

Nonparametric Simulation Extrapolation for Measurement Error Models

Dylan Spicker*, Michael Wallace, Grace Yi

University of Waterloo

Thursday June 2, 2022

My Goals Today

1. Introduce **what** measurement error is.
2. Demonstrate **how** measurement error impacts analyses.
3. Introduce how **simulation extrapolation** can be used to overcome these concerns.
4. Provide a **nonparametric** extension to these methods.

An Illustrative Example

Suppose we want to determine the relationship between BMI and hypertensive status, controlling demographic factors.

An Illustrative Example

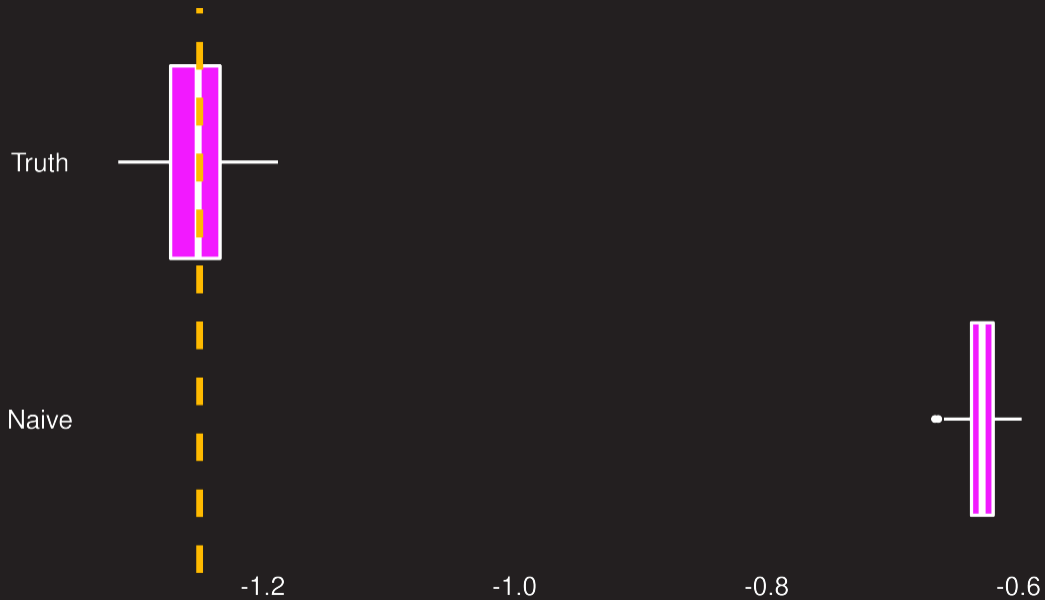
The study data that we are using has self-reported weight and height, in place of clinical measurements for most patients.

$$\text{Self-Reported BMI} = \text{True BMI} + \text{Noise}$$

Simulated Dataset

Patient #	Hypertension	Age	True BMI	Reported BMI $X + U$
1	Y_1	W_1	X_1	X_1^*
2	Y_2	W_2	X_2	X_2^*
		\vdots		
n_1	Y_{n_1}	W_{n_1}	X_{n_1}	$X_{n_1}^*$
$n_1 + 1$	Y_{n_1+1}	W_{n_1+1}	—	$X_{n_1+1}^*$
		\vdots		
n	Y_n	W_n	—	X_n^*

Goal: Determine the relationship given by $E[Y|X, W]$.



Basic Correction: Simulation Extrapolation

Step 1: Simulate additional measurement error, and compute the estimators of interest.

Step 2: Extrapolate this relationship to the case where no error is present.

