.66	1.66
1.82	1.82
0.43	0.43
1.37	1.37
0.82	0.82
0.11	0.11
1.27	1.27
3.10	3.10
0.95	0.95
1.20	-100

$$Mean_1 = 1.27$$
$$Median_1 = 1.24$$

$$Mean_2 = -8.85$$
$$Median_2 = 1.24$$

Tools to assess robustness

- 1. Empirical influence function (Cook & Weisberg, 1982)
  - Add outlier x^c with different values
- 2. Breakdown point (Donoho & Huber, 1982)
  - Add outlier x^c with different fractions

Empirical influence function

Original Data
1.66
1.82
0.43
1.37
0.82
0.11
1.27
3.10
0.95
1.20

$Mean_1 = 1.27$

Contaminated Data
1.66
1.82
0.43
1.37
0.82
0.11
1.27
3.10
0.95
-100

x^c

$Mean_2 = -8.85$

$$\begin{aligned}
&\mathrm{EIF}_{-100} \\
&= Mean_2 - Mean_1 \\
&= -8.85 - 1.27 \\
&= -10.12
\end{aligned}$$

Empirical influence function

Original Data
…
…
…
…
…
…
…
…
…
…

$Mean_1$

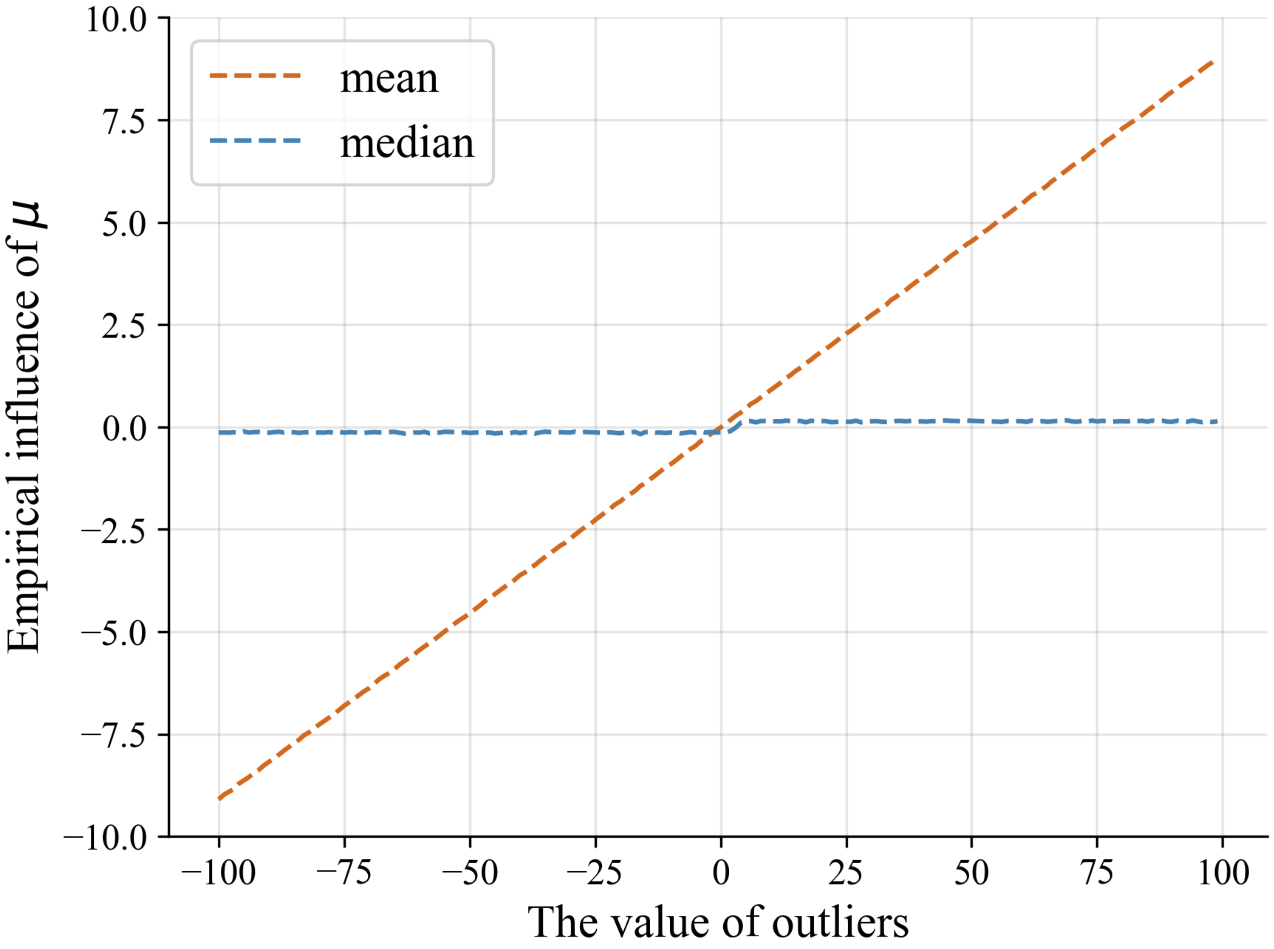
Contaminated Data
…
…
…
…
…
…
…
…
…
-100

x^c

$Mean_2$

$$\overline{\text{EIF}}_{-100} = \overline{Mean_1 - Mean_2}$$

Empirical influence function



For each $x^c = k$,
200 datasets are simulated.

Tools to estimate robustness

- 1. Empirical influence function (Cook & Weisberg, 1982)
  - Add outlier x^c with different values
- 2. Breakdown point (Donoho & Huber, 1982)
  - Add outlier x^c with different fractions

Breakdown point

Contaminated Data
1.66
1.82
0.43
1.37
0.82
0.11
1.27
3.10
-100
-100

x^c

$Mean = -18.9$

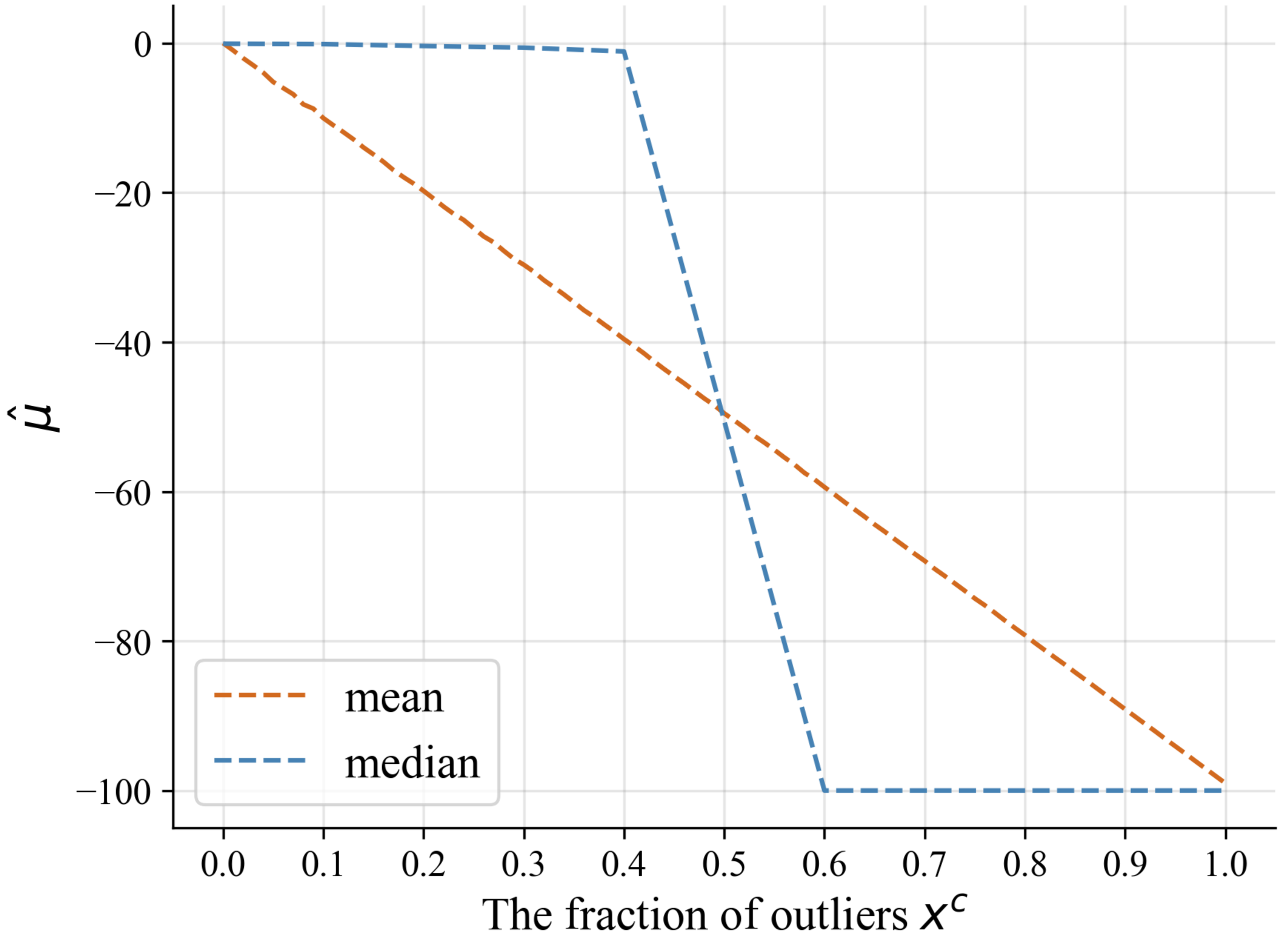
Breakdown point

Contaminated Data
…
…
…
…
…
…
…
…
-100
-100

x^c

$\overline{Mean}$

Breakdown point ($x^c = -100$)

