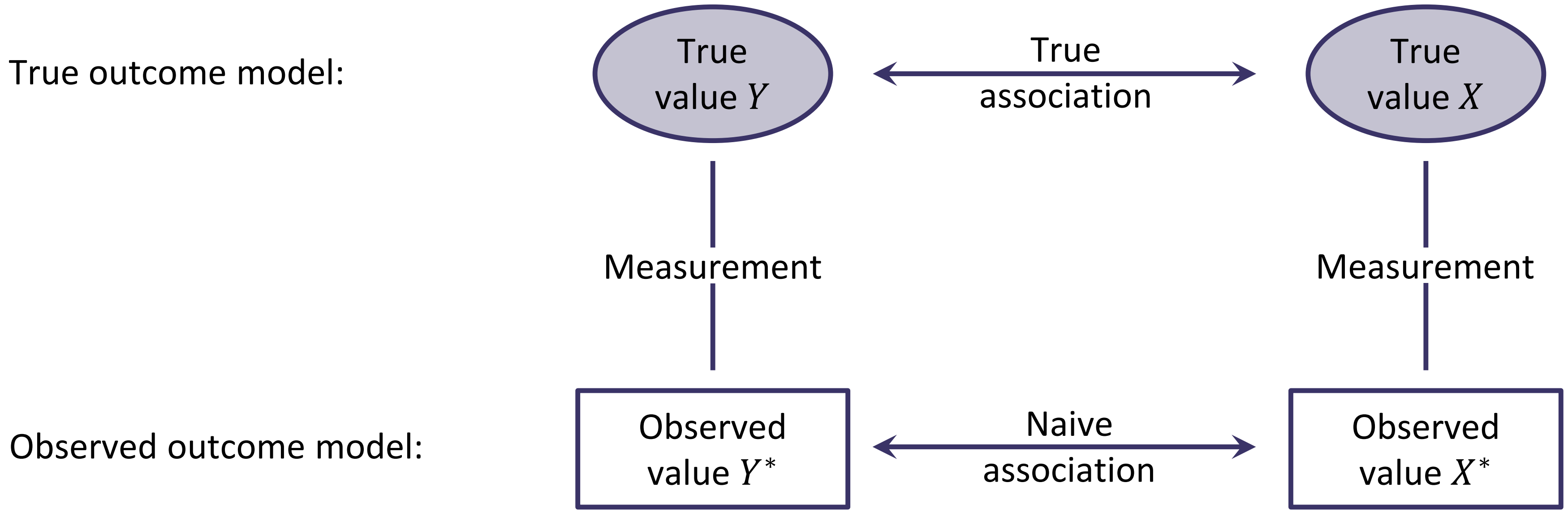
MEM-Explorer

Teaching tool for exploring measurement error and misclassification in statistical analyses

Veronika Deffner

TG 4 on measurement error and misclassification of the STRATOS Initiative

Analyses with erroneous observations



STRATOS TG 4: Measurement error and misclassification

Aims

1. Increase the **awareness** of the problems caused by measurement error and misclassification in statistical analyses
2. Remove barriers to use **statistical methods** that deal with such problems

Activities

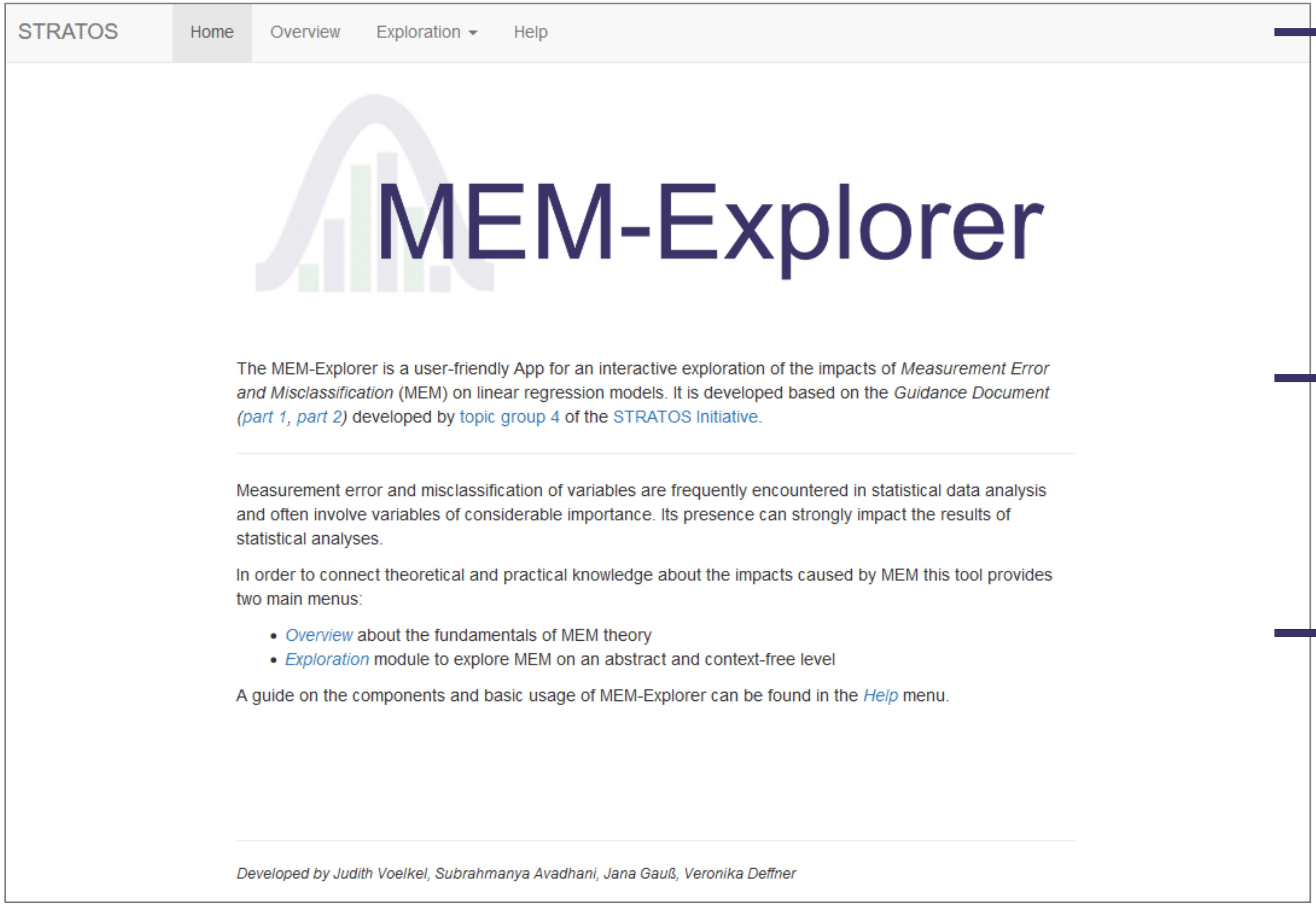
- Article on the **general relevance** of the topic (Wallace 2020)
- **Literature survey** about current practice (Shaw et al. 2018)
- **STRATOS guidance document:**