Mathematical Intuition

$$X^* = X + U$$

Mathematical Intuition

$$X^* = X + U$$

$N(0, \sigma_U^2)$
↑
 U

Mathematical Intuition

$$X^*(\lambda) = X + \overset{N(0, \sigma_U^2)}{\uparrow} U + \overset{N(0, \lambda\sigma_U^2)}{\uparrow} \sqrt{\lambda}\sigma_U \epsilon$$

Mathematical Intuition

$$X^*(\lambda) = X + U + \sqrt{\lambda} \sigma_U \epsilon$$

The diagram illustrates the decomposition of the variance of $X^*(\lambda)$. The top node is $N(0, (1 + \lambda)\sigma_U^2)$. Two arrows point down to $N(0, \sigma_U^2)$ and $N(0, \lambda\sigma_U^2)$. From $N(0, \sigma_U^2)$, an arrow points down to U . From $N(0, \lambda\sigma_U^2)$, an arrow points down to $\sqrt{\lambda}\sigma_U\epsilon$.

Mathematical Intuition

$$X^*(\lambda) = X + U + \sqrt{\lambda}\sigma_U\epsilon$$

$N(0, \sigma_U^2)$ $N(0, \lambda\sigma_U^2)$

$N(0, (1 + \lambda)\sigma_U^2)$

$E[X^*(\lambda)|X] = X$

Mathematical Intuition

$$X^*(\lambda) = X + U + \sqrt{\lambda}\sigma_U\epsilon$$

$N(0, \sigma_U^2)$ $N(0, \lambda\sigma_U^2)$

$N(0, (1 + \lambda)\sigma_U^2)$

$E[X^*(\lambda)|X] = X$

$\text{var}[X^*(\lambda)|X] = (1 + \lambda)\sigma_U^2$

Mathematical Intuition

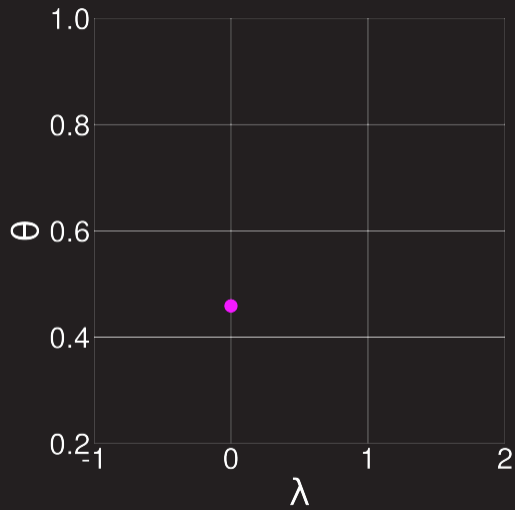
$$\begin{array}{ccc} N(0, (1 + \lambda)\sigma_U^2) & & \\ \nearrow & & \nwarrow \\ N(0, \sigma_U^2) & & N(0, \lambda\sigma_U^2) \end{array}$$

As $\lambda \rightarrow -1$, this tends to X .

$$E[X^*(\lambda)|X] = X$$

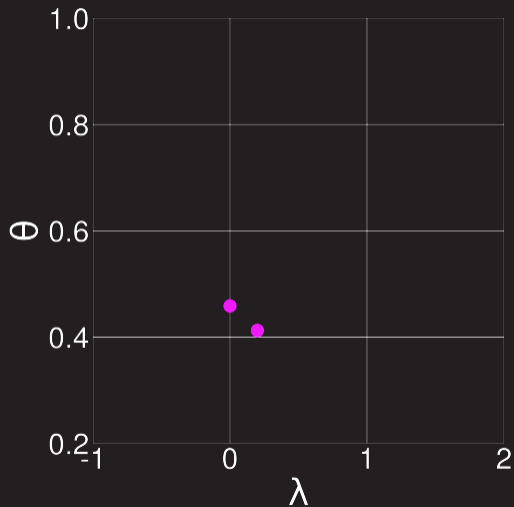
$$\text{var}[X^*(\lambda)|X] = (1 + \lambda)\sigma_U^2$$