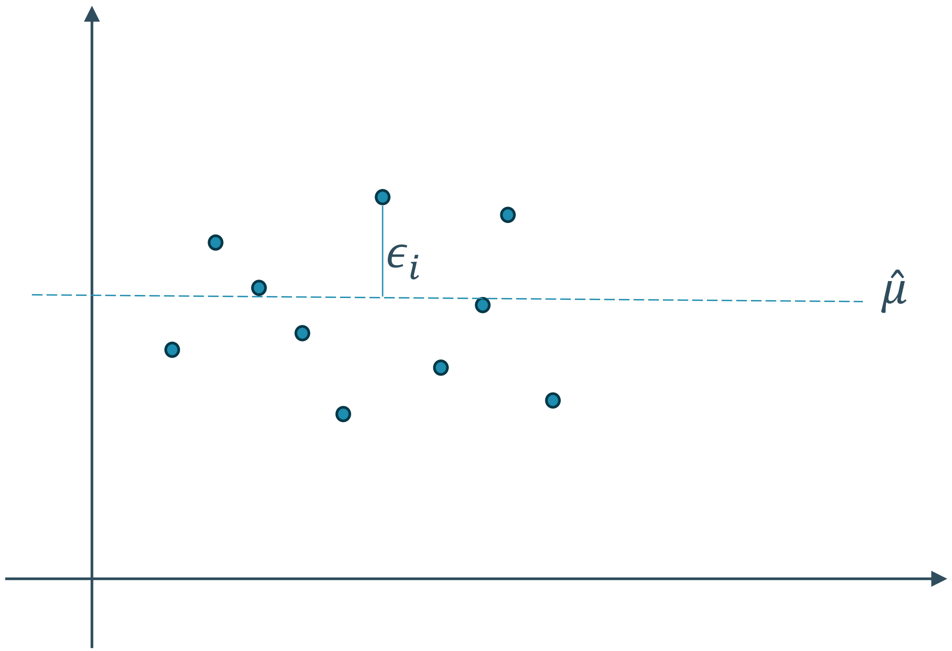


Tools to estimate robustness

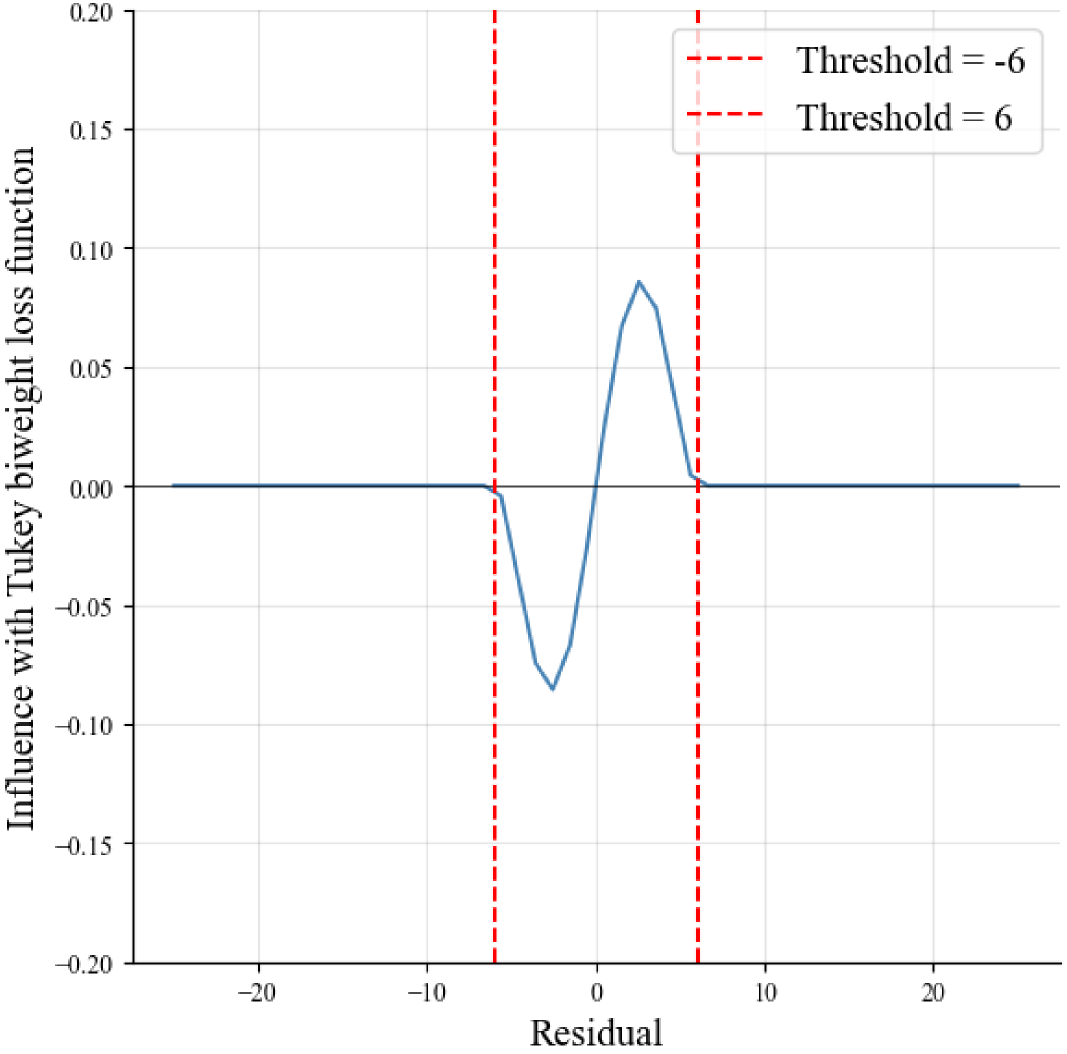
- 1. Empirical influence function (Cook & Weisberg, 1982)
    - Add outlier x^c with different values
- 2. Breakdown point (Donoho & Huber, 1982)
    - Add outlier x^c with different fractions

How to make $\hat{\mu}$ more robust?

- **Minimizing the loss function:** $\sum_{i=1}^{n} \rho(\epsilon_i)$
    - **Mean:** $\rho(\epsilon_i) = \epsilon_i^2$
    - **Median:** $\rho(\epsilon_i) = |\epsilon_i|$
    - **Tukey's biweight function** (Tukey, 1979):
    - $\rho(\epsilon_i) = \begin{cases} \left(1 - \left(\frac{\epsilon_i}{k}\right)^2\right)^2, & if\ |\epsilon_i| \leq k \\ 0, & if\ |\epsilon_i| > k \end{cases}$



Tukey’s biweight function



But…

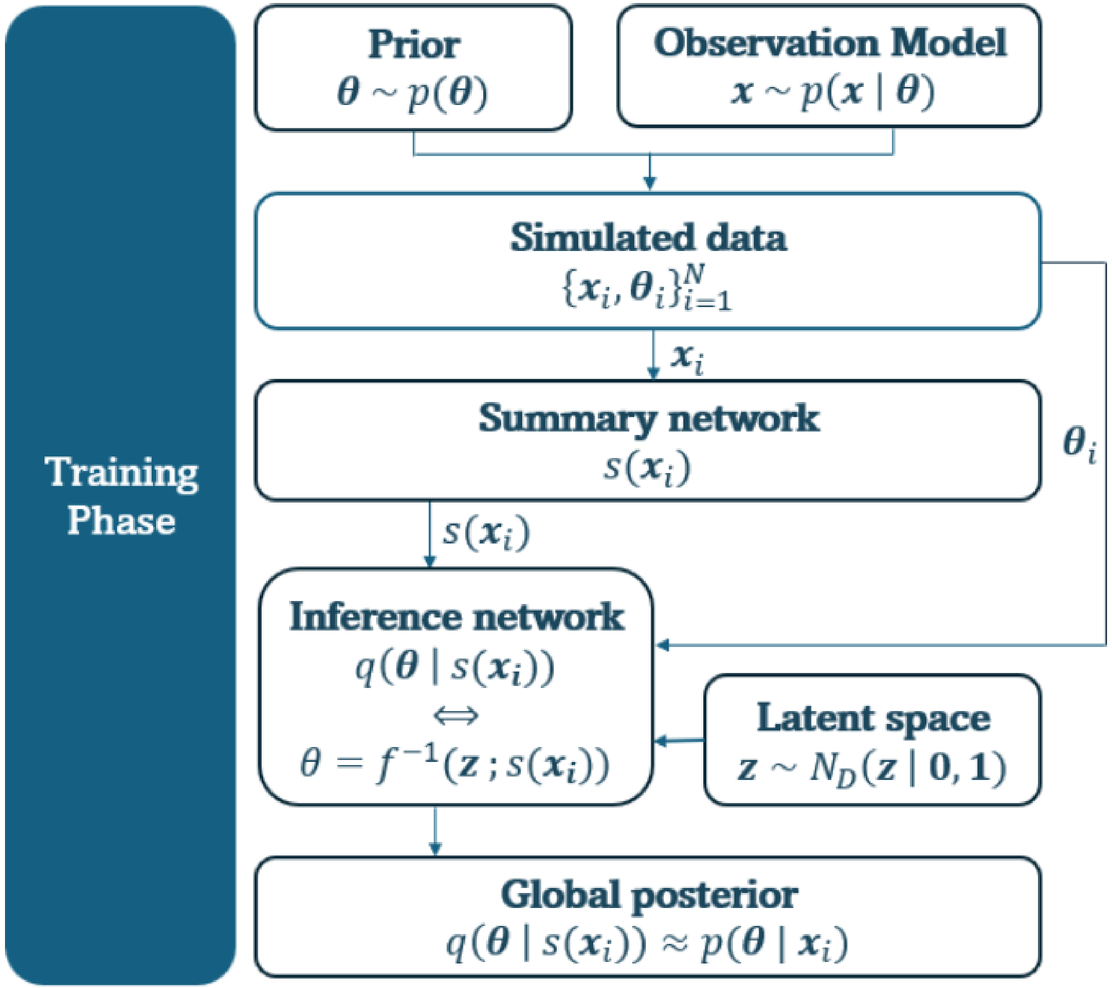
- 1. Modifying loss function can be difficult

Can we bypass the likehood-based framework?

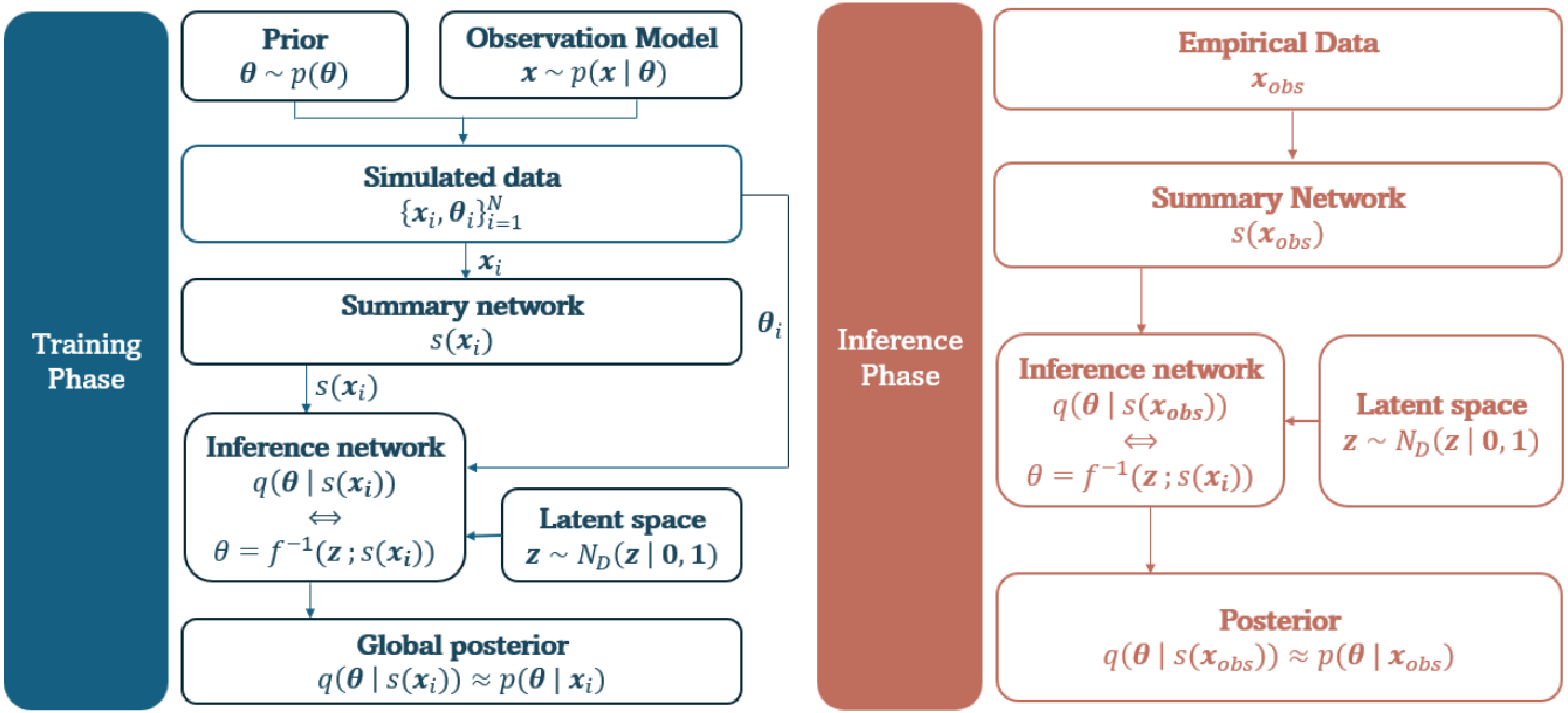
2

Testing and improving the robustness of amortized Bayesian inference for Cognitive Models

Training


Training Phase
Prior
$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$
Observation Model
$\boldsymbol{x} \sim p(\boldsymbol{x} \mid \boldsymbol{\theta})$
Simulated data
$\{\boldsymbol{x}_i, \boldsymbol{\theta}_i\}_{i=1}^N$
$\boldsymbol{x}_i$
Summary network
$s(\boldsymbol{x}_i)$
$\boldsymbol{\theta}_i$
$s(\boldsymbol{x}_i)$
Inference network
$q(\boldsymbol{\theta} \mid s(\boldsymbol{x}_i))$
$\Leftrightarrow$
$\theta = f^{-1}(\boldsymbol{z}\,; s(\boldsymbol{x}_i))$
Latent space
$\boldsymbol{z} \sim N_D(\boldsymbol{z} \mid \boldsymbol{0}, \boldsymbol{1})$
Global posterior
$q(\boldsymbol{\theta} \mid s(\boldsymbol{x}_i)) \approx p(\boldsymbol{\theta} \mid \boldsymbol{x}_i)$