  Part 1 – Basic theory and simple methods of adjustment (Keogh et al. 2020)

  Part 2 – More complex methods of adjustment and advanced topics (Shaw et al. 2020)
- Presenting papers and workshops at **conferences**
- **Website:** http://www.stratostg4.statistik.uni-muenchen.de
- Interactive Shiny application **"MEM-Explorer":**

  https://mem-explorer.shinyapps.io/MEMExplorer-v5/

Home

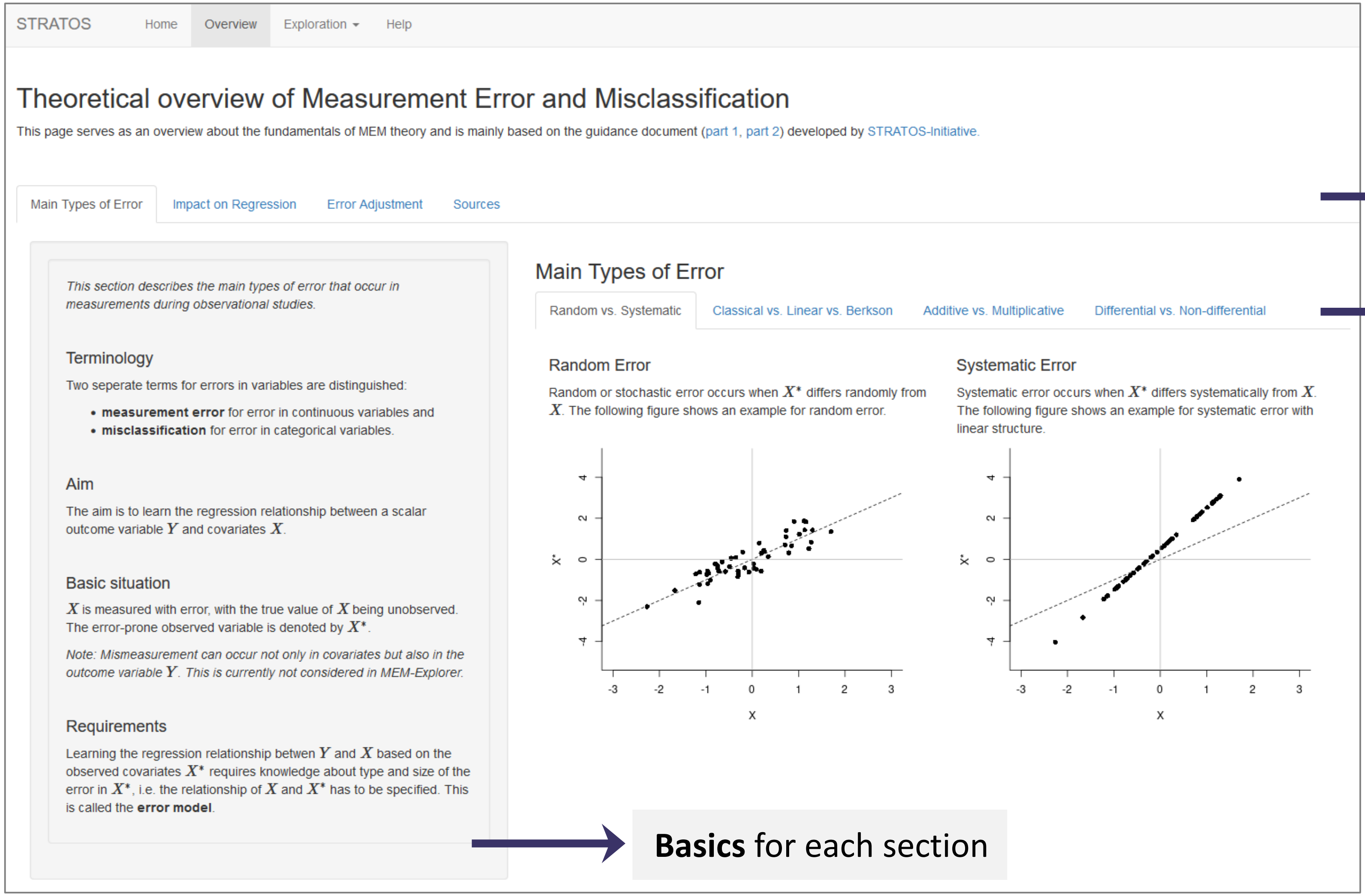


Navigation bar
with 4 tabs

MEM-Explorer is based on the **Guidance Document of TG 4**

Two main menus:
- **Overview** about theory
- **Exploration** module to explore MEM

