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

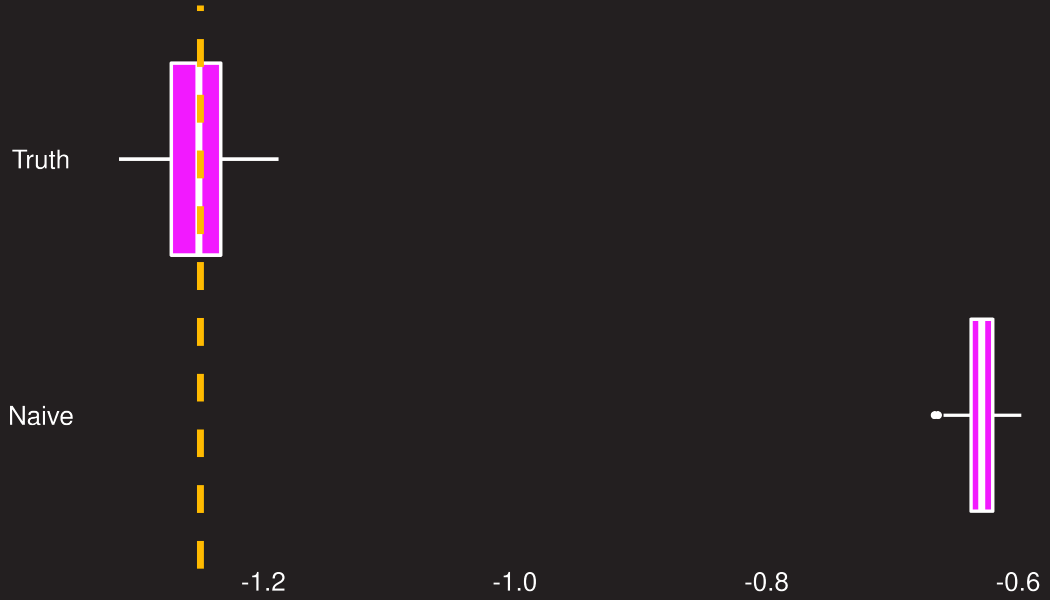
The study data that we are using has self-reported weight and height, in place of clinical measurements for most patients.

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

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 $E[Y|X, W]$.



Truth
Naive
-1.2
-1.0
-0.8
-0.6

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

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

$$\uparrow$$

$$X^* = X + U$$

Mathematical Intuition

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

Mathematical Intuition

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

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

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

Mathematical Intuition

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

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

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

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

Mathematical Intuition

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

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

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

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

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

Mathematical Intuition

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

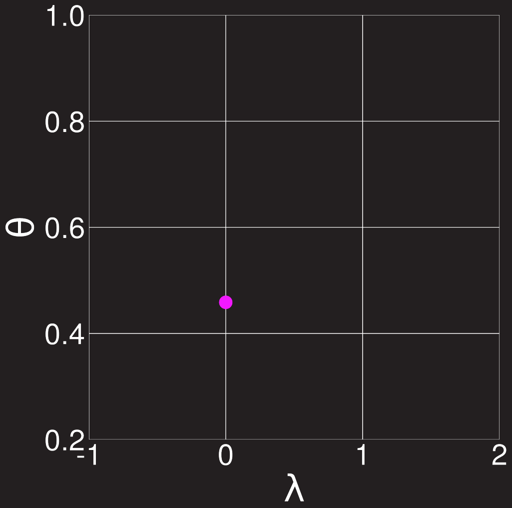
$$N(0, \sigma_U^2) \qquad N(0, \lambda\sigma_U^2)$$

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

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

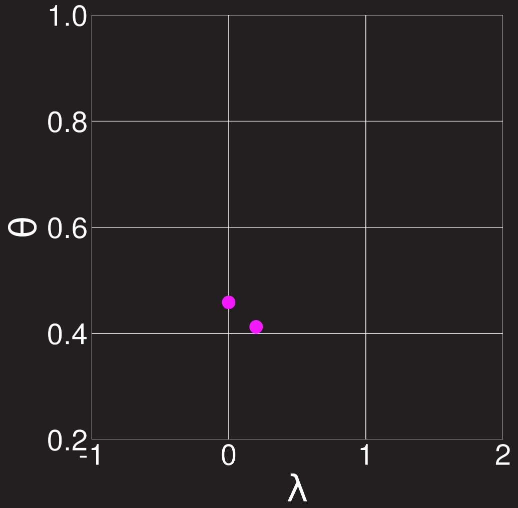
$$\text{var}[X^*(\lambda)|X] = (1+\lambda)\sigma_U^2$$