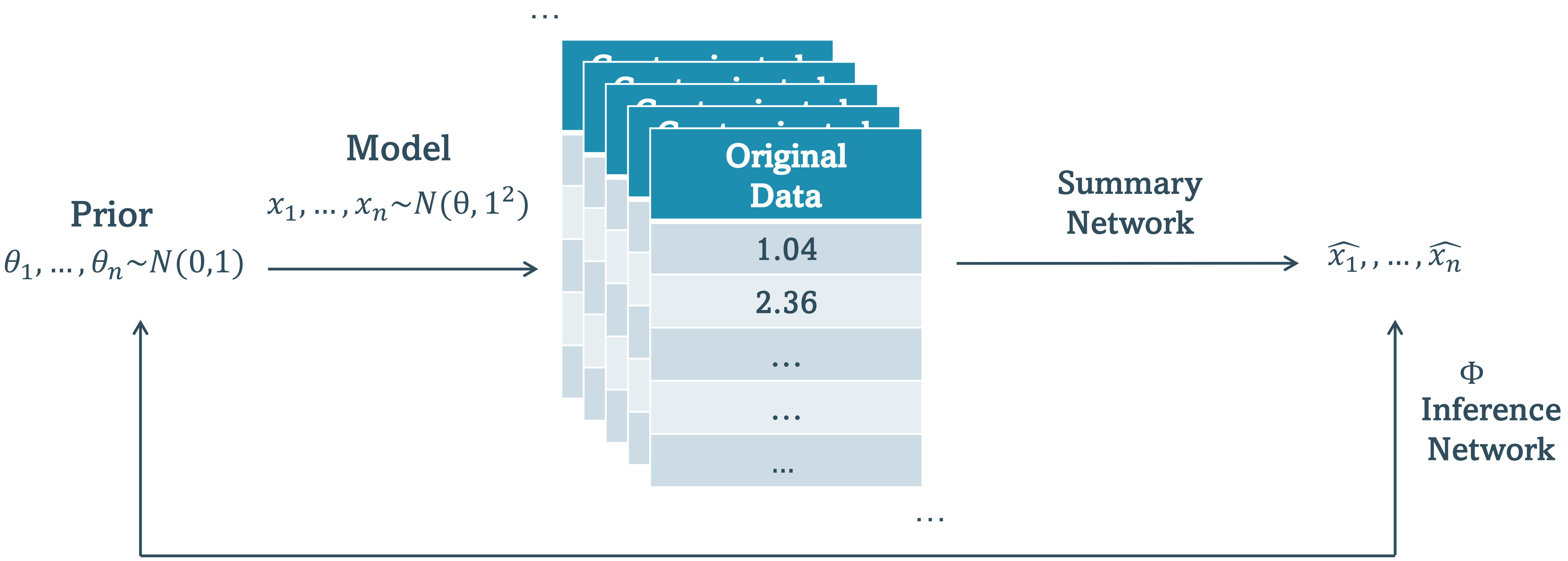
Inference


Training Phase
Prior
$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$
Observation Model
$\boldsymbol{x} \sim p(\boldsymbol{x} \mid \boldsymbol{\theta})$
Simulated data
$\{\boldsymbol{x}_i, \boldsymbol{\theta}_i\}_{i=1}^{N}$
$\boldsymbol{x}_i$
$\boldsymbol{\theta}_i$
Summary network
$s(\boldsymbol{x}_i)$
$s(\boldsymbol{x}_i)$
Inference network
$q(\boldsymbol{\theta} \mid s(\boldsymbol{x}_i))$
$\Leftrightarrow$
$\theta = f^{-1}(\boldsymbol{z}\,; s(\boldsymbol{x}_i))$
Latent space
$\boldsymbol{z} \sim N_D(\boldsymbol{z} \mid \boldsymbol{0}, \boldsymbol{1})$
Global posterior
$q(\boldsymbol{\theta} \mid s(\boldsymbol{x}_i)) \approx p(\boldsymbol{\theta} \mid \boldsymbol{x}_i)$
Inference Phase
Empirical Data
$\boldsymbol{x}_{obs}$
Summary Network
$s(\boldsymbol{x}_{obs})$
Inference network
$q(\boldsymbol{\theta} \mid s(\boldsymbol{x}_{obs}))$
$\Leftrightarrow$
$\theta = f^{-1}(\boldsymbol{z}\,; s(\boldsymbol{x}_i))$
Latent space
$\boldsymbol{z} \sim N_D(\boldsymbol{z} \mid \boldsymbol{0}, \boldsymbol{1})$
Posterior
$q(\boldsymbol{\theta} \mid s(\boldsymbol{x}_{obs})) \approx p(\boldsymbol{\theta} \mid \boldsymbol{x}_{obs})$

Advantages

- Expensive training, cheap inference
- Likelihood-free

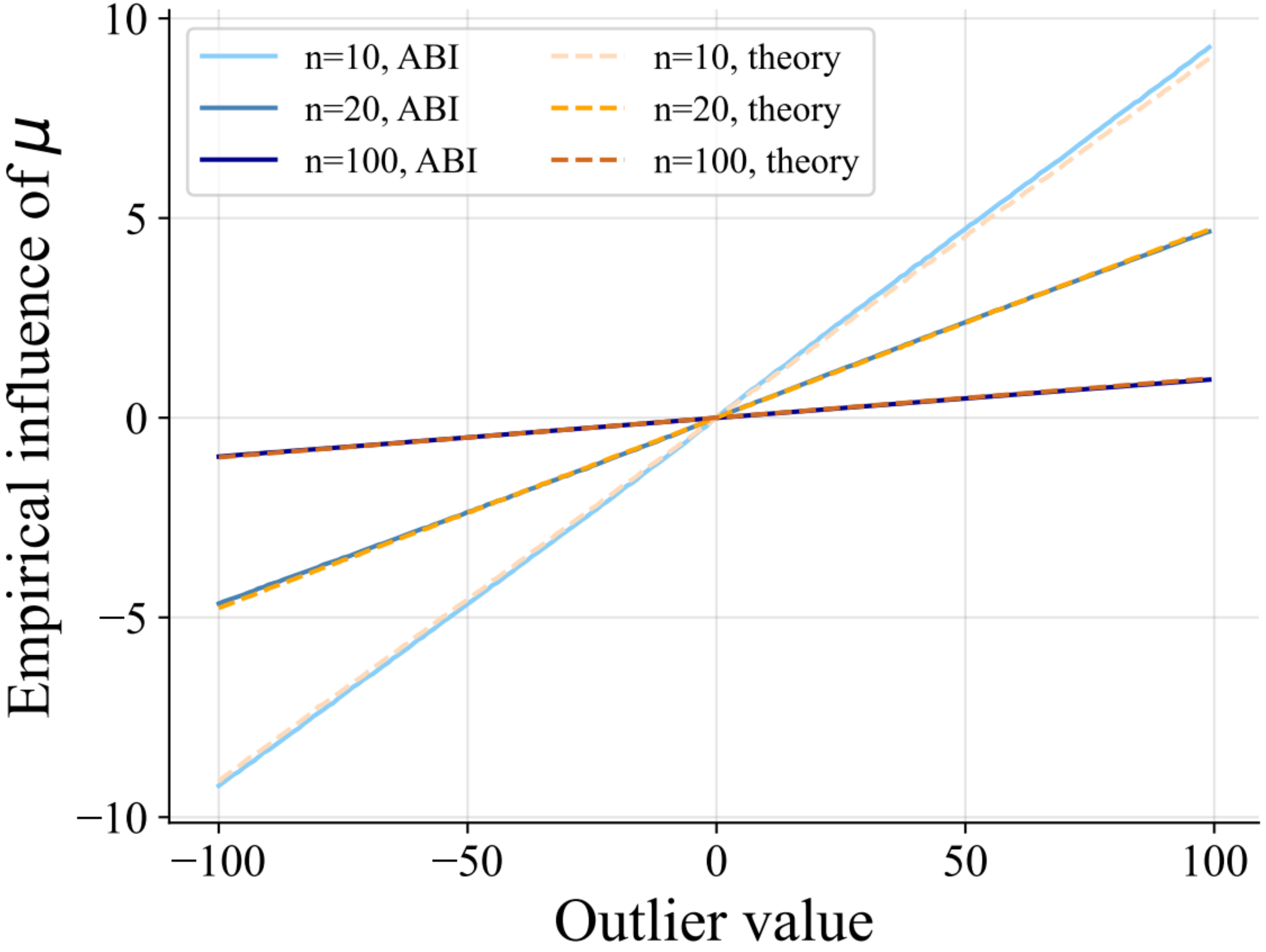
Standard estimator training



Standard estimator training

- Summary network: DeepSet(summary_dim=2) (other settings stay the same as in bayesflow 1.1.4/ bayesflow stable-legacy branch)
- Inference network: Normalizingflow (Couplingflow)

Empirical influence function



Assuming outliers in simulation

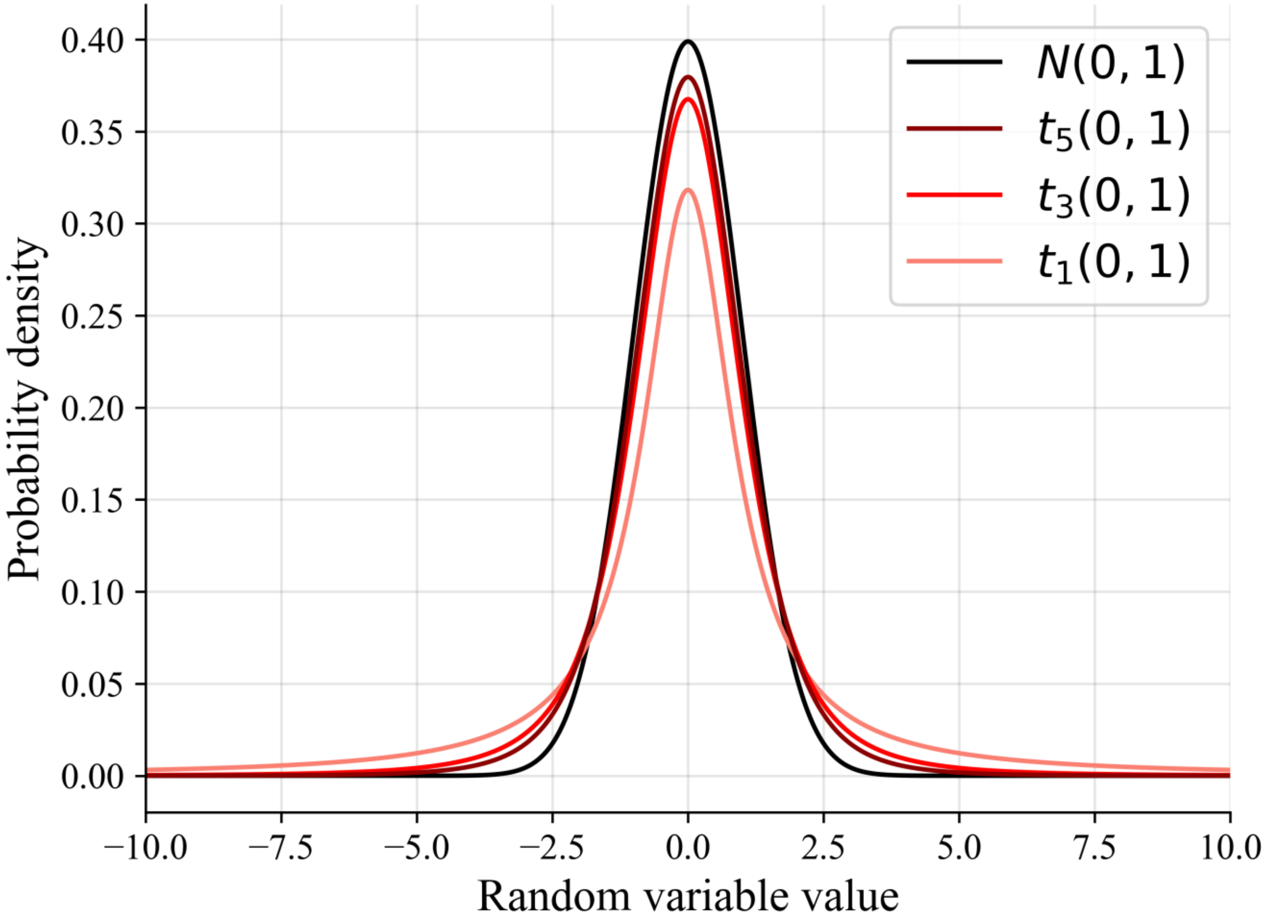
$\boldsymbol{x} \sim N(\mu, 1^2)$

Research Group of Quantitative Psychology and Individual Differences, PPW, KU Leuven

