A JOINT BAYESIAN FRAMEWORK FOR MEASUREMENT ERROR AND MISSING DATA

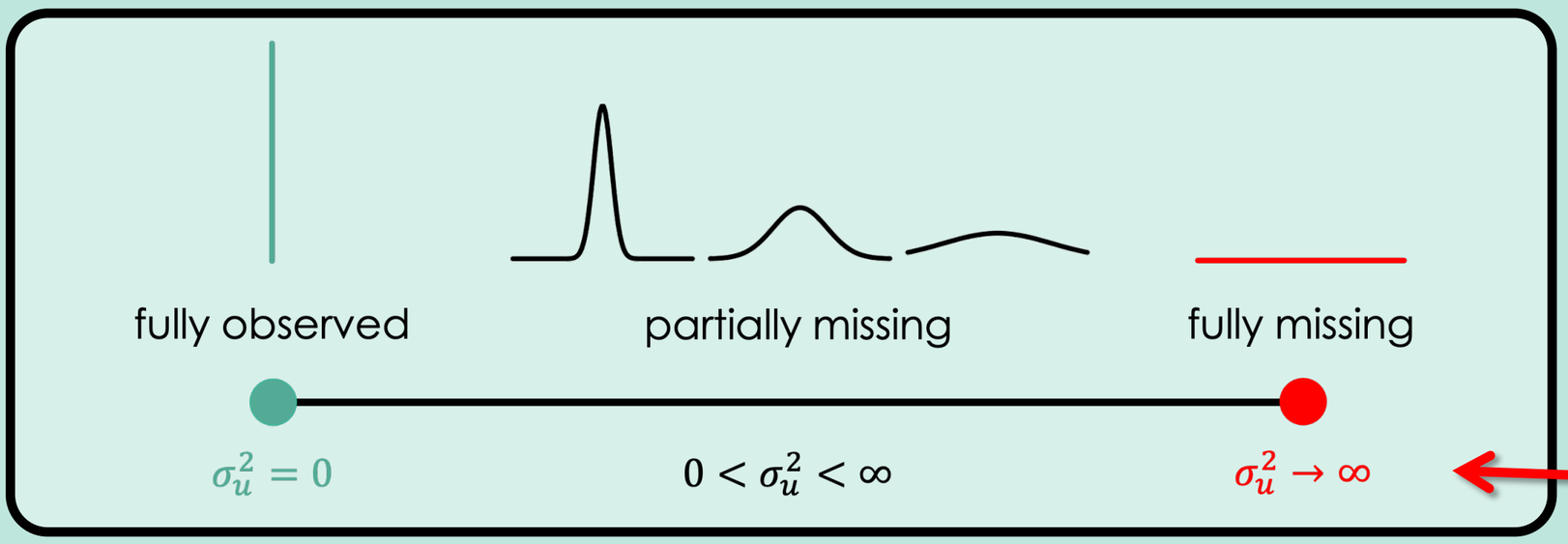
EMMA SKARSTEIN[1] AND STEFANIE MUFF[1,2]

[1]Department of Mathematical Sciences (IMF, NTNU, Trondheim), [2]Centre for Biodiversity Dynamics (CBD, NTNU, Trondheim)

1. MOTIVATION

Although measurement error in covariates has been studied in debth, many applied scientists still do not (know how to) deal with it.

- We aim to develop a method for viewing missing data as a limiting case of measurement error, allowing these two problems to be handled in the same framework.
- A Bayesian hierarchical structure provides a natural flexible framework in order to also use prior knowledge about the measurement error.
- The methods can then be efficiently implemented for potentially complex models using integrated nested Laplace approximations (INLA)(Rue et al., 2009, Muff et al., 2015).



2. CLASSICAL ERROR MODEL

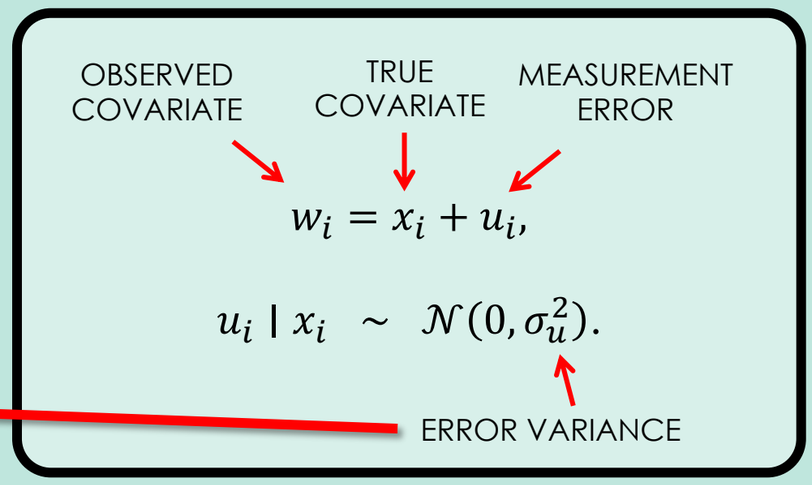


Figure 1: The continuum of measurement error, with observation-level priors illustrated in the top row. From Blackwell et al. (2017)