Overview



Basics for each section

Sections:

- Main Types of error
- Impact on Regression
- Error Adjustment

Subsections for each section

Overview

Main Types of Error

- Random vs. systematic
- Classical vs. linear vs. Berkson
  - Measurement error
  - Misclassification
- Additive vs. Multiplicative
- Differential vs. Non-differential

Impact on Regression

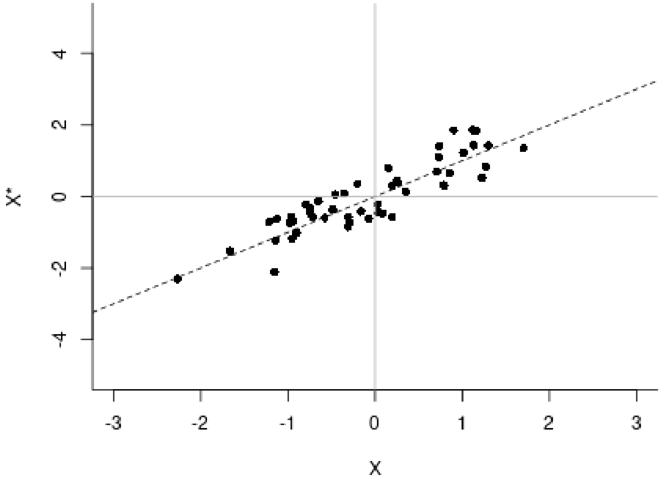
Error Adjustment

Main Types of Error

Random vs. Systematic | Classical vs. Linear vs. Berkson | Additive vs. Multiplicative | Differential vs. Non-differential

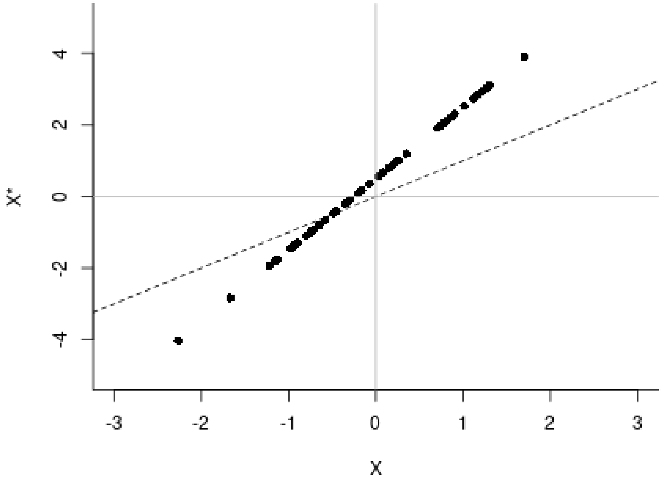
Random Error

Random or stochastic error occurs when X^* differs randomly from X. The following figure shows an example for random error.



Systematic Error

Systematic error occurs when X^* differs systematically from X. The following figure shows an example for systematic error with linear structure.



Overview

Main Types of Error

Impact on Regression

(a) Measurement error

- Classical vs. Linear vs. Berkson
- Impact on
  - Regression coefficient
  - Test of null hypothesis
  - Power

(b) Misclassification

- Classical vs. Berkson

Error Adjustment

Impact on Regression

(a) Measurement error

(b) Misclassification

Impacts of Measurement Errors on Linear Regression

Assume the following:

- X, X^* are continuous covariates.
- U is a random variable with mean 0.
- $U \perp X$ in linear and classiscal error models.
- $U \perp X^*$ in Berkson error model.
- Errors U are nondifferential with respect to outcome Y.
- ρ_{XX^*} is the correlation coefficient between X and X^*.

	Classical measurement error model $X^* = X + U$	Linear measurement error model $X^* = \alpha_0 + \alpha_X X + U$	Berkson error model $X = X^* + U$
Single covariate regression			
Regression coefficient	Underestimated $\lambda = \frac{\mathrm{var}(X)}{\mathrm{var}(X) + \mathrm{var}(U)}$	Biased in either direction $\lambda = \frac{\alpha_X \mathrm{var}(X)}{\alpha_X^2 \mathrm{var}(X) + \mathrm{var}(U)}$	Unbiased $\lambda = 1$
Test of null hypothesis	Valid	Valid	Valid
Power	Reduced → *effective sample size reduced by approximately* $\rho_{XX^*}^2 = \lambda$	Reduced → *effective sample size reduced by approximately* $\rho_{XX^*}^2$	Reduced → *effective sample size reduced by approximately* $\rho_{XX^*}^2$
Regression with multiple error prone covariates			

Overview

Main Types of Error

Impact on Regression

Error Adjustment

- SIMEX
  - (a) Measurement error
  - (b) Misclassification

Error Adjustment

SIMEX

The basic idea of SIMEX (Simulation extrapolation) is to add more error to the error prone covariate X^*, to see the impact this has on the parameter estimates, and then extrapolate back to the situation with no measurement error. This is equivalent to estimating the relationship between the measurement error variance $var(U)$ and the parameter estimates and to extrapolate back to the situation in which the error variance is 0.

(a) Measurement error

(b) Misclassification

Assume a *classical measurement error* model, that is $X^* = X + U$ (U is a random variable with mean 0 and $U \perp X$).

The theoretical relationship between the biased regression slope β^*_X, based on the regression of Y on X^*, can be described as a function of $var(U)$, $\beta^*_X(var(U))$. This function is estimated from simulated datasets obtained by adding further measurement error to X^*. To estimate β_X, the function $\beta^*_X(var(U))$ is extrapolated back to $var(U) = 0$ (as $\beta^*_X(0) = \beta_X$).