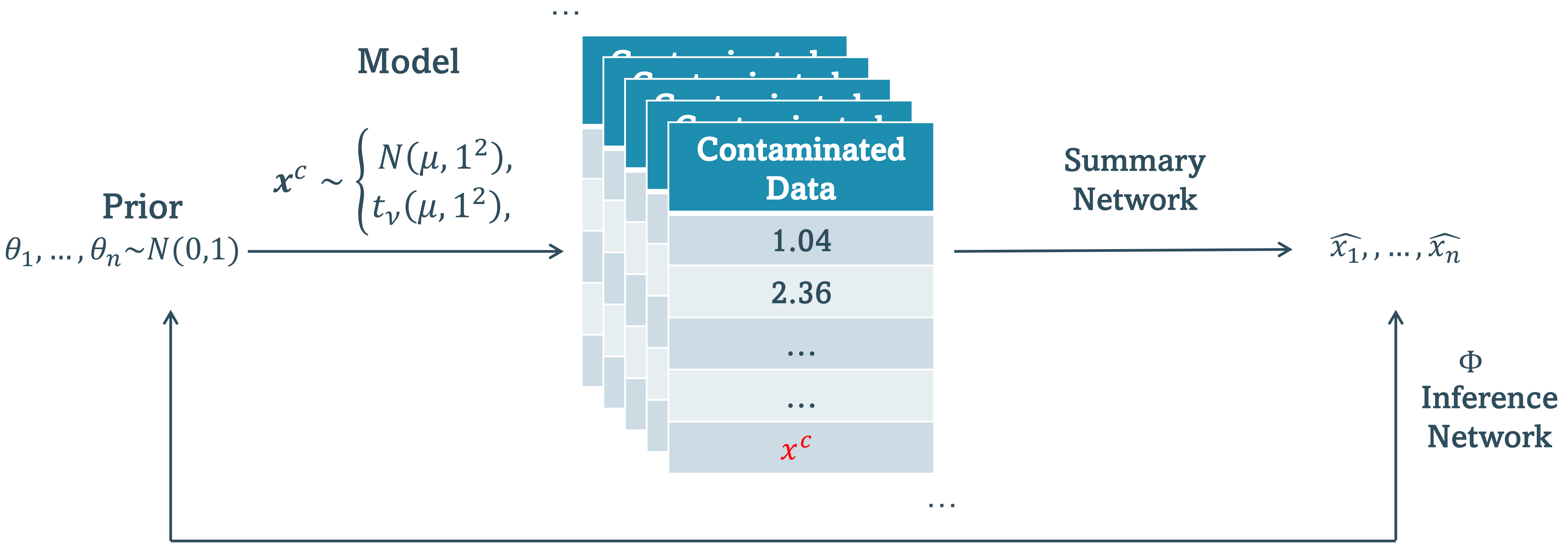
Assuming outliers in simulation

$x \sim N(\mu, 1^2)$

$$x^c \sim \begin{cases} N(\mu, 1^2), & with\ probablity\ 1 - \pi = 0.9 \\ t_\nu(\mu, 1^2), & with\ probablity\ \pi = 0.1 \end{cases}$$



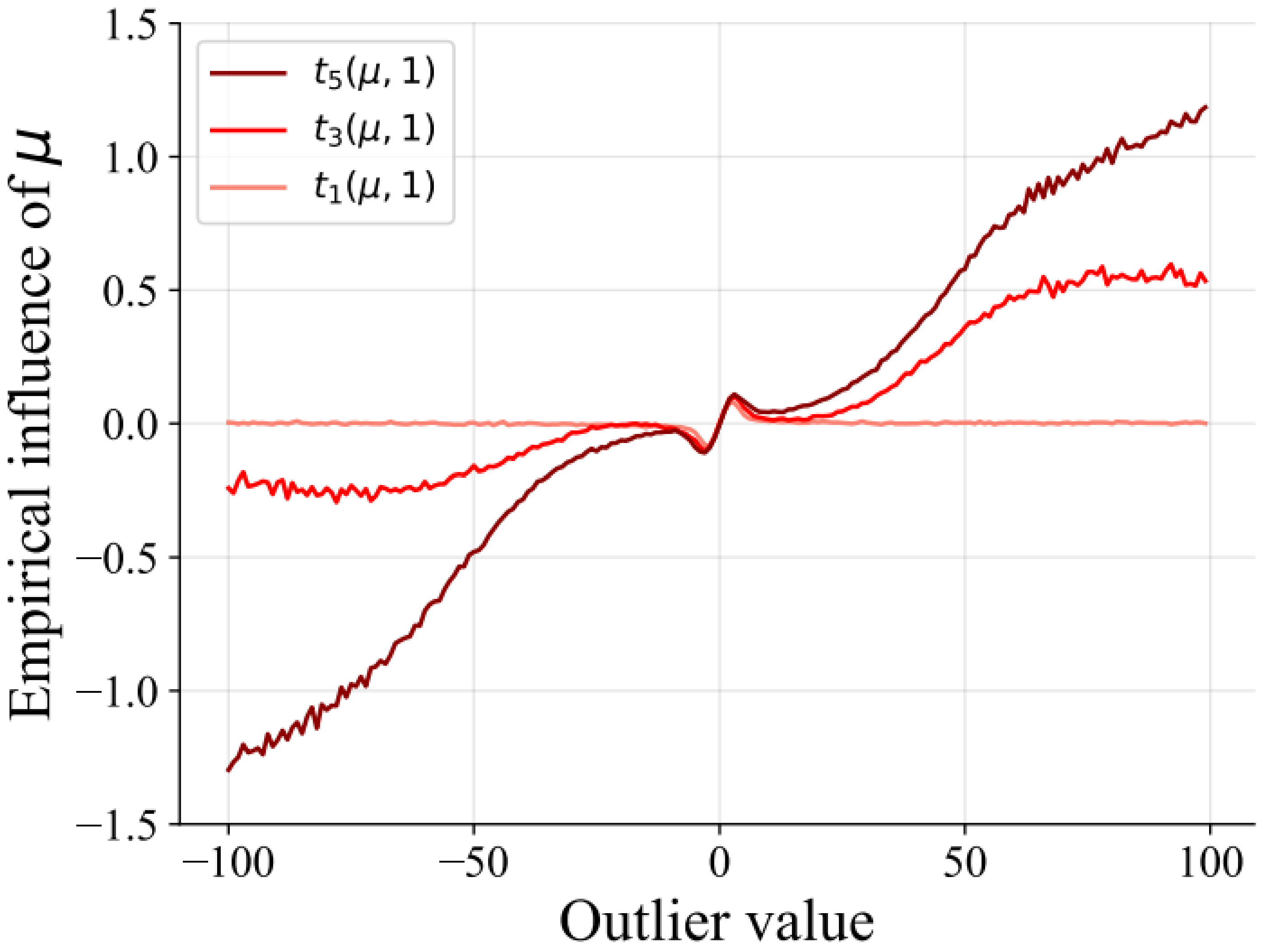
Robust estimator training



Empirical influence function

The smaller the ν,

the more robust the estimator.

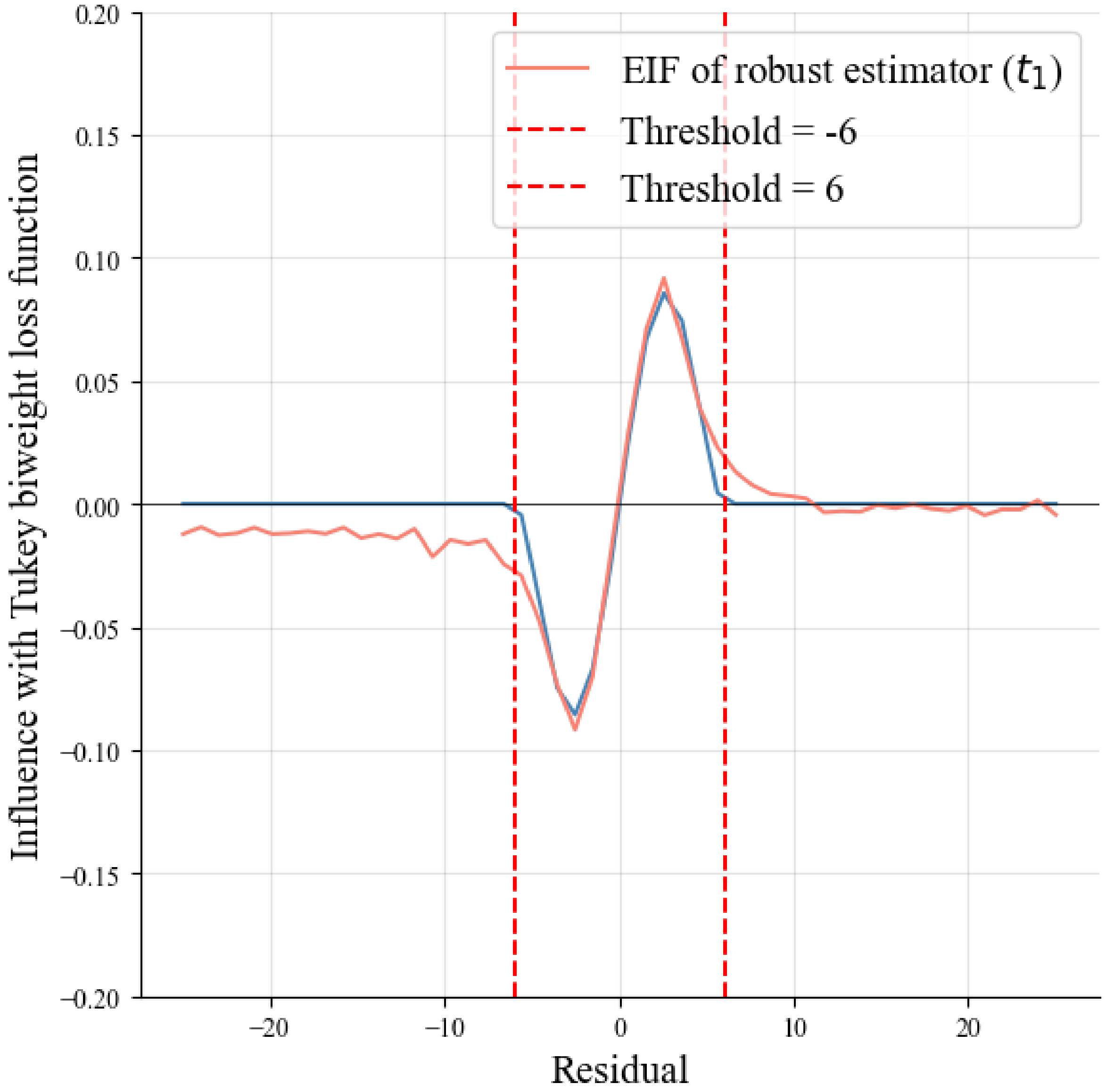


How many outliers during training?

