Robust Bayesian Inference for Berkson and Classical Measurement Error Models

Harita Dellaporta, Theo Damoulas

March 2024

Outline

Motivation

Measurement error

- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N$$
$$N \sim \mathbb{F}_N^0, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E$$
$$E \sim \mathbb{F}_E^0, \quad \mathbb{E}[E] = 0.$$

Goal: to estimate θ_0 from (W, Y) while incorporating prior beliefs about $\mathbb{F}_N^0$ when:

1. Data from (W, Y) is available
2. $\mathbb{F}_N^0$ and $\mathbb{F}_E^0$ are unknown but some prior beliefs might be available
3. We assume additive homoscedastic errors in both X and Y
4. We assume non-differential ME, i.e. $Y \perp\!\!\!\perp W \mid X$

Measurement error

- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N$$
$$N \sim \mathbb{F}_N^0, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E$$
$$E \sim \mathbb{F}_E^0, \quad \mathbb{E}[E] = 0.$$

⬇

Goal: to estimate θ_0 from (W, Y) while incorporating prior beliefs about $\mathbb{F}_N^0$ when:

1. Data from (W, Y) is available
2. $\mathbb{F}_N^0$ and $\mathbb{F}_E^0$ are unknown but some prior beliefs might be available
3. We assume additive homoscedastic errors in both X and Y
4. We assume non-differential ME, i.e. $Y \perp\!\!\!\perp W \mid X$

Measurement error

- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N$$
$$N \sim \mathbb{F}_N^0, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E$$
$$E \sim \mathbb{F}_E^0, \quad \mathbb{E}[E] = 0.$$

Goal: to estimate θ_0 from (W, Y) while incorporating prior beliefs about $\mathbb{F}_N^0$ when:

1. Data from (W, Y) is available
2. $\mathbb{F}_N^0$ and $\mathbb{F}_E^0$ are unknown but some prior beliefs might be available
3. We assume additive homoscedastic errors in both X and Y
4. We assume non-differential ME, i.e. $Y \perp\!\!\!\perp W \mid X$

Measurement error

- Covariate X is only observed via a *noisy* proxy W such that:

$$X = W + N$$
$$N \sim \mathbb{F}_N^0, \quad \mathbb{E}[N] = 0$$

- Function $g : \Theta \times \mathcal{X} \to \mathbb{R}$ explains the relationship between X and Y such that:

$$Y = g(\theta_0, X) + E$$
$$E \sim \mathbb{F}_E^0, \quad \mathbb{E}[E] = 0.$$

Goal: to estimate θ_0 from (W, Y) while incorporating prior beliefs about $\mathbb{F}_N^0$ when:

1. Data from (W, Y) is available
2. $\mathbb{F}_N^0$ and $\mathbb{F}_E^0$ are unknown but some prior beliefs might be available
3. We assume additive homoscedastic errors in both X and Y
4. We assume non-differential ME, i.e. $Y \perp\!\!\!\perp W \mid X$

Two types of Measurement Error

Classical

Two types of Measurement Error

Classical

$$W = X + N$$
$$X \perp\!\!\!\perp N, \quad W \not\!\perp\!\!\!\perp N$$

Two types of Measurement Error

Classical

Noisy (observed)

$$W = X + N$$
$$X \perp\!\!\!\perp N, \quad W \not\perp\!\!\!\perp N$$

Two types of Measurement Error

Classical

Noise-free (unobserved)