Simulation Extrapolation



Correction Procedure

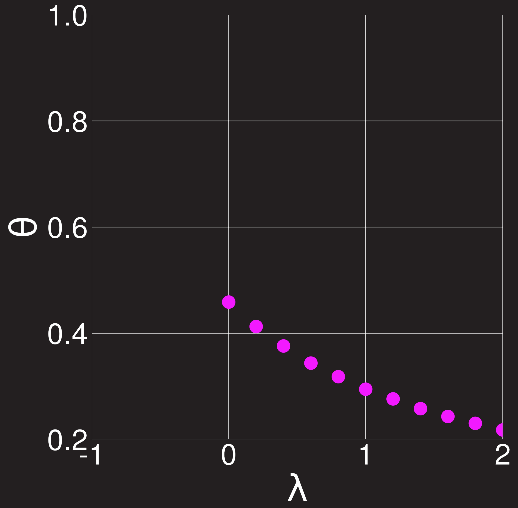
Simulation Extrapolation



Correction Procedure

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

Simulation Extrapolation



Correction Procedure

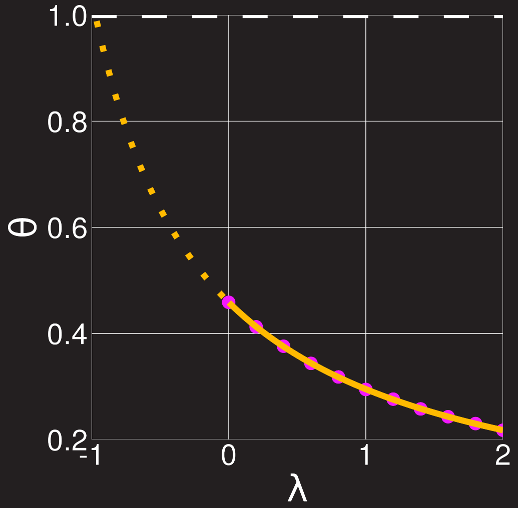
1. Add extra measurement error and fit the model of interest.
2. Repeat this for progressively more measurement error.

Simulation Extrapolation

Correction Procedure

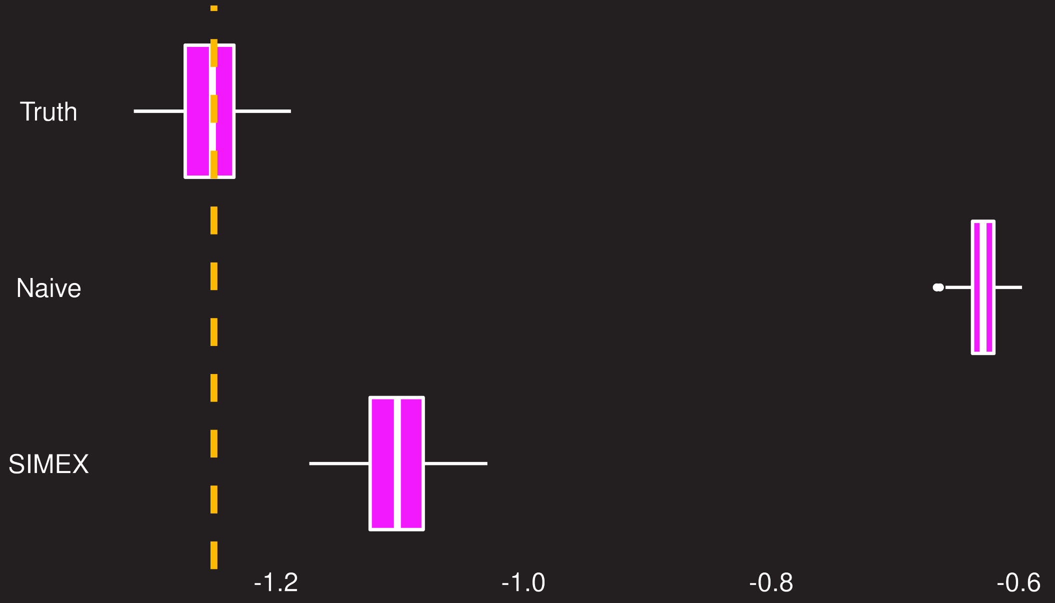
1. Add extra measurement error and fit the model of interest.
2. Repeat this for progressively more measurement error.
3. Predict the outcome based on the amount of extra error used.