Estimator		Regular outliers $Q_3 + 1.5 \cdot IQR$	%	Far outliers $Q_3 + 3 \cdot IQR$	%
Robust	normal + 10% t_1	2.828	2.585	4.949	1.269
	normal + 10% t_3	2.734	1.280	4.785	0.174
	normal + 10% t_5	2.719	1.008	4.758	0.051
Standard	normal	2.698	0.698	4.721	0.0002

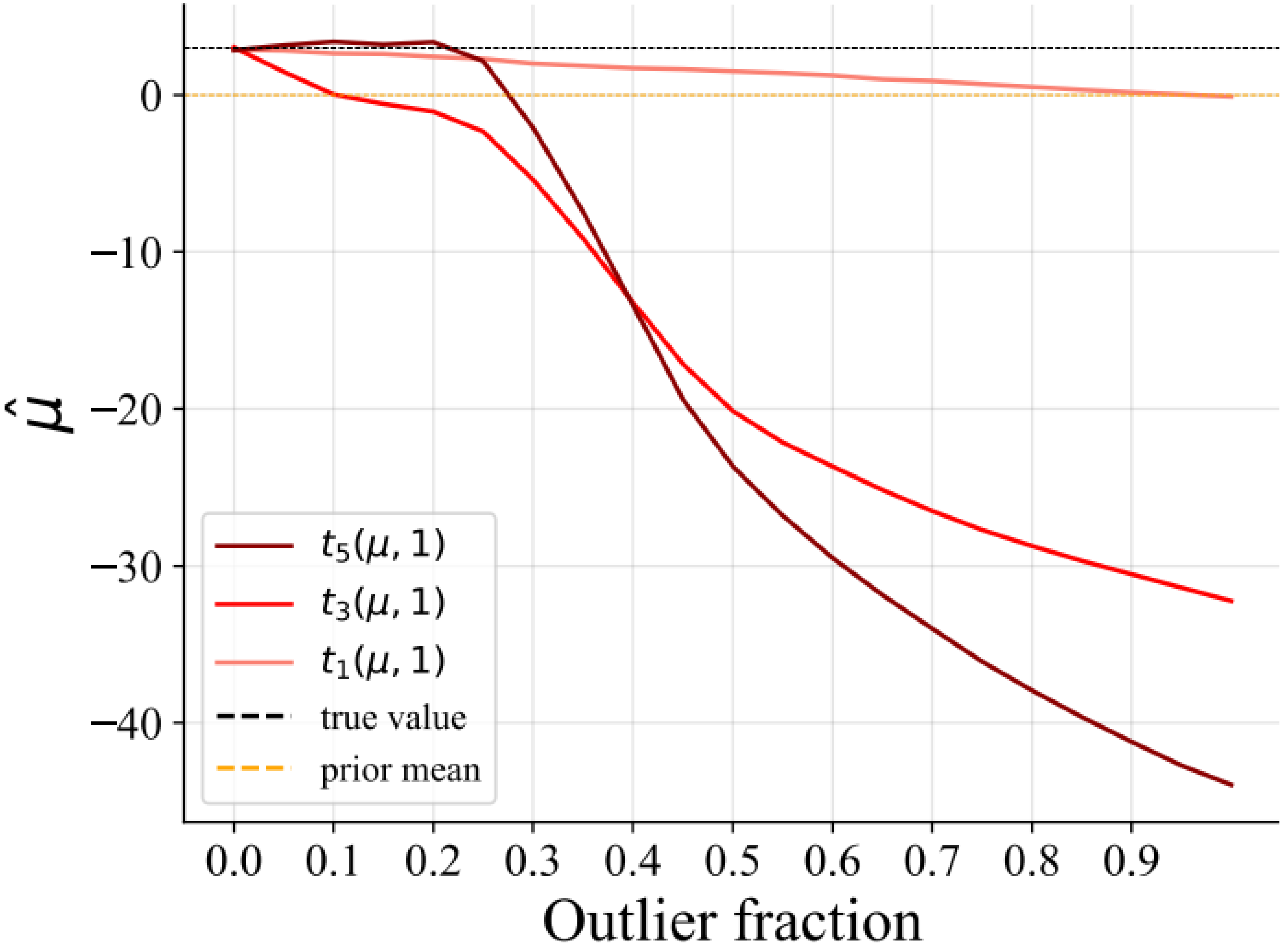
Coincidence

- Those (specific) neural networks works in a similar way as an M-estimator in traditional robust statistics.



Breakdown point ($x^c = -100$)

When there is no information, estimation goes to the 0, the prior mean.



Research Group of Quantitative Psychology and Individual Differences, PPW, KU Leuven

KU LEUVEN

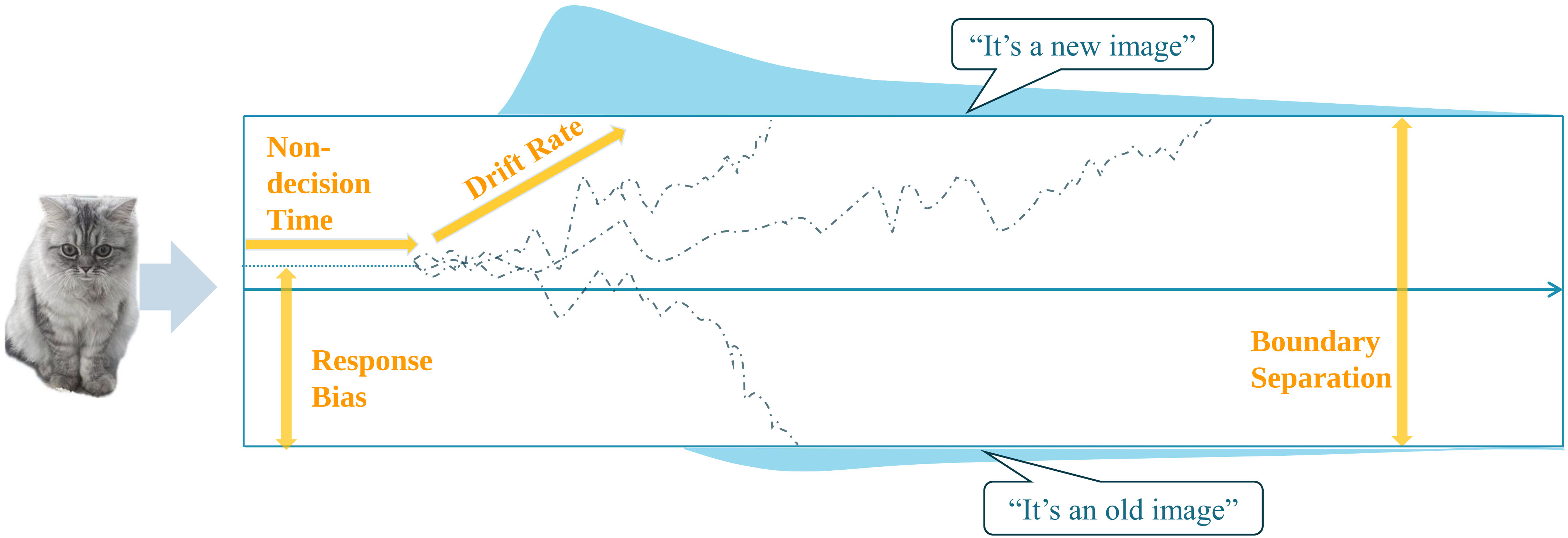
3

Robust Amortized Bayesian Inference of Cognitive Model Parameters

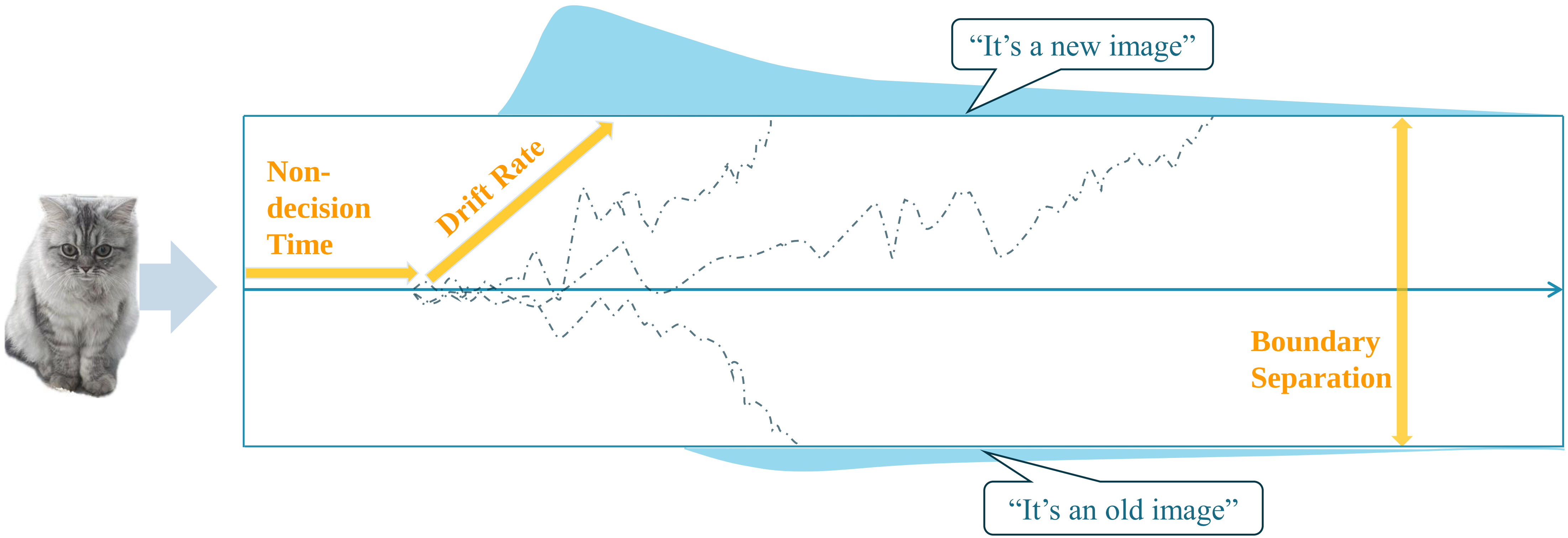
Research Group of Quantitative Psychology and Individual Differences, PPW, KU Leuven

Drift diffusion model (DDM)

Condition	Response	Reaction Time
0	0	0.63
1	1	0.89
0	1	0.23
…	…	…



Drift diffusion model (EZ version)



Sufficient summary statistics

$$P_c = \frac{1}{1+\exp(h)}$$

$$M_{RT} = \left(\frac{a}{2v}\right)\frac{1-\exp(h)}{1+\exp(h)} + T_{er}$$

$$V_{RT} = \left(\frac{as^2}{2v^3}\right)\frac{2h\exp(h)-\exp(2h)+1}{[\exp(h)+1]^2}$$

$$h = -va/s^2$$

Sufficient summary statistics

- P_c, M_{RT}, V_{RT}

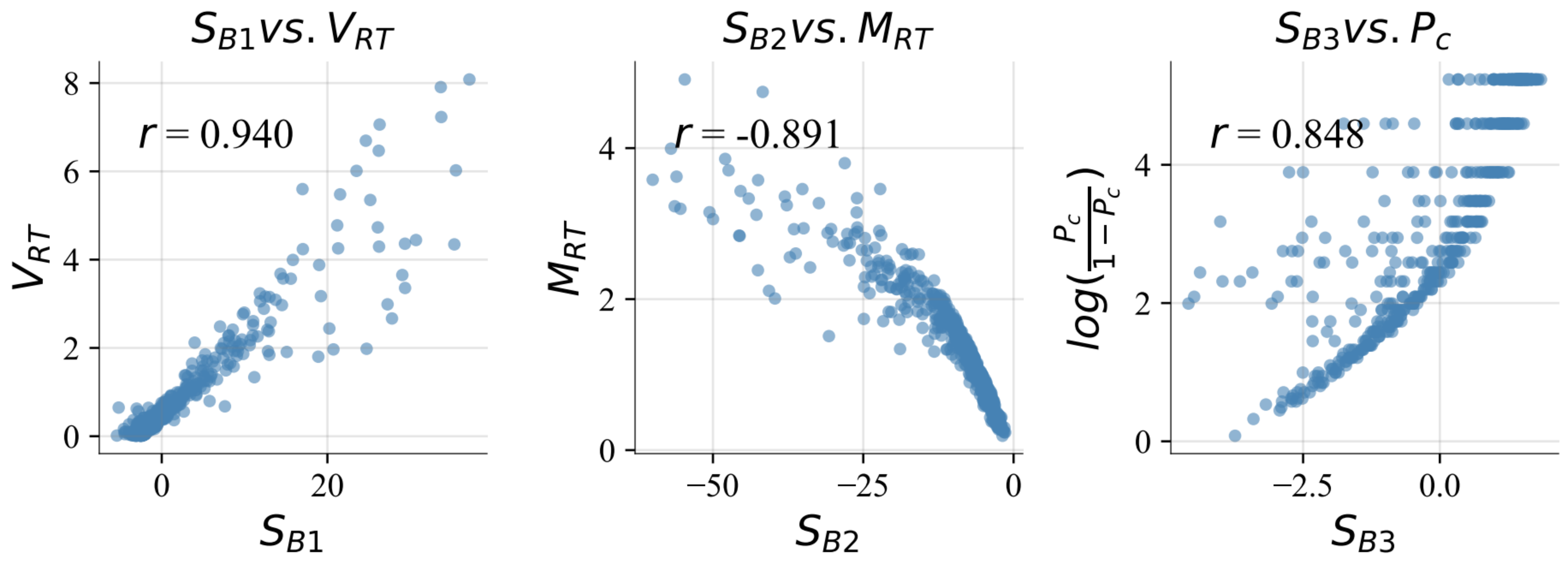
EZ version DDM estimator training

- Summary network: SetTransformer(summary_dim=3) (other settings stay the same as in bayesflow 1.1.4/ bayesflow stable-legacy branch)
- Inference network: Normalizingflow (Couplingflow)