Noisy (observed)

$$W = X + N$$

$$X \perp\!\!\!\perp N, \quad W \not\perp\!\!\!\perp N$$

Two types of Measurement Error

Classical

$$\overset{\text{Noisy (observed)}}{W} = \overset{\text{Noise-free (unobserved)}}{X} + \overset{\text{ME}}{N}$$

$$X \perp\!\!\!\perp N, \quad W \not\!\perp\!\!\!\perp N$$

Two types of Measurement Error

Classical

$$\overset{\text{Noisy (observed)}}{W} = \overset{\text{Noise-free (unobserved)}}{X} + \overset{\text{ME}}{N}$$

$$X \perp\!\!\!\perp N, \quad W \not\perp\!\!\!\perp N$$

Berkson

Two types of Measurement Error

Classical

$$\overset{\text{Noisy (observed)}}{W} = \overset{\text{Noise-free (unobserved)}}{X} + \overset{\text{ME}}{N}$$

$$X \perp\!\!\!\perp N, \quad W \not\perp\!\!\!\perp N$$

Berkson

$$X = W + N$$

$$W \perp\!\!\!\perp N, \quad X \not\perp\!\!\!\perp N$$

Example: Linear Regression

- Special case:

$$g(\theta_0, X) = \alpha X + \beta, \quad \theta_0 := (\alpha, \beta)$$

- For observations $\{w_i, y_i\}_{i=1}^n$ and associated unobserved errors $\{\nu_i\}_{i=1}^n \overset{\text{iid}}{\sim} \mathbb{F}_N^0$ and $\{\epsilon_i\}_{i=1}^n \overset{\text{iid}}{\sim} \mathbb{F}_E^0$ we have that for each $i = 1, \ldots, n$:

$$y_i = \alpha x_i + \beta + \epsilon_i, \quad w_i = x_i + \nu_i$$

- Naively using OLS would result in endogeneity bias since:

$$y_i = \alpha w_i + \beta + (\epsilon_i - \alpha \nu_i)$$

Covariate and error term are correlated!

Example: Linear Regression

- Special case:

$$g(\theta_0, X) = \alpha X + \beta, \quad \theta_0 := (\alpha, \beta)$$

- For observations $\{w_i, y_i\}_{i=1}^n$ and associated unobserved errors $\{\nu_i\}_{i=1}^n \overset{\text{iid}}{\sim} \mathbb{F}_N^0$ and $\{\epsilon_i\}_{i=1}^n \overset{\text{iid}}{\sim} \mathbb{F}_E^0$ we have that for each $i = 1, \ldots, n$:

$$y_i = \alpha x_i + \beta + \epsilon_i, \quad w_i = x_i + \nu_i$$

- Naively using OLS would result in endogeneity bias since:

$$y_i = \alpha w_i + \beta + (\epsilon_i - \alpha \nu_i)$$

Covariate and error term are correlated!

Example: Linear Regression

- Special case:

$$g(\theta_0, X) = \alpha X + \beta, \quad \theta_0 := (\alpha, \beta)$$

- For observations $\{w_i, y_i\}_{i=1}^n$ and associated unobserved errors $\{\nu_i\}_{i=1}^n \overset{\text{iid}}{\sim} \mathbb{F}_N^0$ and $\{\epsilon_i\}_{i=1}^n \overset{\text{iid}}{\sim} \mathbb{F}_E^0$ we have that for each $i = 1, \ldots, n$:

$$y_i = \alpha x_i + \beta + \epsilon_i, \quad w_i = x_i + \nu_i$$

- Naively using OLS would result in endogeneity bias since:

$$y_i = \alpha w_i + \beta + (\epsilon_i - \alpha \nu_i)$$

Covariate and error term are correlated!

Background

Existing approaches

- Deming regression (Deming, 1943)
- Simulation Extrapolation method (SIMEX) (Cook and Stefanski, 1994) which assumes knowledge or estimates measurement error variance)
- Instrumental Variable approaches (Newhouse and McClellan, 1998)
- Nonparametric approaches: Deconvolution Kernel Estimator (Fan and Truong, 1993; Wang and Wang, 2011), Gaussian Processes regression Cervone and Pillai (2015); Zhou et al. (2023)
- Bayesian semi-parametric approach with penalised splines (Berry et al., 2002; Sarkar et al., 2014)

Bayesian Nonparametric Learning (NPL) framework (Lyddon et al., 2018; Fong et al., 2019)