Simulation Extrapolation



Correction Procedure

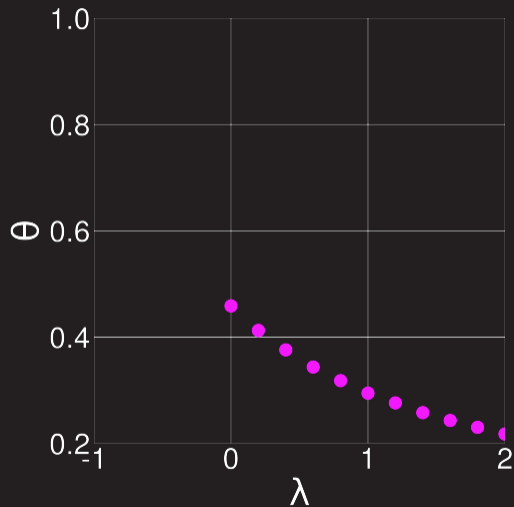
Simulation Extrapolation



Correction Procedure

1. Add extra measurement error and fit the model of interest.

Simulation Extrapolation



Correction Procedure

1. Add **extra measurement error** and fit the **model of interest**.
2. Repeat this for progressively **more measurement error**.

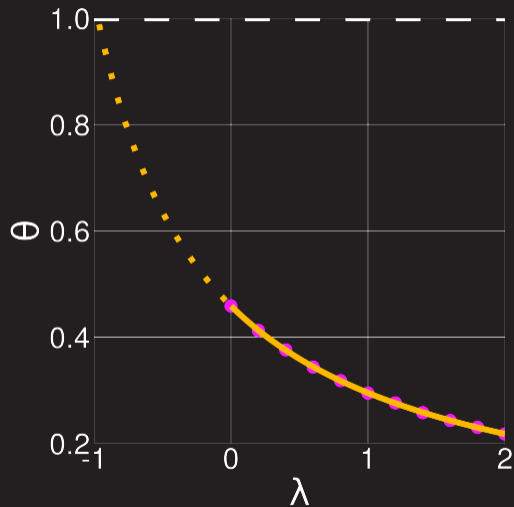
Simulation Extrapolation



Correction Procedure

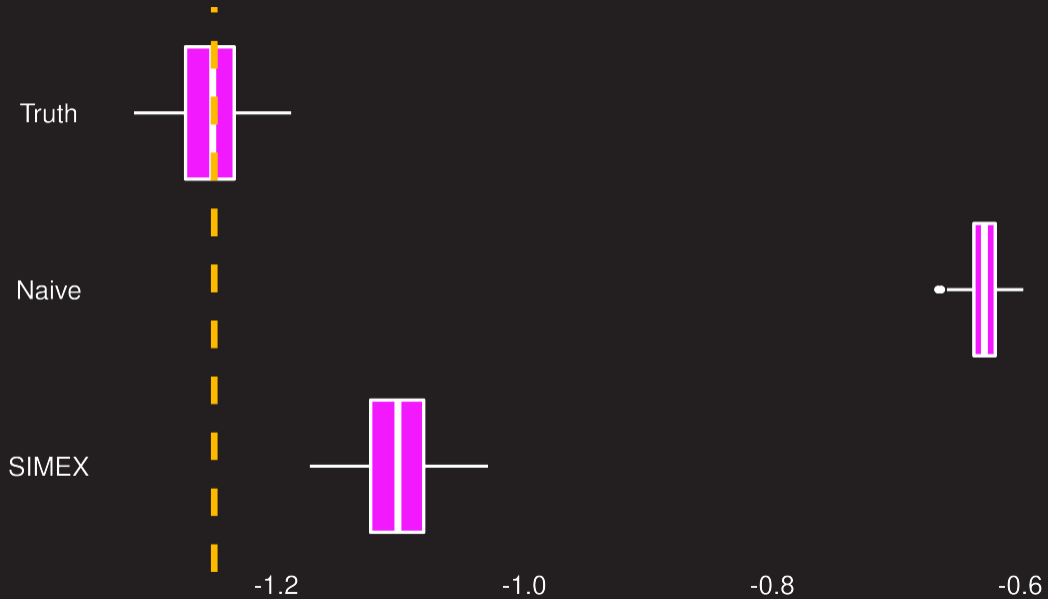
1. Add **extra measurement error** and fit the **model of interest**.
2. Repeat this for progressively **more measurement error**.
3. Predict the **outcome** based on the amount of **extra error** used.