Let $var(U)$ be known or assumed, e.g. through comparison measurements.

- **Simulation step:**
  For each value $s_1, \dots, s_m \geq 0$, B new pseudo-datasets are simulated by

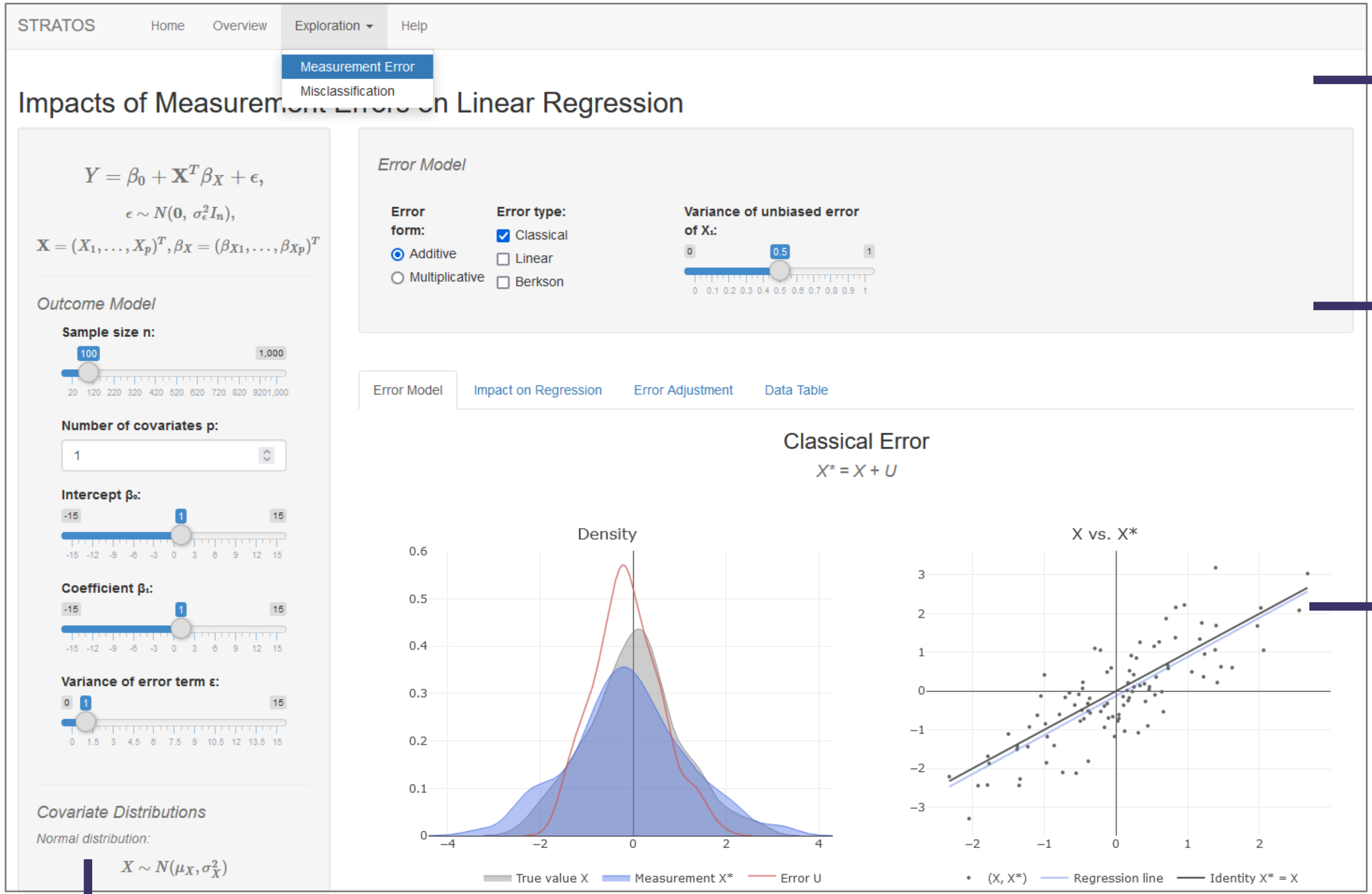
$$X^*_{ib}(s_k) = X^*_i + \sqrt{s_k var(U)} U_{ibk},\ i = 1, \dots, n; b = 1, \dots, B; k = 1, \dots, m$$

  where U_{ibk} are *iid* standard normal variables. The measurement error variance of $X^*_{ib}(s_k)$ is therefore $(1 + s_k) var(U)$.
  For each pseudo-dataset, the unadjusted estimator based on Y and $X^*_b(s_k)$ is calculated and the results are averaged over the B repetitions, which leads to $\widehat{\beta}^*_{X s_k}$.
- **Extrapolation step:**
  We use a parametric approximation for the function $\beta^*_X((1+s)var(U))$, denoted by $G(s, \Gamma)$, where Γ denotes the parameters of the function, e.g. in case of a quadratic approximation $G(s, \Gamma) = \gamma_0 + \gamma_1 s + \gamma_2 s^2$.
  The parameters Γ are estimated by least squares and the estimated parametric function is then extrapolated to $s = -1$, the case of no measurement error, yielding the SIMEX estimator:

$$\widehat{\beta}_{SIMEX} = G(-1, \widehat{\Gamma})$$

When β_X is a vector, the SIMEX-procedure can be applied separately for each component.