1. Standard Bayesian inference assumes that the model is well-specified and sets uncertainty directly on the parameter of interest
2. Place a nonparametric prior *directly* on the data-generating mechanism $\mathbb{P}^*$:

$$\mathbb{P} \sim DP(\alpha, \mathbb{F}), \quad \mathbb{P}|x_{1:n} \sim DP\left(\alpha', \mathbb{F}'\right)$$

where

$$\alpha' = \alpha + n, \quad \mathbb{F}' := \frac{\alpha}{\alpha+n}\mathbb{F} + \frac{n}{n+\alpha}\mathbb{P}_n, \quad \mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$$

3. For a loss function $l(x, \theta)$ *propagate uncertainty* from $\mathbb{P}^*$ to the parameter of interest θ through

$$\boxed{\theta_l^*(\mathbb{P}^*) := \arg\inf_{\theta\in\Theta} \mathbb{E}_{X\sim\mathbb{P}^*}[l(X, \theta)]}$$

Bayesian Nonparametric Learning (NPL) framework (Lyddon et al., 2018; Fong et al., 2019)

1. Standard Bayesian inference assumes that the model is well-specified and sets uncertainty directly on the parameter of interest
2. Place a nonparametric prior *directly* on the data-generating mechanism $\mathbb{P}^*$:

$$\mathbb{P} \sim DP(\alpha, \mathbb{F}), \quad \mathbb{P}|x_{1:n} \sim DP\left(\alpha', \mathbb{F}'\right)$$

   where

$$\alpha' = \alpha + n, \quad \mathbb{F}' := \frac{\alpha}{\alpha+n}\mathbb{F} + \frac{n}{n+\alpha}\mathbb{P}_n, \quad \mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$$

3. For a loss function $l(x, \theta)$ *propagate uncertainty* from $\mathbb{P}^*$ to the parameter of interest θ through

$$\boxed{\theta_l^*(\mathbb{P}^*) := \arg\inf_{\theta\in\Theta} \mathbb{E}_{X\sim\mathbb{P}^*}[l(X, \theta)]}$$

Bayesian Nonparametric Learning (NPL) framework (Lyddon et al., 2018; Fong et al., 2019)

1. Standard Bayesian inference assumes that the model is well-specified and sets uncertainty directly on the parameter of interest

2. Place a nonparametric prior *directly* on the data-generating mechanism $\mathbb{P}^*$:

$$\mathbb{P} \sim DP(\alpha, \mathbb{F}), \quad \mathbb{P}|x_{1:n} \sim DP\left(\alpha', \mathbb{F}'\right)$$

where

$$\alpha' = \alpha + n, \quad \mathbb{F}' := \frac{\alpha}{\alpha+n}\mathbb{F} + \frac{n}{n+\alpha}\mathbb{P}_n, \quad \mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$$

3. For a loss function $l(x, \theta)$ *propagate uncertainty* from $\mathbb{P}^*$ to the parameter of interest θ through

$$\boxed{\theta_l^*(\mathbb{P}^*) := \arg\inf_{\theta \in \Theta} \mathbb{E}_{X \sim \mathbb{P}^*}[l(X, \theta)]}$$

Methodology

Robustness with respect to ME

Conditional model family: $\boxed{\mathcal{P}_\Theta^{Y\mid X} = \{(\mathbb{P}_{g(\theta,x)})_{x\in\mathcal{X}} : \theta \in \Theta\}}$

- $\mathbb{P}_X := \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \rightarrow \boxed{\mathcal{P}_\Theta^{X,Y} = \{\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\, \mathbb{P}_{g(\theta,x_i)} : \theta \in \Theta\}}$
- $\mathbb{P}_W := \frac{1}{n}\sum_{i=1}^n \delta_{w_i} \rightarrow \boxed{\mathcal{P}_\Theta^{W,Y} = \{\frac{1}{n}\sum_{i=1}^n \delta_{w_i}\, \mathbb{P}_{g(\theta,w_i)} : \theta \in \Theta\}}$