Simulation Extrapolation



Correction Procedure

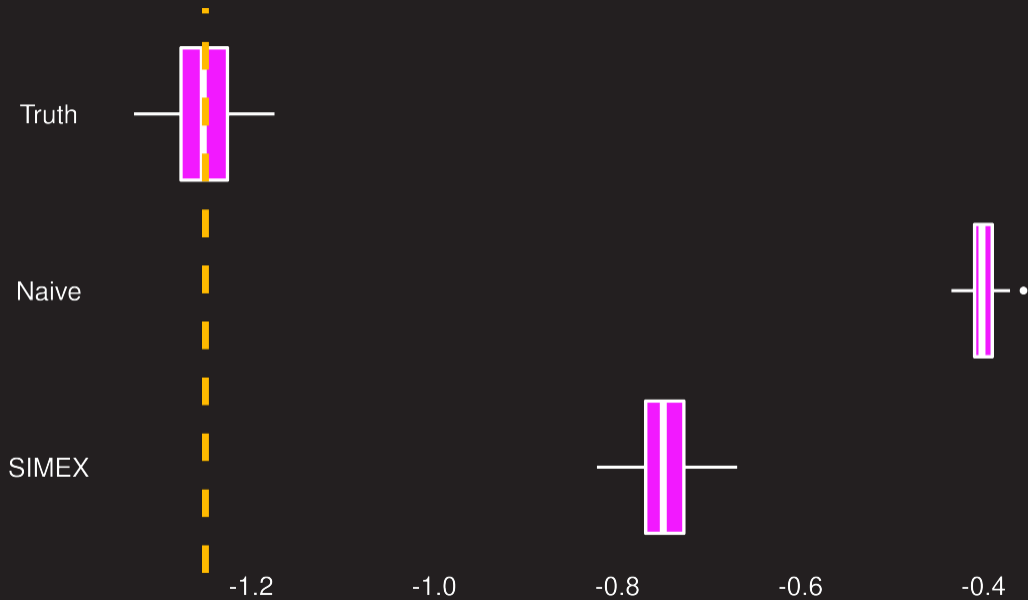
1. Add **extra measurement error** and fit the **model of interest**.
2. Repeat this for progressively **more measurement error**.
3. Predict the **outcome** based on the amount of **extra error** used.
4. **Extrapolate** to the case where there is **no error**.



The Problem

Is the assumption that U is normally distributed reasonable?

Oftentimes, no.