Exploration



Exploration:

- **Measurement Error**
- **Misclassification**

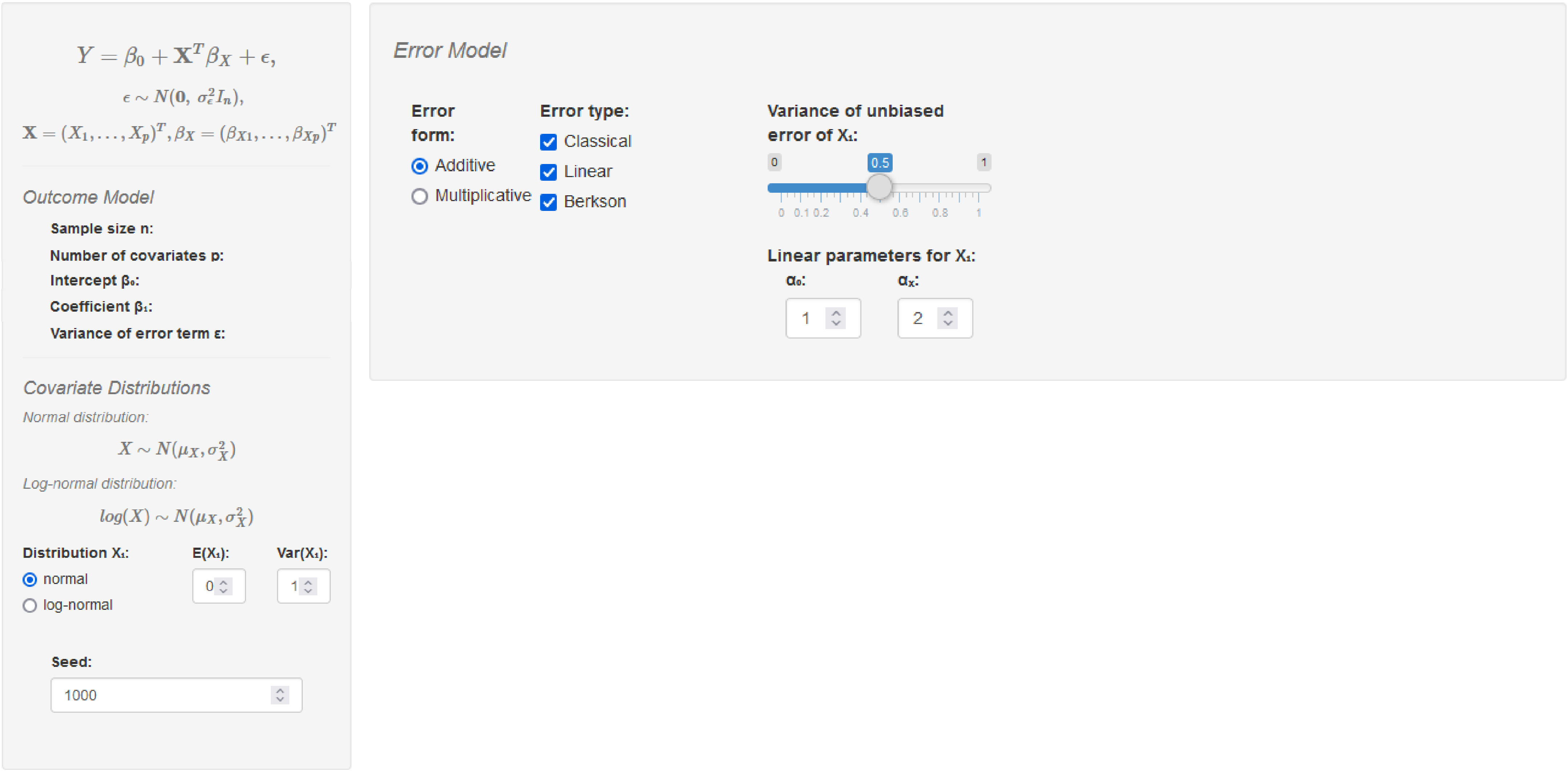
Specification of **error model**

Subsections with output:

- Error Model
- Impact on Regression
- Error Adjustment
- Data Table

Specification of **sample size, outcome model** and **covariate distributions**

Exploration – Input



Exploration – Output

Error Model

Empirical distributions:

- True covariate
- Error-prone covariate
- Measurement error

Impact on Regression

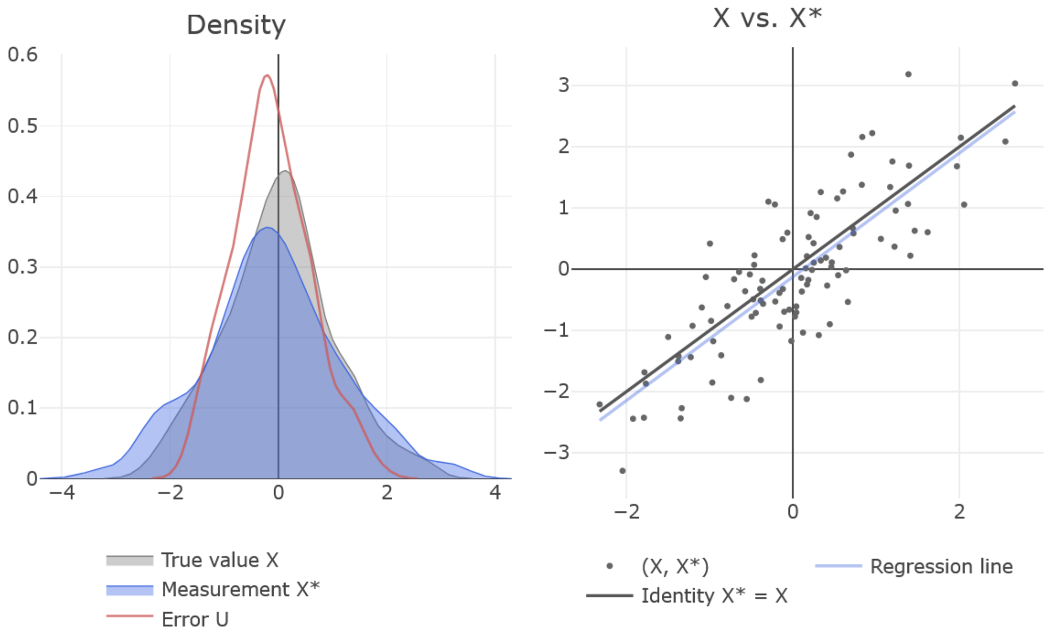
Error Adjustment

Data Table

Error Model | Impact on Regression | Error Adjustment | Data Table

Classical Error

$X^* = X + U$



	Mean	Empirical variance
U1	-0.12	0.51
X1	0.02	1.01
X1*	-0.11	1.54

Exploration – Output

Error Model

Impact on Regression

Visualization of the impact of MEM on the regression model

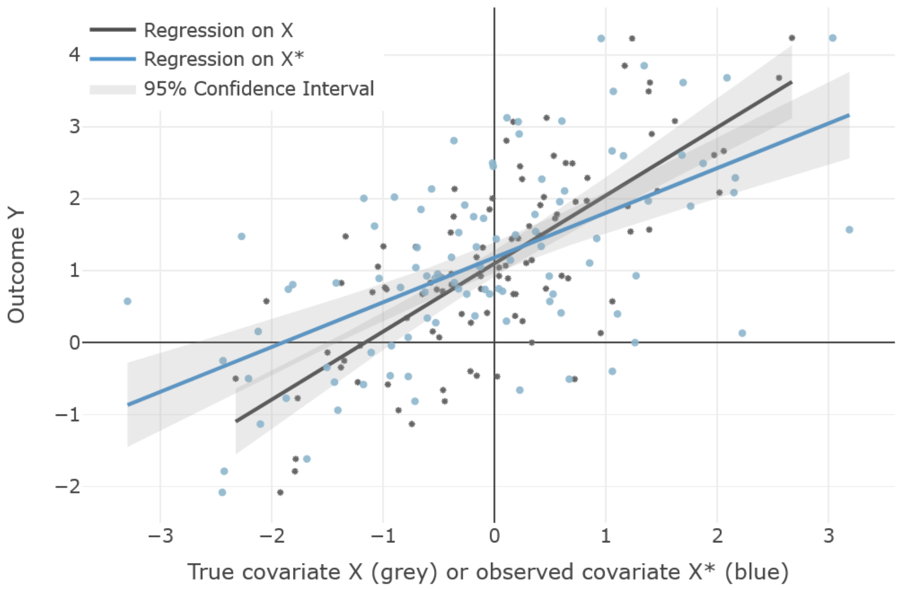
Error Adjustment

Data Table

Error Model | Impact on Regression | Error Adjustment | Data Table

Classical Error

$$X^* = X + U$$



	$Coeff_{Classical}$	$Coeff^*_{Classical}$	$\lambda_{Classical}$
Intercept	1.1	1.18	
X1	0.95	0.62	0.65

	$SE_{Classical}$	$SE^*_{Classical}$
Intercept	0.09	0.11
X1	0.09	0.09

	$p_{Classical}$	$p^*_{Classical}$
Intercept	< 0.001	< 0.001
X1	< 0.001	< 0.001

Exploration – Output

Error Model

Impact on Regression

Error Adjustment

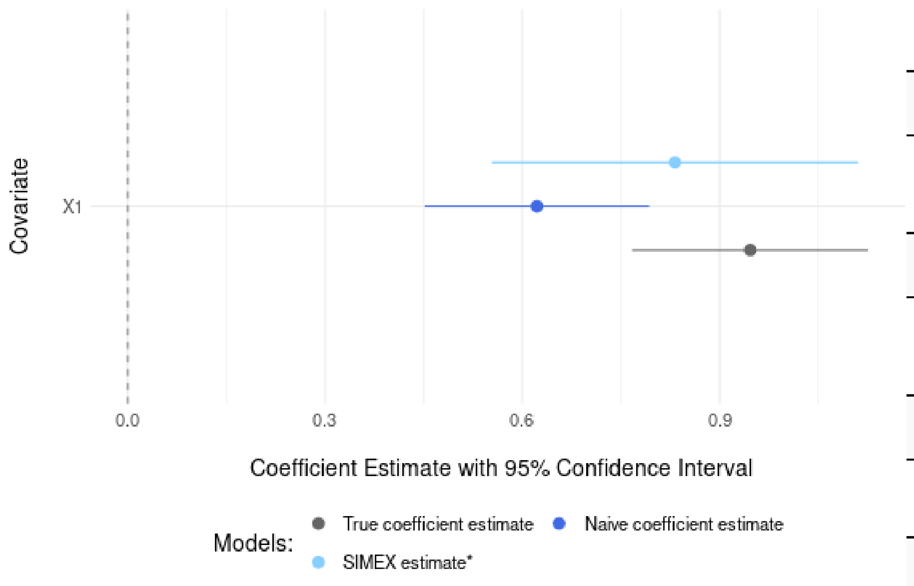
Visualization of regression models using methods that adjust for bias in regression coefficients caused by MEM

Data Table

Error Model | Impact on Regression | Error Adjustment | Data Table

Classical Error

$X^* = X + U$



$Coeff_{Classical}$	$Coeff^*_{Classical}$	$Coeff\_SIMEX_{Classical}$
0.95	0.62	0.83