Robustness with respect to ME

Conditional model family: $\boxed{\mathcal{P}_\Theta^{Y\mid X} = \{(\mathbb{P}_{g(\theta,x)})_{x\in\mathcal{X}} : \theta \in \Theta\}}$

- $\mathbb{P}_X := \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \rightarrow \boxed{\mathcal{P}_\Theta^{X,Y} = \{\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\,\mathbb{P}_{g(\theta,x_i)} : \theta \in \Theta\}}$
- $\mathbb{P}_W := \frac{1}{n}\sum_{i=1}^n \delta_{w_i} \rightarrow \boxed{\mathcal{P}_\Theta^{W,Y} = \{\frac{1}{n}\sum_{i=1}^n \delta_{w_i}\,\mathbb{P}_{g(\theta,w_i)} : \theta \in \Theta\}}$

Robustness with respect to ME

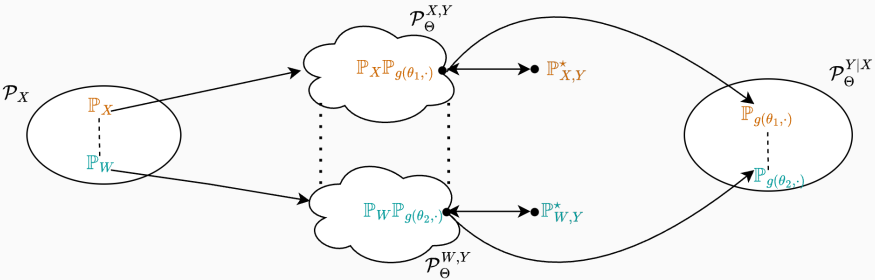
Conditional model family: $\boxed{\mathcal{P}_\Theta^{Y\,|\,X} = \{(\mathbb{P}_{g(\theta,x)})_{x\in\mathcal{X}} : \theta \in \Theta\}}$

- $\mathbb{P}_X := \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \to \boxed{\mathcal{P}_\Theta^{X,Y} = \{\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\,\mathbb{P}_{g(\theta,x_i)} : \theta \in \Theta\}}$
- $\mathbb{P}_W := \frac{1}{n}\sum_{i=1}^n \delta_{w_i} \to \boxed{\mathcal{P}_\Theta^{W,Y} = \{\frac{1}{n}\sum_{i=1}^n \delta_{w_i}\,\mathbb{P}_{g(\theta,w_i)} : \theta \in \Theta\}}$

Robustness with respect to ME

Conditional model family: $\boxed{\mathcal{P}_\Theta^{Y \mid X} = \{(\mathbb{P}_{g(\theta,x)})_{x\in\mathcal{X}} : \theta \in \Theta\}}$

- $\mathbb{P}_X := \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \rightarrow \boxed{\mathcal{P}_\Theta^{X,Y} = \{\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\,\mathbb{P}_{g(\theta,x_i)} : \theta \in \Theta\}}$
- $\mathbb{P}_W := \frac{1}{n}\sum_{i=1}^n \delta_{w_i} \rightarrow \boxed{\mathcal{P}_\Theta^{W,Y} = \{\frac{1}{n}\sum_{i=1}^n \delta_{w_i}\,\mathbb{P}_{g(\theta,w_i)} : \theta \in \Theta\}}$



Robust-MEM

- Target parameter:

$$\theta_l^*(\mathbb{P}^\star_{X,Y}) = \arg\min_{\theta\in\Theta} \mathbb{E}_{(X,Y)\sim\mathbb{P}^\star_{X,Y}}[l(x,y;\theta)]$$

- Set uncertainty on the true distribution of $X \mid W = w_i$, denoted by $\mathbb{P}^*_{X \mid w_i}$
- Dirichlet Process (DP) prior for each $i = 1, \ldots, n$:

$$\mathbb{P}_i \sim \mathrm{DP}(c, \mathbb{F}_{w_i})$$

- DP posterior:

$$\mathbb{P}_i \mid w_i \sim \mathrm{DP}(c+1, \mathbb{F}'_{w_i}), \quad \mathbb{F}'_{w_i} = \frac{1}{c+1}\delta_{w_i} + \frac{c}{c+1}\mathbb{F}_{w_i}$$