Simulation Extrapolation



Correction Procedure

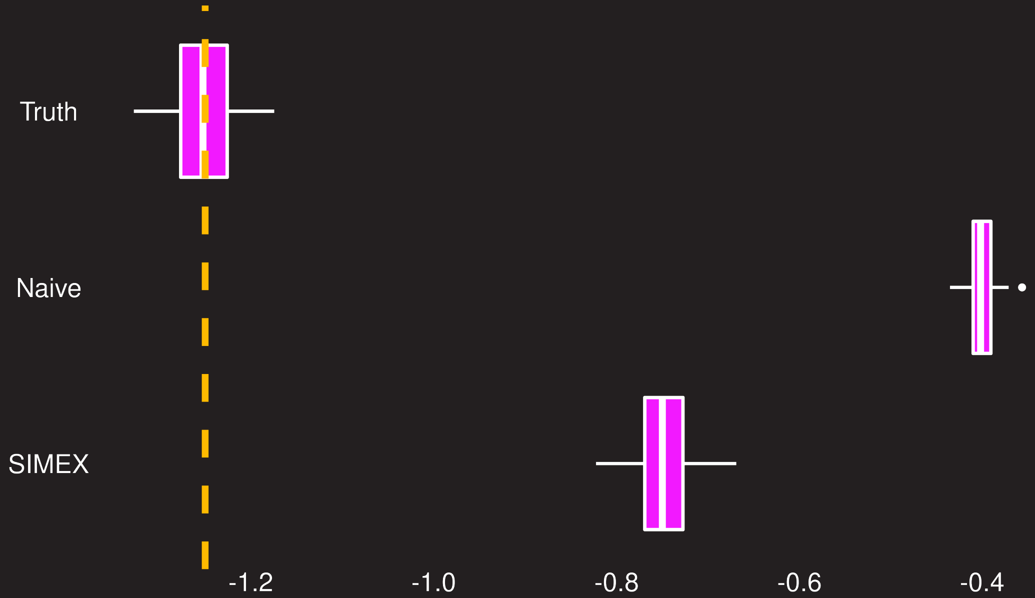
1. Add extra measurement error and fit the model of interest.
2. Repeat this for progressively more measurement error.
3. Predict the outcome based on the amount of extra error used.
4. Extrapolate to the case where there is no error.


Truth
Naive
SIMEX
-1.2
-1.0
-0.8
-0.6

The Problem

Is the assumption that U is normally distributed reasonable?

Oftentimes, no.