Learned summary statistics

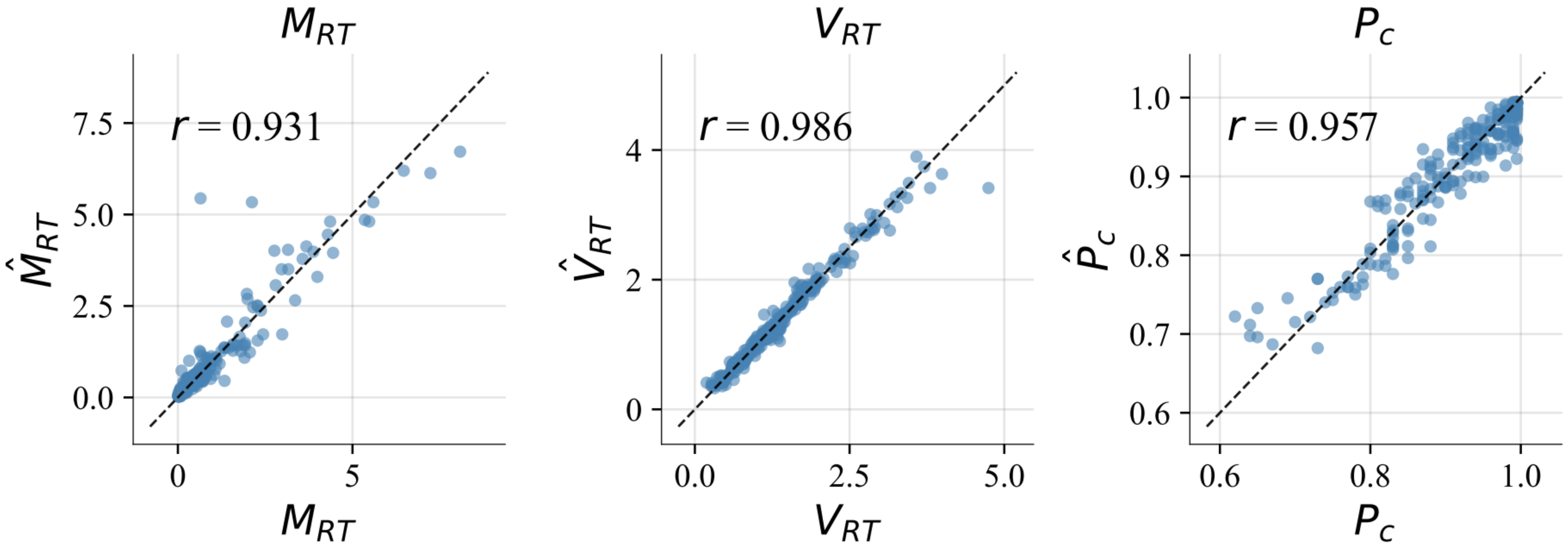
- S_{B1}, S_{B2}, S_{B3}

Sufficient vs. learned summary statistics



Research Group of Quantitative Psychology and Individual Differences, PPW, KU Leuven

Sufficient vs. learned summary statistics

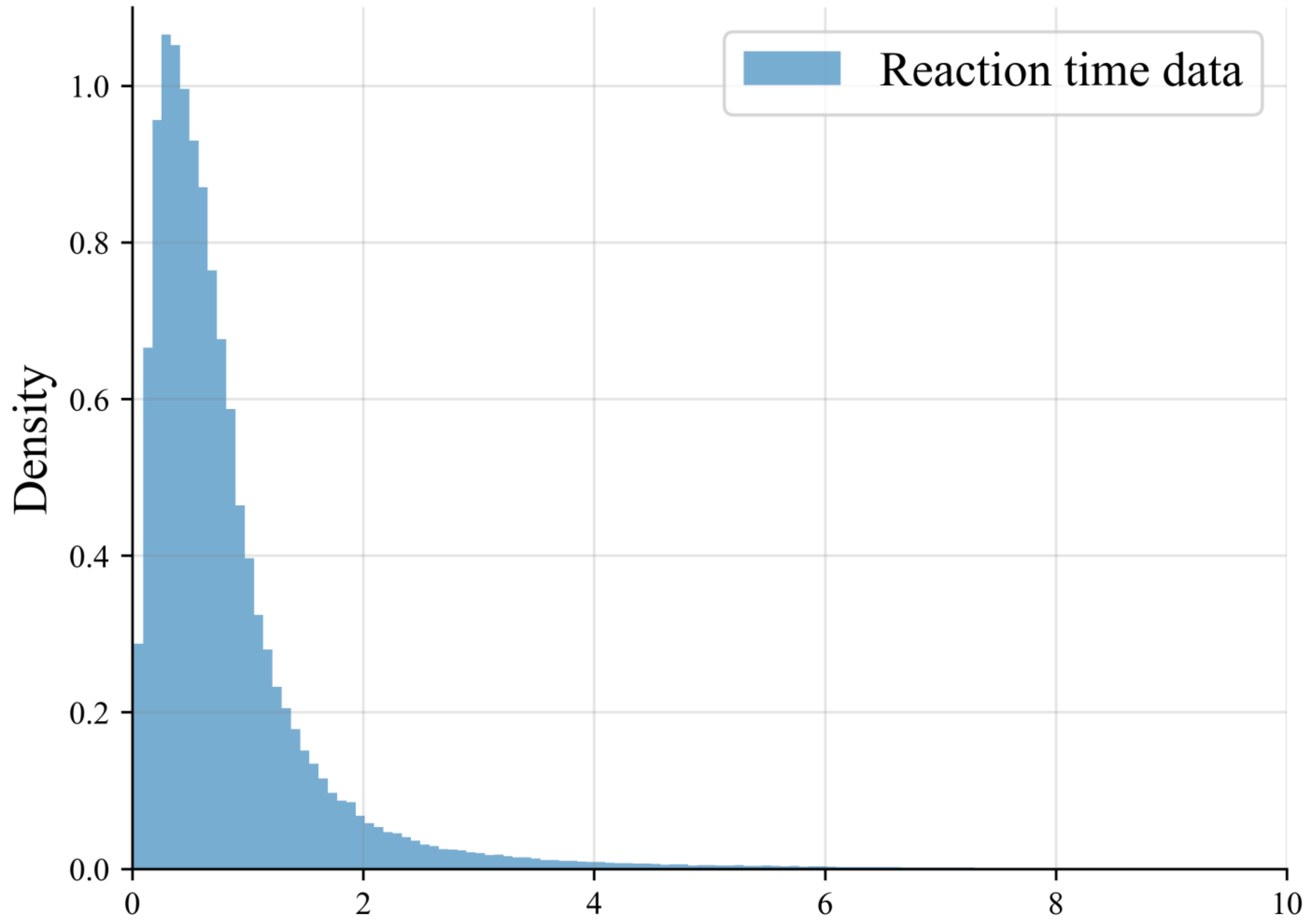


Sensitive to outliers by nature

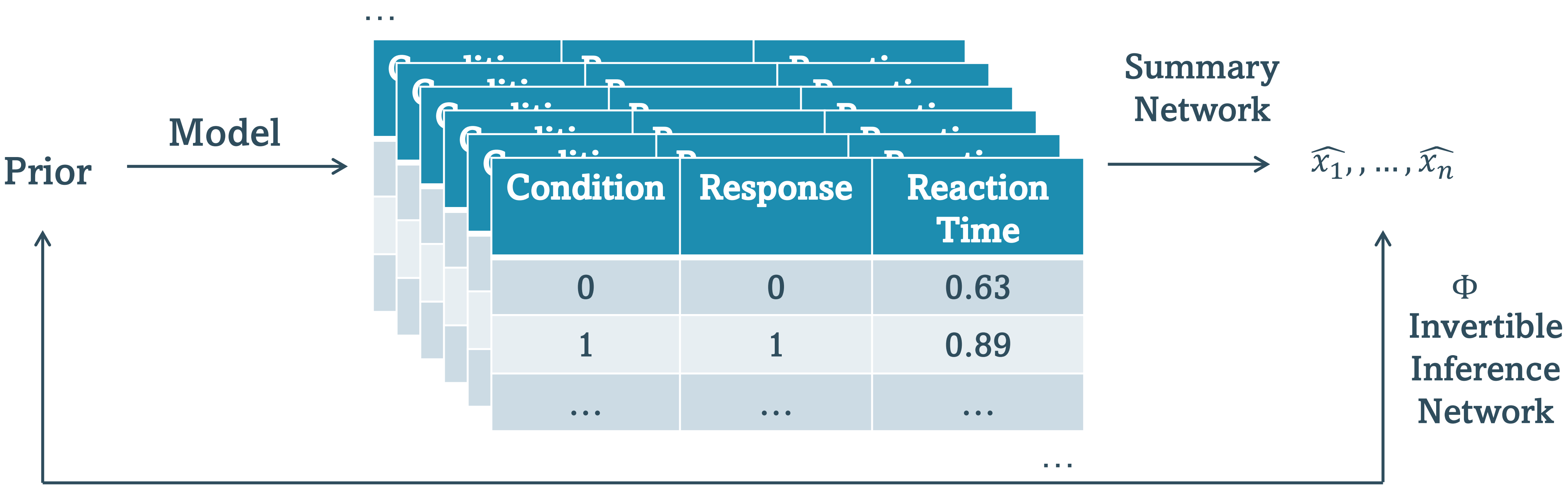
- non-decision time is shorter than the shortest reaction time

Simulation

$$(rt, r) \sim Wiener(v, a, z, T_{er})$$



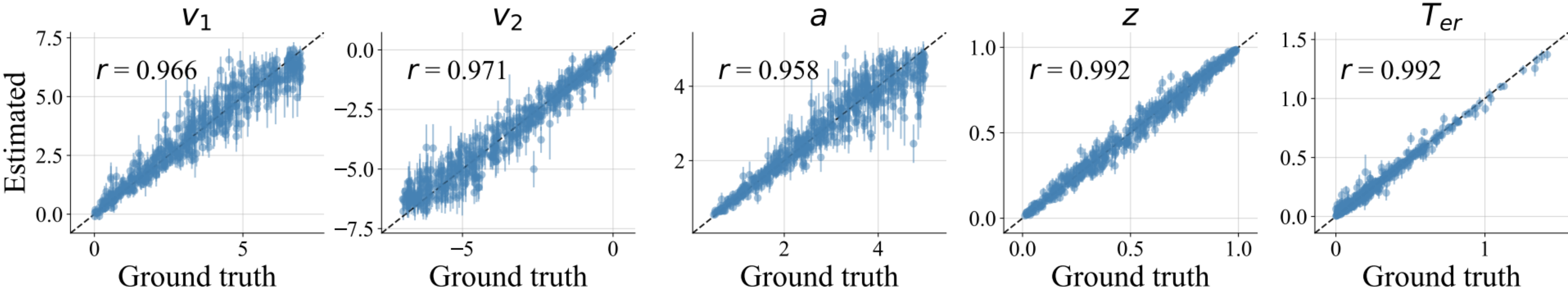
Standard estimator training



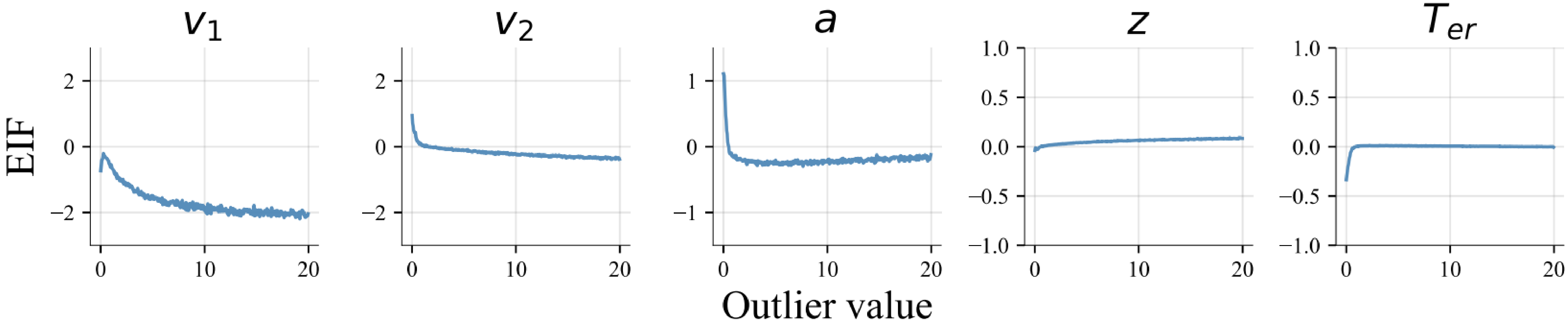
Condition	Response	Reaction Time
0	0	0.63
1	1	0.89
…	…	…