Prior specification

$$\boxed{\mathbb{P}_i \mid w_i \sim \mathrm{DP}(c + 1, \mathbb{F}'_{w_i}), \quad \mathbb{F}'_{w_i} = \frac{1}{c+1}\delta_{w_i} + \frac{c}{c+1}\mathbb{F}_{w_i}}$$

- $c = 0$: all the mass of the posterior centering measure concentrated at the observation δ_{x_i} $\rightarrow$ no measurement error on the covariates
- $c = 1$: weighting in $\mathbb{F}'$ equally split between the prior centering measure $\mathbb{F}$ and the Dirac measure on the observation x_i.
- $c \rightarrow \infty$: no weighting is imposed on x_i

Prior specification

$$\boxed{\mathbb{P}_i \mid w_i \sim \mathrm{DP}(c+1, \mathbb{F}'_{w_i}), \quad \mathbb{F}'_{w_i} = \frac{1}{c+1}\delta_{w_i} + \frac{c}{c+1}\mathbb{F}_{w_i}}$$

- Berkson ($W \perp\!\!\!\perp N$):
  - $\mathbb{F}_N$ represents prior beliefs about the ME distribution
  - $x \sim \mathbb{F}_{w_i}$ is such that $x = w_i + \nu$ where $\nu \sim \mathbb{F}_N$
- Classical ($X \perp\!\!\!\perp N$):
  - $\mathbb{F}_{W \mid X}$ with density $f_{W \mid X}$ represents prior beliefs about the ME distribution, $\mathbb{F}_X$ with density f_X represents prior beliefs about the marginal distribution of *true* covariate X
  - $x \sim \mathbb{F}_{w_i}$ is such that

$$f_{w_i}(x \mid w_i) = \frac{f_{W \mid X}(w_i \mid x) f_X(x)}{\int f_{W \mid X}(w_i \mid x) f_X(x) dx}$$

Uncertainty propagation

- Target parameter:

$$\theta_l^*(\mathbb{P}^\star_{X,Y}) = \arg\min_{\theta\in\Theta} \mathbb{E}_{(x,y)\sim\mathbb{P}^\star_{X,Y}}[l(x,y;\theta)]$$

- DP posterior on the true distribution of $X \mid W = w_i$, denoted by $\mathbb{P}^*_{X \mid w_i}$:

$$\mathbb{P}_i \mid w_i \sim \mathrm{DP}(c+1, \mathbb{F}'_{w_i}), \quad \mathbb{F}'_{w_i} = \frac{1}{c+1}\delta_{w_i} + \frac{c}{c+1}\mathbb{F}_{w_i}$$

- Posterior Bootstrap; at iteration j:

1. For $i = 1, \ldots, n:$ sample $\mathbb{P}^{(i,j)} \sim \mathrm{DP}(c+1, \mathbb{F}'_{w_i})$.
2. Calculate $\theta_l^*(\mathbb{P}^{(1,j)}, \ldots, \mathbb{P}^{(n,j)}) = \arg\min_{\theta\in\Theta} \mathbb{E}_{(x,y)\sim\frac{1}{n}\sum_{i=1}^n \mathbb{P}^{(i,j)}\,\delta_{y_i}}[l(x,y,\theta)]$.