$\lambda_{Classical}$	$\lambda\_SIMEX_{Classical}$
0.65	0.87

$SE_{Classical}$	$SE^*_{Classical}$	$SE\_SIMEX_{Classical}$
0.09	0.09	0.14

$p_{Classical}$	$p^*_{Classical}$	$p\_SIMEX_{Classical}$
< 0.001	< 0.001	< 0.001

Exploration – Output

Error Model

Impact on Regression

Error Adjustment

Data Table

- Excerpt of simulated dataset
- Download

Error Model | Impact on Regression | Error Adjustment | Data Table

Download Data Table

Header of simulated Data

y	x1.x	x1.u	x1.classical
-0.82	-0.45	-0.27	-0.71
-0.04	-1.21	0.28	-0.92
0.93	0.04	-0.65	-0.61
2.50	0.64	-0.66	-0.02
0.34	-0.79	0.18	-0.60
0.81	-0.39	-1.42	-1.81

Example

Sample size

$$n = 100$$

Outcome model

Simple linear regression model with

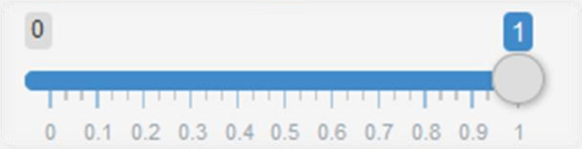
- coefficients $\beta_0 = 0$ and $\beta_1 = 1$
- error variance $\mathrm{Var}(\varepsilon) = 1$

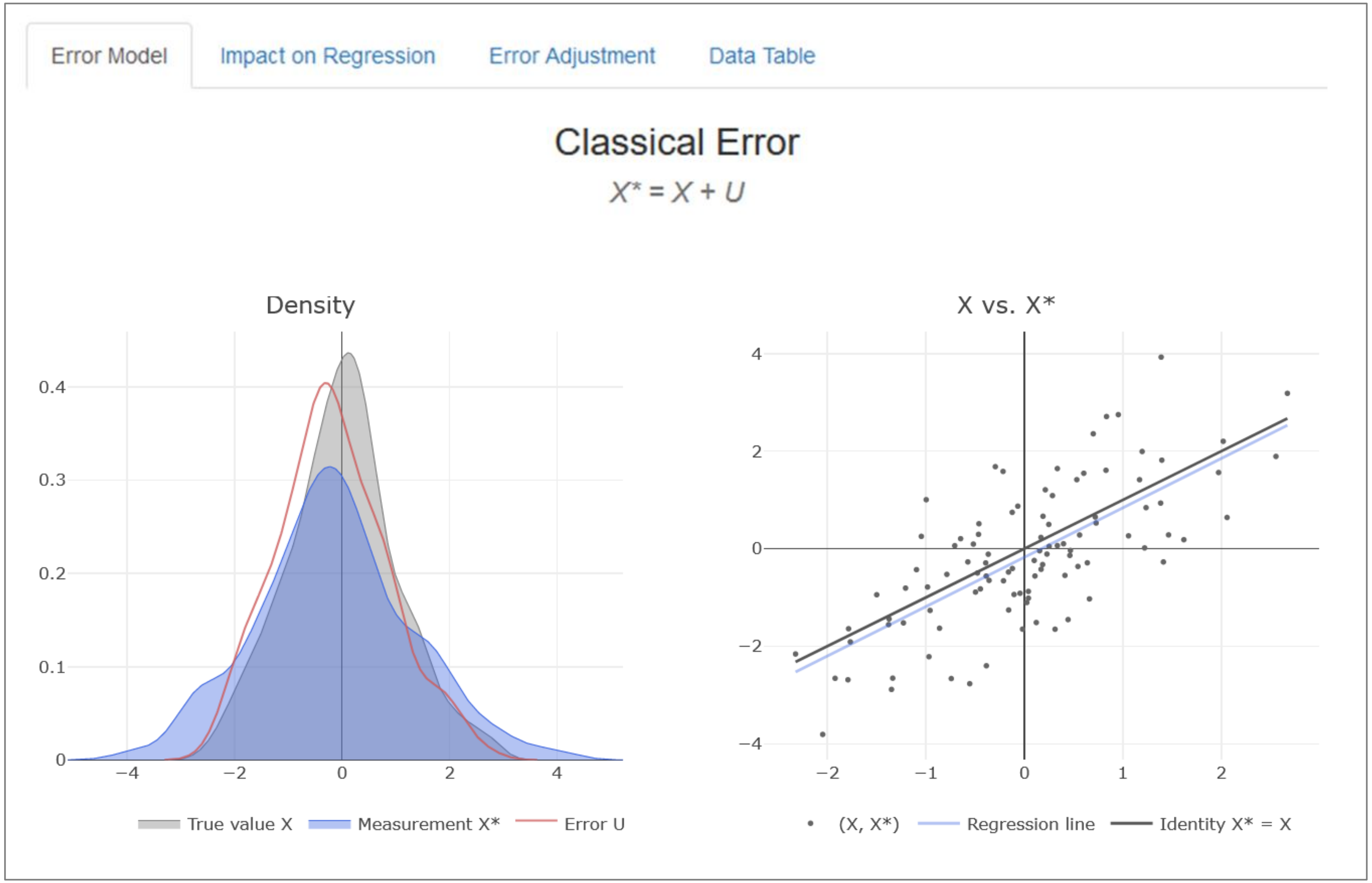
Covariate distribution

$$X \sim N(0,1)$$

Error model

- Additive, classical error
- Error variance





Example

Sample size

$$n = 100$$

Outcome model

Simple linear regression model with

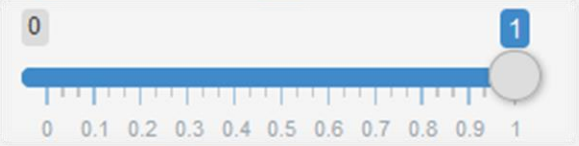
- coefficients $\beta_0 = 0$ and $\beta_1 = 1$
- error variance $\mathrm{Var}(\varepsilon) = 1$