Research Group of Quantitative Psychology and Individual Differences, PPW, KU Leuven

Parameter recovery



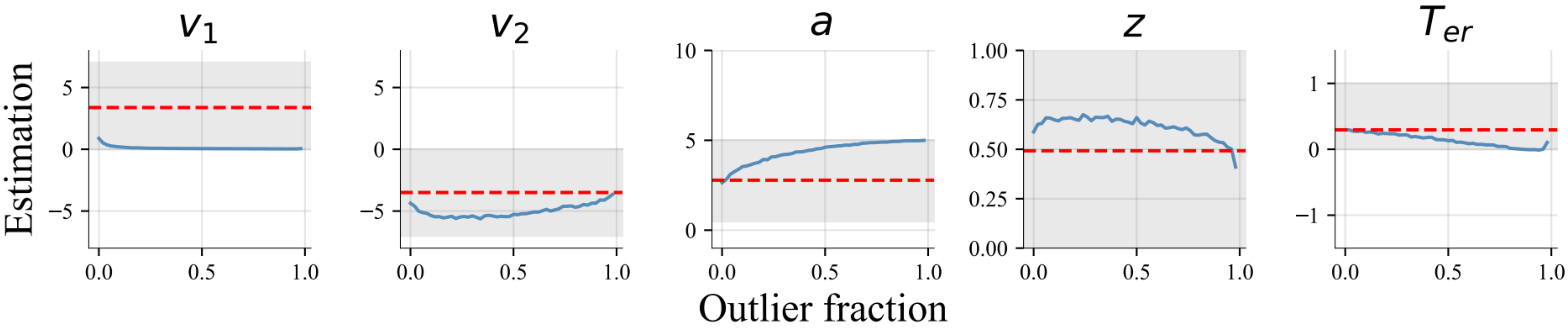
Empirical influence function



Empirical influence function

- Short outlier:
  - Underestimated non-decision time
  - Overestimated time for evidence accumulation process in both conditions
  - Reflected on the rising boundary separation
- Long outlier:
  - Overestimated time for evidence accumulation process in condition 1
  - Drift rate in condition 1 is underestimated

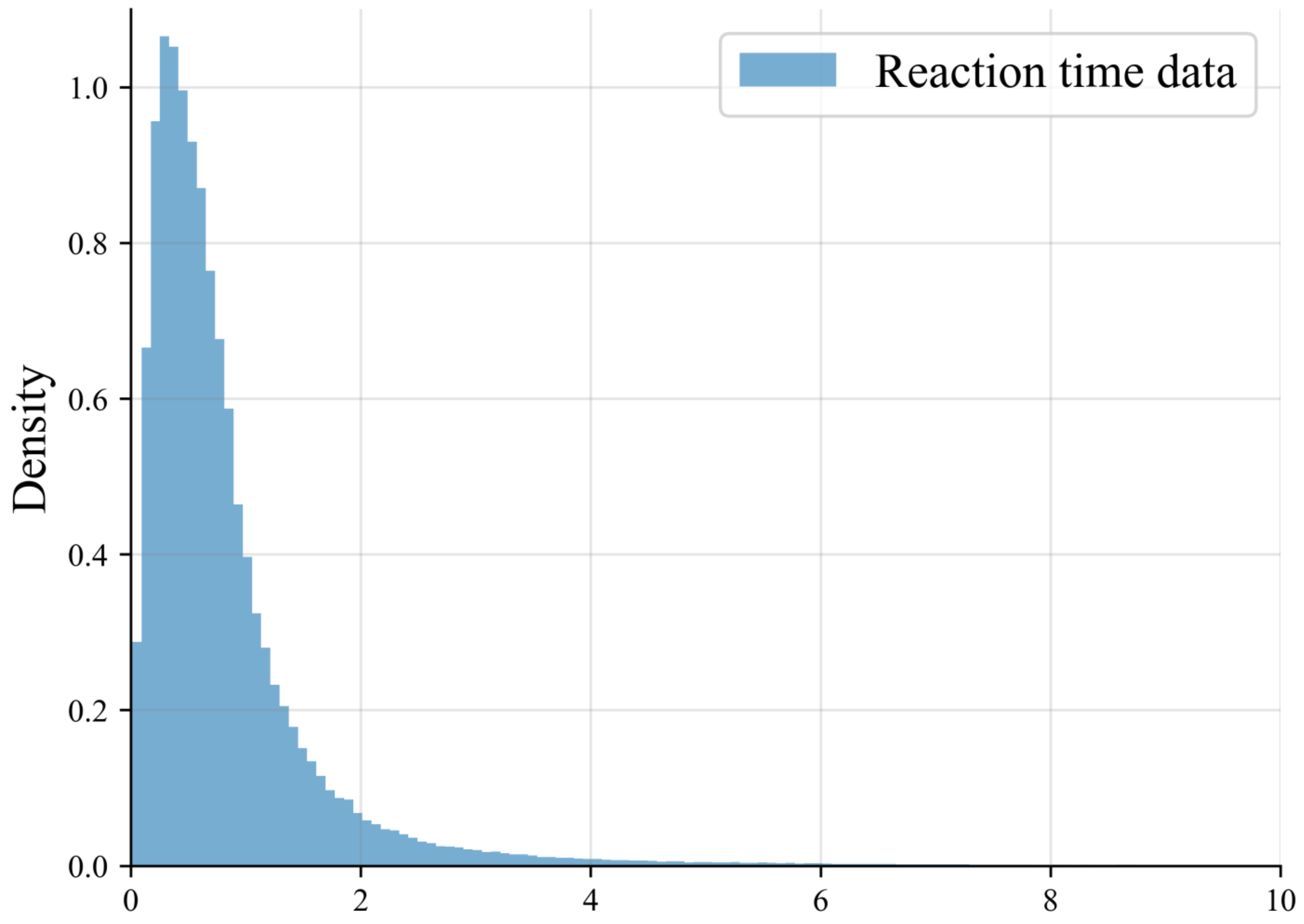
Breakdown point ($x^c = 20s$)



Simulation

$$(rt, r) \sim Wiener(v, a, z, T_{er})$$

$$(rt^c, r^c) \sim \begin{cases} Wiener(v, a, z, T_{er}), & 1 - \pi = 0.9 \\ contamination\ distribution & \pi = 0.1 \end{cases}$$



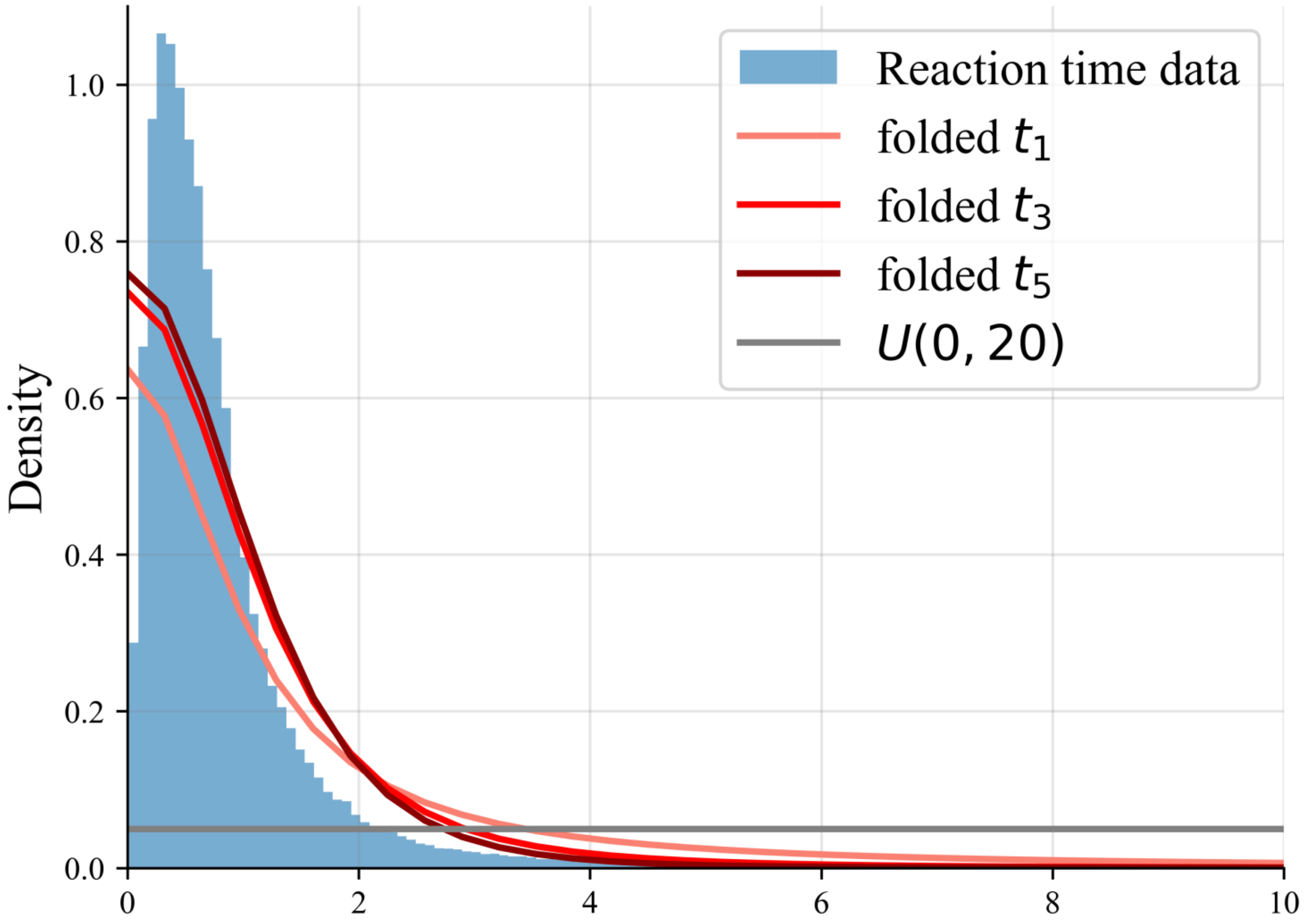
Research Group of Quantitative Psychology and Individual Differences, PPW, KU Leuven