Our Solution: Nonparametric Simulation Extrapolation

$$X^*(\lambda) = X + U + \sum_{l=1}^{\lambda} \tilde{U}_l$$

Our Solution: Nonparametric Simulation Extrapolation

$$X^*(\lambda) = X + U + \sum_{l=1}^{\lambda} \tilde{U}_l$$

Drawn from \hat{F}_U



Our Solution: Nonparametric Simulation Extrapolation

$$X^*(\lambda) = X + U + \sum_{\ell=1}^{\lambda} \tilde{U}_{\ell}$$

Drawn from \hat{F}_U

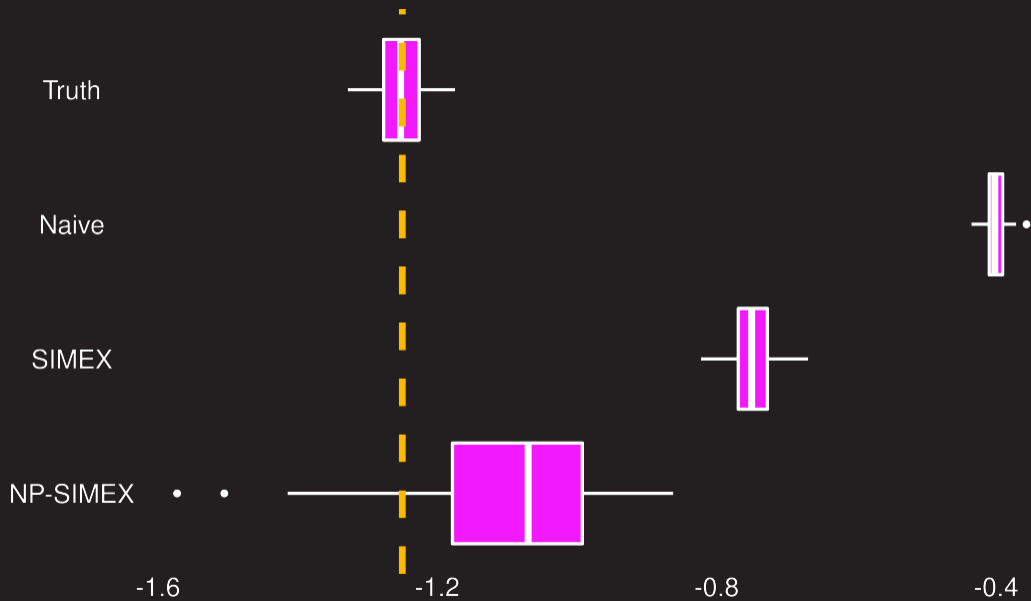


As $\lambda \rightarrow -1$, this has the same properties as before.

Software Implementation

An R implementation is available at
<https://github.com/DylanSpicker/np-simex>.

- `B` (defaults to `50`) - The number of re-sampled iterations to average over for each value of lambda.
- `parallel` (defaults to `TRUE`) - Whether the re-sampling should be parallelized or not. Implementations exist for both the `foreach` package and `parallel::mclapply`.
- `numCores` (defaults to `parallel::detectCores()/2`) - The number of cores to be used, if parallelization occurs.
- `est.variance` (defaults to `"none"`) - The method for estimating the variance. If it is provided as `"jackknife"` then the modified Jackknife procedure is used, otherwise no asymptotic variances are estimated.
- `parPackage` (defaults to `"foreach"`) - Which method for parallelizing is used, if this is anything other than `"foreach"`, and `parallel = TRUE`, then `parallel::mclapply` will be used.
- `smoothed` (defaults to `FALSE`) - Should smoothed density estimators be used, if so samples are drawn from the KDE estimate of the distribution of $\hat{\theta}$, rather than from the empirical distribution.
- `het` (defaults to `FALSE`) - Are the errors heterogenous, in that $\hat{\theta}$ and $\hat{\kappa}$ are dependent. If so conditional KDEs are used in place of the empirical error distribution.



Contributions of the NP-SIMEX

- ▶ Can account for errors with any distributional assumption assuming validation data are available.

Contributions of the NP-SIMEX

- ▶ Can account for errors with any distributional assumption assuming validation data are available.
- ▶ Can account for errors with any symmetric distributional assumption using replicate data.

Contributions of the NP-SIMEX

- ▶ Can account for errors with any distributional assumption assuming validation data are available.
- ▶ Can account for errors with any symmetric distributional assumption using replicate data.
- ▶ Can accommodate dependent errors which differ based on the true value of X .

Contributions of the NP-SIMEX

- ▶ Can account for errors with any distributional assumption assuming validation data are available.
- ▶ Can account for errors with any symmetric distributional assumption using replicate data.
- ▶ Can accommodate dependent errors which differ based on the true value of X .
- ▶ Results in asymptotically normal estimators.

Conclusions

By re-sampling from the empirical error distribution we can render the SIMEX estimators nonparametric, while maintaining their same, familiar form.

This is done with little additional complexity.

Thank You.

Dylan Spicker

dylan.spicker@uwaterloo.ca | www.dylanspicker.com