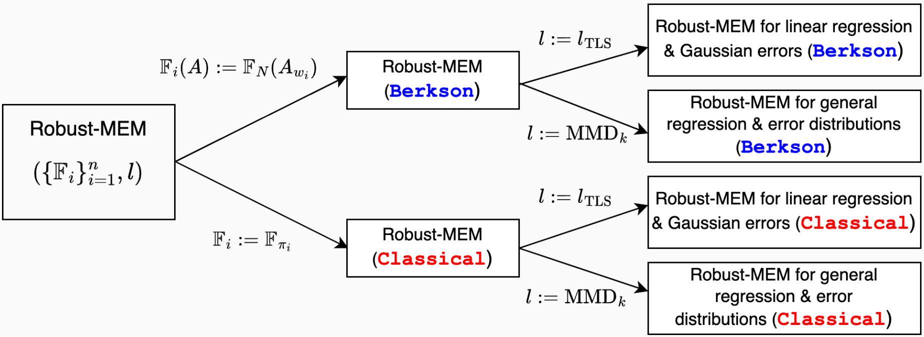
Loss function

- *Total Least Squares* (Golub and Loan, 1979; Golub and Van Loan, 1980)
  - Linear Regression
  - Underlying Gaussian errors assumption
- *Maximum Mean Discrepancy* (Briol et al., 2019; Chérief-Abdellatif and Alquier, 2022; Alquier and Gerber, 2020)
  - Nonlinear regression
  - Suitable for any error distribution
  - Able to handle likelihood-free models
  - Robust towards model misspecification on error distribution of Y

Loss function

- *Total Least Squares* (Golub and Loan, 1979; Golub and Van Loan, 1980)
    - Linear Regression
    - Underlying Gaussian errors assumption
- *Maximum Mean Discrepancy* (Briol et al., 2019; Chérief-Abdellatif and Alquier, 2022; Alquier and Gerber, 2020)
    - Nonlinear regression
    - Suitable for any error distribution
    - Able to handle likelihood-free models
    - Robust towards model misspecification on error distribution of Y

Summary of Robust MEM



Generalisation error

Generalisation error

Theorem (Berkson)

$$\overbrace{\mathbb{E}\left[\mathrm{MMD}_k(\mathbb{P}^*_{X,Y}, \mathbb{P}_{\theta^\star(\mathbb{P})})\right] - \inf_{\theta\in\Theta} \mathrm{MMD}_k(\mathbb{P}^*_{X,Y}, \mathbb{P}_\theta)}^{\textit{Generalisation error}}$$

$$\leq \frac{2}{\sqrt{n}} + 2(1+\Lambda)\left(\frac{1}{\sqrt{n}(c+1)} + \sqrt{\frac{1}{c+2}} + \right.$$

$$\left. + \underbrace{\frac{c}{c+1}\mathrm{MMD}_{k^2_X}(\mathbb{F}_N, \mathbb{F}^0_N)}_{\textit{Prior specification}} + \underbrace{\frac{1}{c+1}\mathrm{MMD}_{k^2_X}(\mathbb{F}^0_N, \delta_0)}_{\textit{ME deviation from 0}}\right).$$