Covariate distribution

$$X \sim N(0,1)$$

Error model

- Additive, classical error
- Error variance



Example

Sample size

$$n = 100$$

Outcome model

Simple linear regression model with

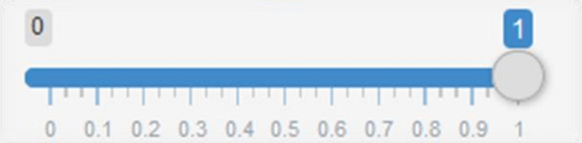
- coefficients $\beta_0 = 0$ and $\beta_1 = 1$
- error variance $\mathrm{Var}(\varepsilon) = 1$

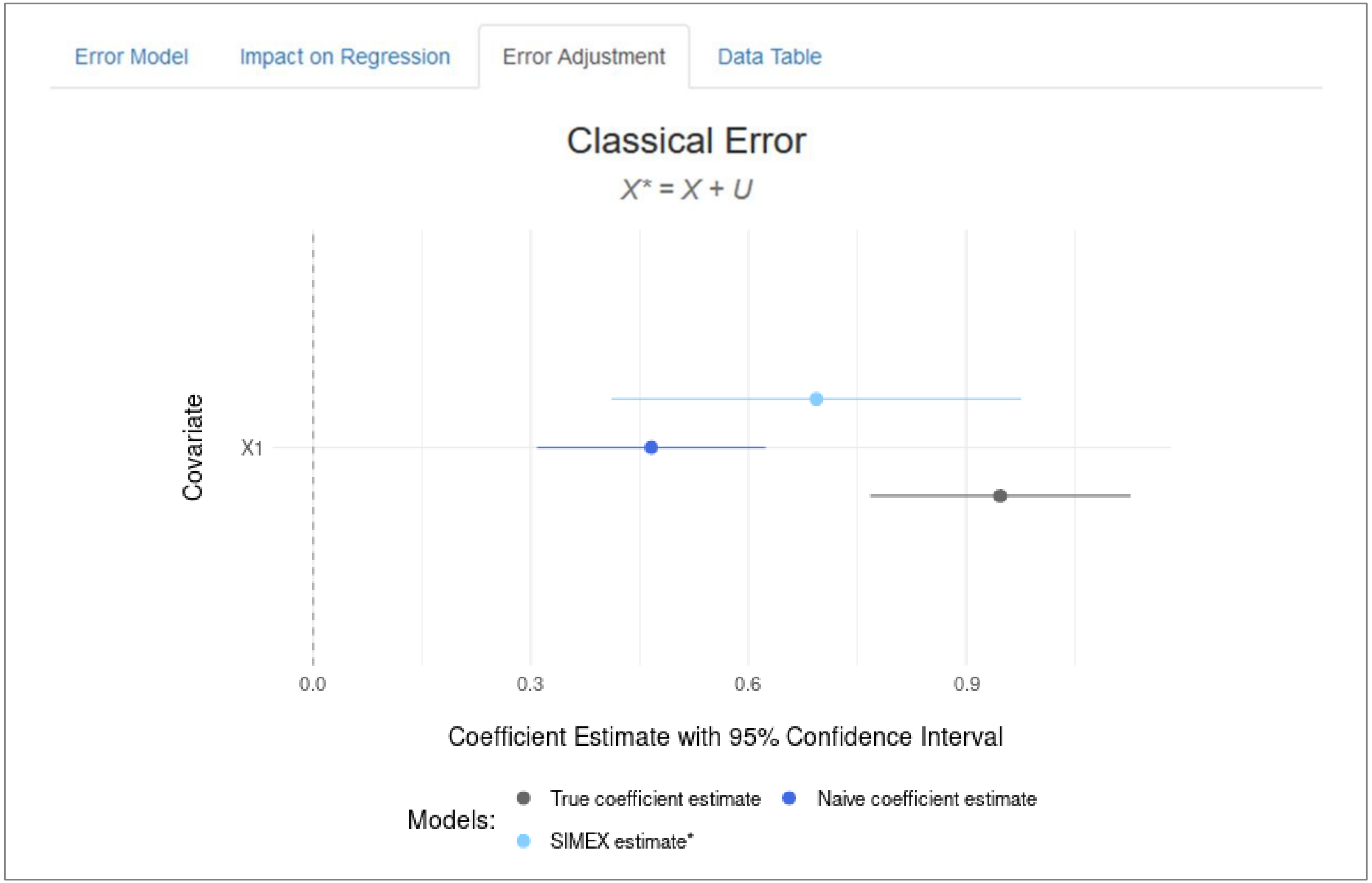
Covariate distribution

$$X \sim N(0,1)$$

Error model

- Additive, classical error
- Error variance





Summary & outlook

MEM-Explorer

- ▶ Shiny app for interactive exploration of measurement error and misclassification

- ▶ Find MEM-Explorer here: https://mem-explorer.shinyapps.io/MEMExplorer-v5/
- ▶ Ideas for improvement are welcome: vdeffner@bfs.de

Future work

- Options to specify parameters in MEM adjustment methods
- Integration of further MEM adjustment methods
- Measurement error in the outcome

Bundesamt für Strahlenschutz

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Impressum

Bundesamt für Strahlenschutz
Postfach 10 01 49
38201 Salzgitter

Tel.: +49 30 18333-0
Fax: +49 30 18333-1885
E-Mail: ePost@bfs.de

www.bfs.de

Kontakt für Rückfragen
Dr. Veronika Deffner
vdeffner@bfs.de
030/18333-2251