KU LEUVEN

Assuming outliers in simulation

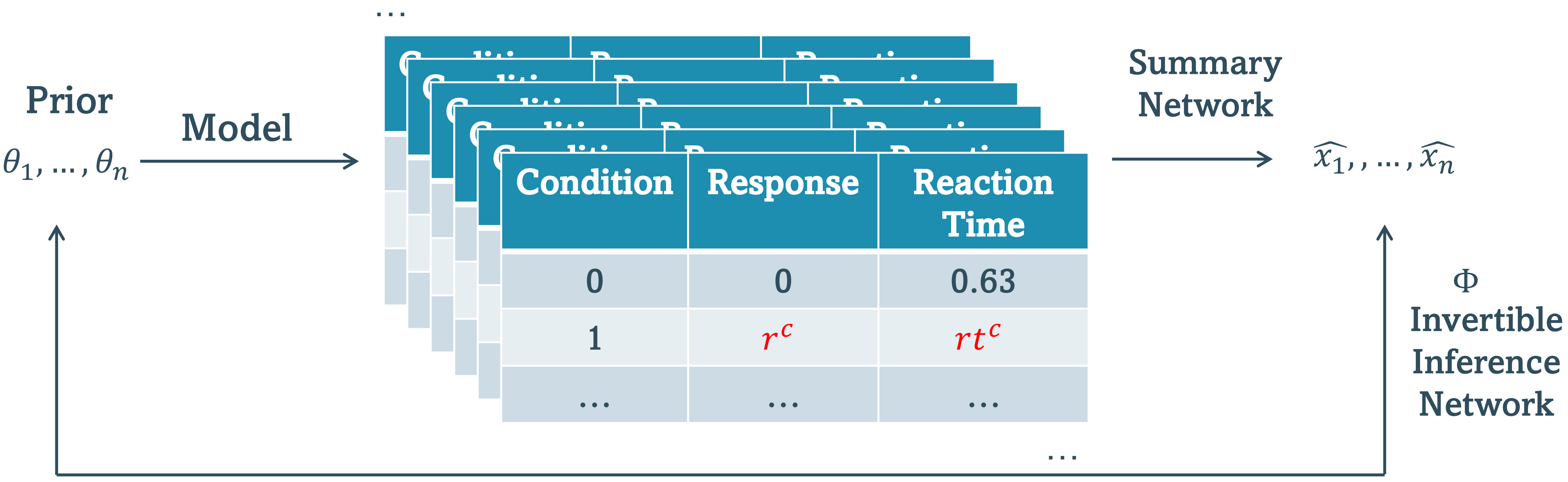
$$rt^c \sim \begin{cases} folded\ t_1 \\ folded\ t_3 \\ folded\ t_5 \\ U(0,20) \end{cases}$$

$$r^c \sim Bern(0.5)$$

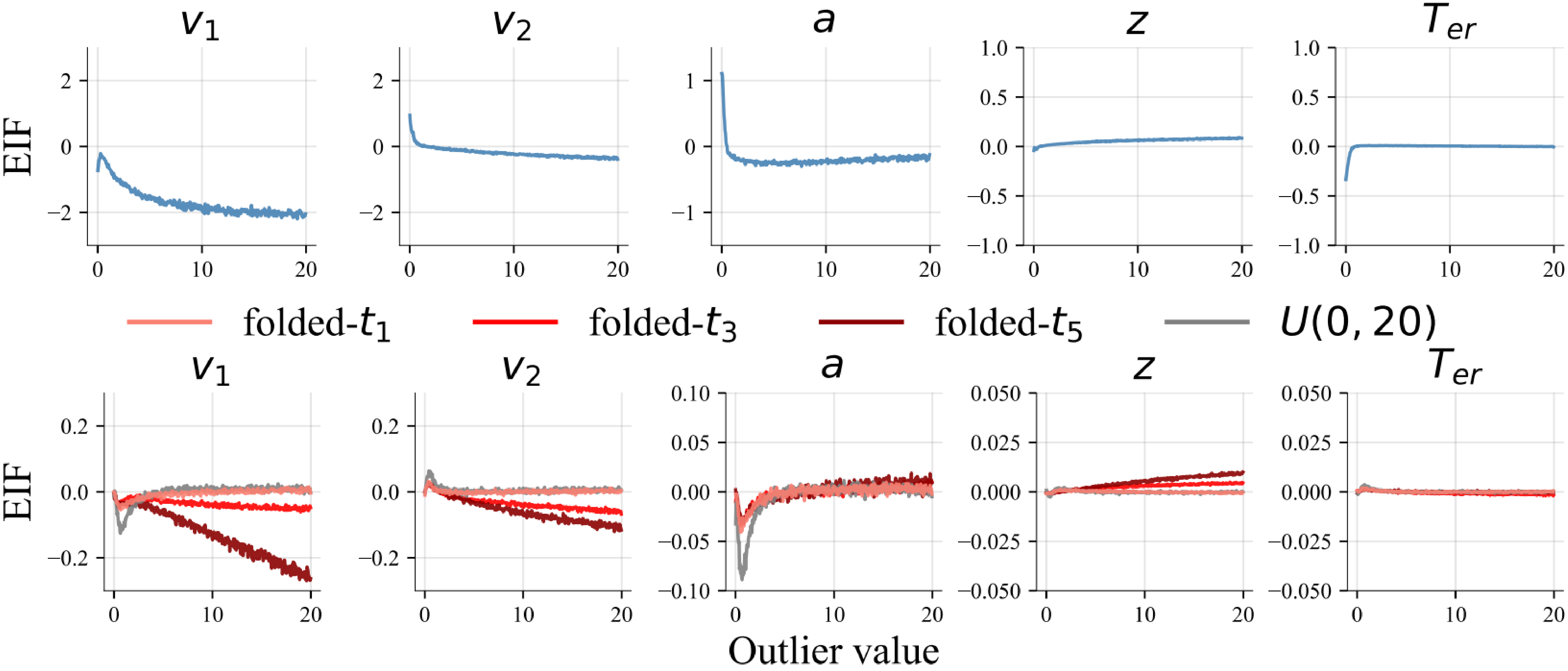


KU LEUVEN

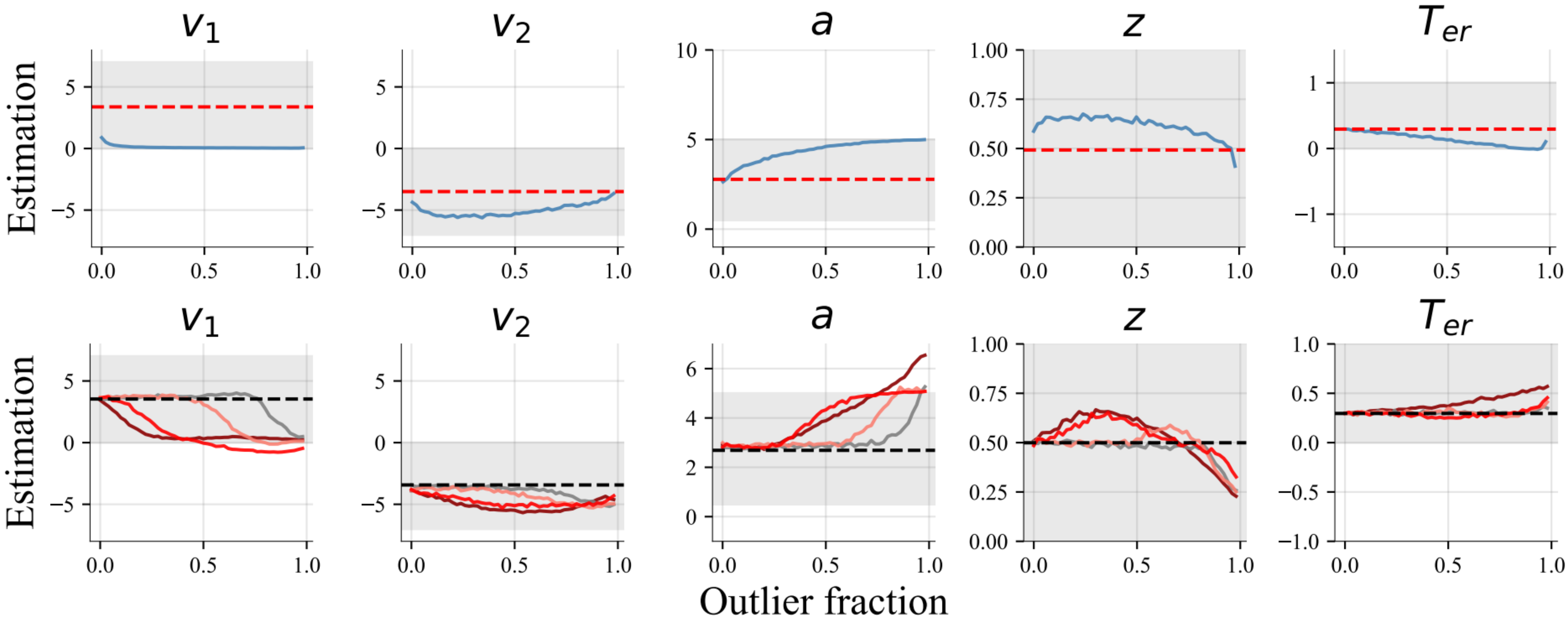
Robust estimator training



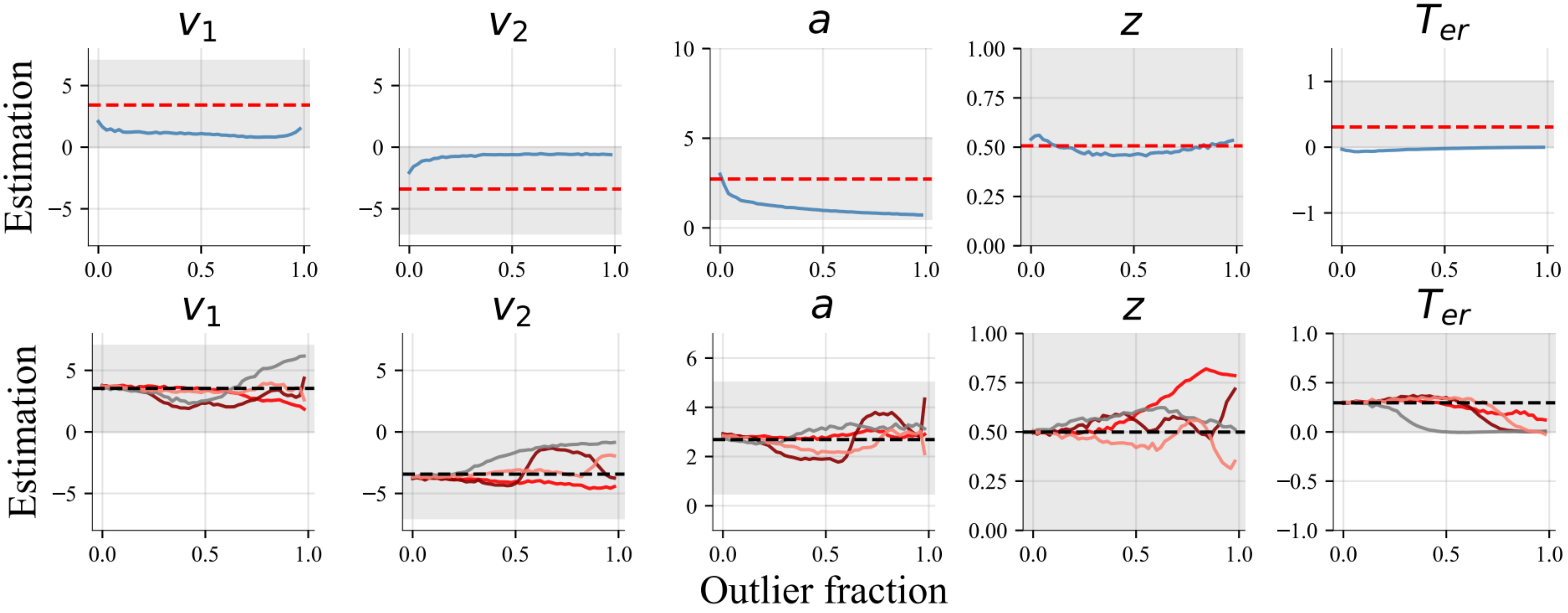
Empirical influence function



Breakdown point ($x^c = 20s$)



Breakdown point ($x^c = 0.01s$)



KU LEUVEN

The cost of robustness

$$\text{MAE Ratio}_j^{S,R} = \frac{1}{B}\sum_{b=1}^{B}\frac{|\hat{\theta}_j^S(\boldsymbol{x}_b) - \theta_{j(b)}|}{|\hat{\theta}_j^R(\boldsymbol{x}_b) - \theta_{j(b)}|},$$

Robust estimator	v_1	v_2	a	z	T_{er}
t_1	0.76	0.79	0.769	0.88	0.792
t_3	0.731	0.722	0.728	0.865	0.764
t_5	0.728	0.74	0.734	0.862	0.733
$U(0,20)$	0.876	0.894	0.907	0.936	0.848

$$\text{Posterior Variance Ratio}_j^{S,R} = \frac{1}{B}\sum_{b=1}^{B}\frac{\text{var}(\hat{\theta}_j^S \mid \boldsymbol{x}_b)}{\text{var}(\hat{\theta}_j^R \mid \boldsymbol{x}_b)},$$

Robust estimator	v_1	v_2	a	z	T_{er}
t_1	0.613	0.596	0.621	0.732	0.571
t_3	0.637	0.569	0.634	0.694	0.581
t_5	0.618	0.569	0.672	0.713	0.568
$U(0,20)$	0.723	0.672	0.723	0.756	0.636

Summary

- 1. Amortized Bayesian inference works:
  - It makes accurate inference through learning the mapping between data and $P(\theta|x)$
- 2. Amortized Bayesian inference can be robust:
  - Assuming the presence of outliers in simulation process increases the robustness of amortized Bayesian inference
- 3. The robustness depends on the contamination distribution:
  - Different contamination distributions have different effect on robustness
- 4. In our given example, the estimator with t_1 performs the best.

Thanks!