Generalisation Error / Role of c

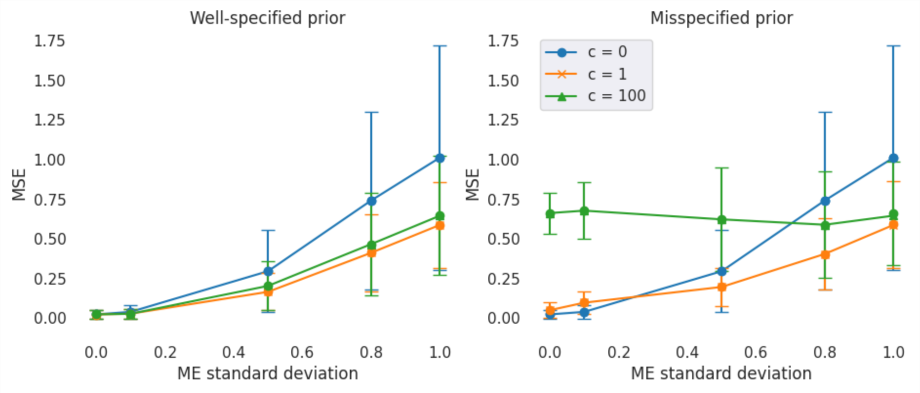
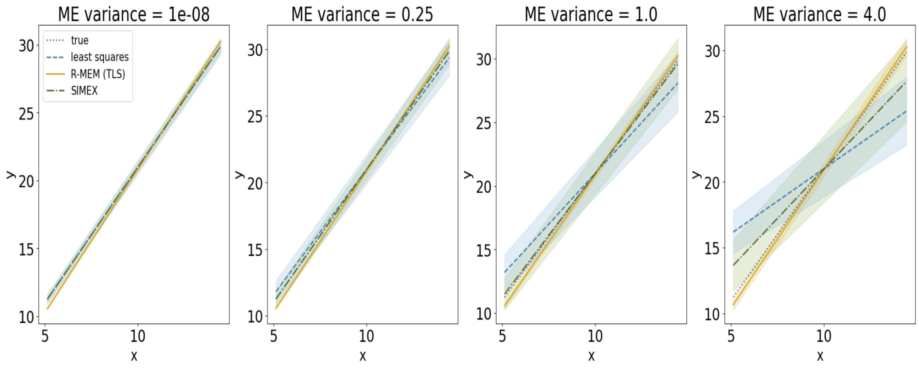


Figure 1: Average MSE of polynomial regression parameters in the presence of Berkson, Gaussian ME with increasing standard deviation over 50 replications.

Examples

Example: Linear Regression with Classical Measurement Error

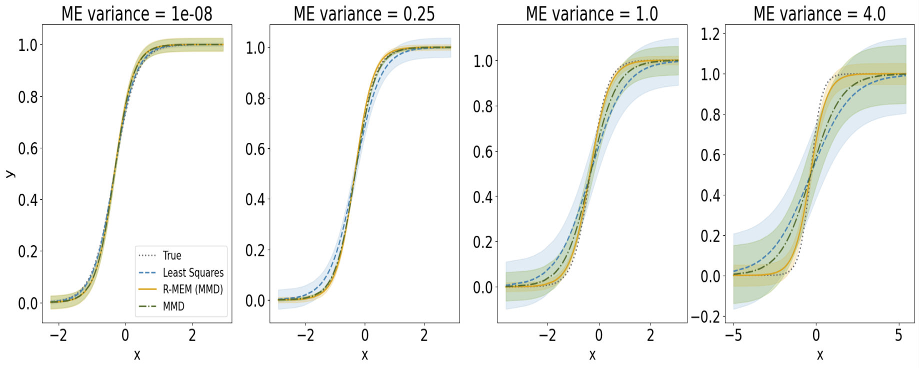
$$y = \theta_1 + \theta_2 x + \epsilon, \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$
$$w = x + \nu, \quad \nu \sim N(0, \sigma_\nu^2).$$



Example: Nonlinear Regression with Berkson Measurement Error

$$y = \frac{\exp(a+bx)}{1+\exp(a+bx)} + \epsilon, \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

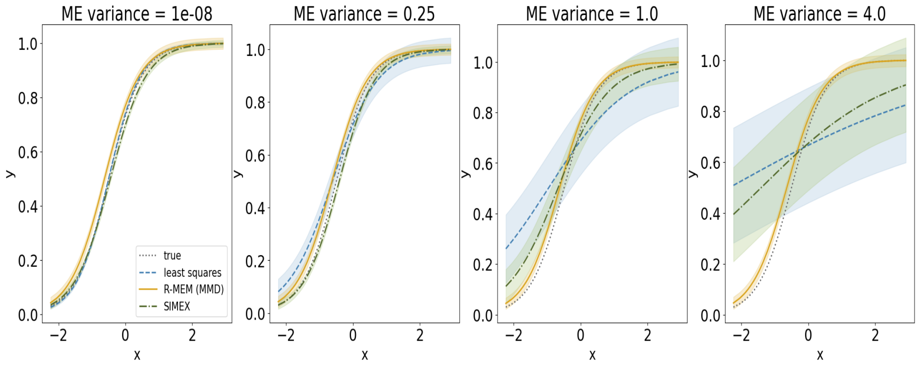
$$x = w + \nu, \quad \nu \sim N(0, \sigma_\nu^2)$$