Truth
Naive
SIMEX
-1.2
-1.0
-0.8
-0.6
-0.4

Our Solution: Nonparametric Simulation Extrapolation

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

Our Solution: Nonparametric Simulation Extrapolation

Drawn from $\widehat{F}_U$

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

Our Solution: Nonparametric Simulation Extrapolation

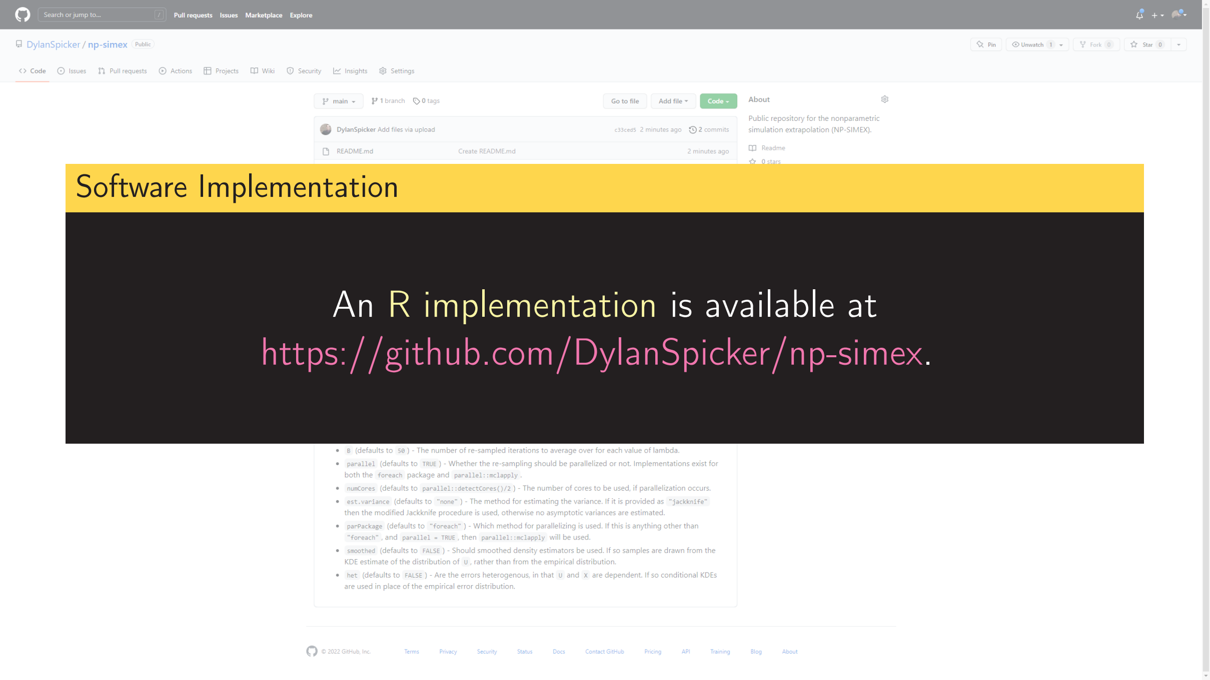
Drawn from $\widehat{F}_U$

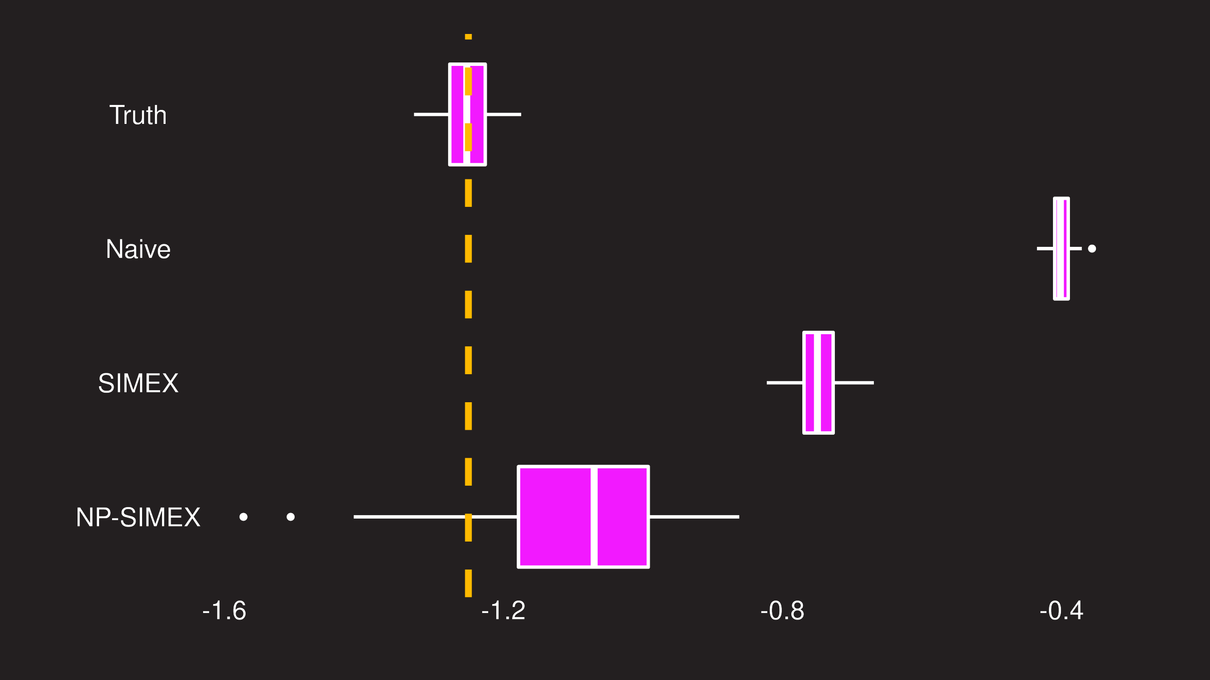
$$X^*(\lambda) = X + U + \sum_{\ell=1}^{\lambda} \widetilde{U}_\ell$$

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

Software Implementation

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





Truth
Naive
SIMEX
NP-SIMEX
-1.6
-1.2
-0.8
-0.4

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