Example: Nonlinear Regression with Classical Measurement Error

$$y = \frac{\exp(a+bx)}{1+\exp(a+bx)} + \epsilon, \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

$$w = x + \nu, \quad \nu \sim N(0, \sigma_\nu^2)$$



Discussion

- Most suitable for when we think there might be (`Berkson` or `Classical`) ME but are uncertain about its distribution or size
- Prior specification is important
- Kernel choice and hyperparameters
- Optimisation convergence

Discussion

- Most suitable for when we think there might be (`Berkson` or `Classical`) ME but are uncertain about its distribution or size
- Prior specification is important
- Kernel choice and hyperparameters
- Optimisation convergence

Discussion

- Most suitable for when we think there might be (`Berkson` or `Classical`) ME but are uncertain about its distribution or size
- Prior specification is important
- Kernel choice and hyperparameters
- Optimisation convergence

Discussion

- Most suitable for when we think there might be (`Berkson` or `Classical`) ME but are uncertain about its distribution or size
- Prior specification is important
- Kernel choice and hyperparameters
- Optimisation convergence

Discussion

- Most suitable for when we think there might be (`Berkson` or `Classical`) ME but are uncertain about its distribution or size
- Prior specification is important
- Kernel choice and hyperparameters
- Optimisation convergence

Alquier, P. and Gerber, M. (2020). Universal robust regression via maximum mean discrepancy. *arXiv preprint arXiv:2006.00840*.

Berry, S. M., Carroll, R. J., and Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, 97(457):160–169.

Briol, F.-X., Barp, A., Duncan, A. B., and Girolami, M. (2019). Statistical inference for generative models with maximum mean discrepancy. *arXiv preprint arXiv:1906.05944*.

Cervone, D. and Pillai, N. S. (2015). Gaussian process regression with location errors. *arXiv preprint arXiv:1506.08256*.

Chérief-Abdellatif, B.-E. and Alquier, P. (2022). Finite sample properties of parametric mmd estimation: robustness to misspecification and dependence. *Bernoulli*, 28(1):181–213.

Cook, J. R. and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical association*, 89(428):1314–1328.

References ii

Deming, W. E. (1943). Statistical adjustment of data.

Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics*, pages 1900–1925.

Fong, E., Lyddon, S., and Holmes, C. (2019). Scalable nonparametric sampling from multimodal posteriors with the posterior bootstrap. In *International Conference on Machine Learning*, pages 1952–1962. PMLR.

Golub, G. H. and Loan, C. V. (1979). Total least squares. In *Smoothing Techniques for Curve Estimation*, pages 69–76. Springer.

Golub, G. H. and Van Loan, C. F. (1980). An analysis of the total least squares problem. *SIAM journal on numerical analysis*, 17(6):883–893.

Lyddon, S., Walker, S., and Holmes, C. C. (2018). Nonparametric learning from bayesian models with randomized objective functions. *Advances in neural information processing systems*, 31.

Newhouse, J. P. and McClellan, M. (1998). Econometrics in outcomes research: the use of instrumental variables. *Annual review of public health*, 19(1):17–34.

References iii

Sarkar, A., Mallick, B. K., and Carroll, R. J. (2014). Bayesian semiparametric regression in the presence of conditionally heteroscedastic measurement and regression errors. *Biometrics*, 70(4):823–834.

Wang, X.-F. and Wang, B. (2011). Deconvolution estimation in measurement error models: the r package decon. *Journal of statistical software*, 39(10).

Zhou, S., Pati, D., Wang, T., Yang, Y., and Carroll, R. J. (2023). Gaussian processes with errors in variables: Theory and computation. *Journal of Machine Learning Research*, 24(87):1–53.