3. HIERARCHICAL CLASSICAL ERROR MODEL

(Muff et al., 2015, Goldstein et al., 2018)

$$\eta_i = \beta_0 + \beta_x x_i + \boldsymbol{z_i}\boldsymbol{\beta_z} + \varepsilon_i \quad \text{model of interest}$$
$$w_i = x_i + u_i \quad \text{error model}$$
$$x_i = \alpha_0 + \boldsymbol{z_i}\boldsymbol{\alpha_z} + \varepsilon_i^{(x)} \quad \text{exposure model}$$

Model of interest: η_i is the linear predictor in a generalized linear model (GLM), given the true covariate values for x_i, as well as other covariates $\boldsymbol{z_i}$, which are observed without error.
Error model: u_i is the error in the observed variable w_i, where $u_i \sim \mathcal{N}(0, \sigma_{u_i}^2)$.
Exposure model: Describes the true covariate x_i, which possibly depends on the correctly observed covariates $\boldsymbol{z_i}$.

4. EXAMPLE: SIMULATION STUDY

DATA SIMULATION

$$\begin{bmatrix} X_1 \\ V \\ X_3 \end{bmatrix} \sim \mathcal{N}(0, \Sigma), \qquad \Sigma = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}$$

$$X_2 = \begin{cases} 0, & \text{if } V < 0 \\ 1, & \text{otherwise} \end{cases}$$

$$Y = 1 + X_1 + X_2 + X_3 + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, 1)$$

Measurement error:
$W = X_1 + U, \quad U \sim \mathcal{N}(0, 0.25)$

Missing data:
20% of observations in W are removed

MODEL

Hierarchichal model to account for measurement error and missing data:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \quad \text{model of interest}$$
$$w_i = x_{1i} + u_i \quad \text{error model}$$
$$x_{1i} = \alpha_0 + \alpha_1 x_{2i} + \alpha_2 x_{3i} + \varepsilon_i^{(x_1)} \quad \text{exposure model}$$

RESULTS

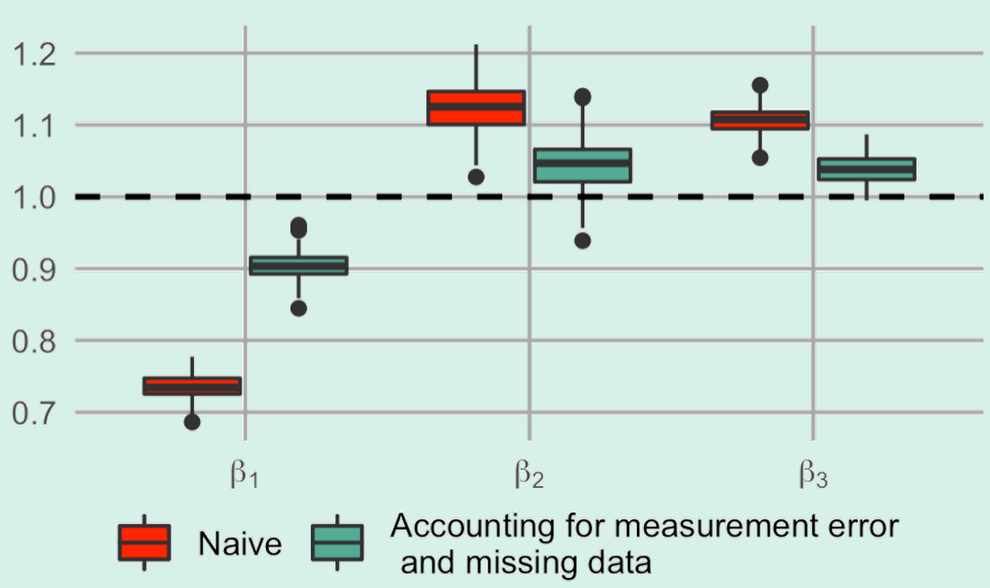


Figure 2: The estimated coefficients for models not accounting for measurement error (in red), and one accounting for measurement error (in green), fitted for 200 simulated datasets of 1000 observations each. True coefficients are $\beta_1 = \beta_2 = \beta_3 = 1$.

5. FUTURE WORK

- **MEASUREMENT ERROR AND MISSING DATA IN INLA R-PACKAGE**: The described work will be implemented in an R-package with extensive documentation.
- **DIFFERENT ERROR TYPES**: Berkson errors and non-differential errors will also be explored.
- **SHOULD MEASUREMENT ERROR BE ACCOUNTED FOR AT ALL?** In some cases, modelling the measurement error will add unneccesary complexity. An additional R-package will be created to aid in evaluating the severity of the measurement error using simulation tools.
- **CATEGORICAL COVARIATES WITH MEASUREMENT ERROR**: It is currently not easy to account for categorical measurement error in INLA.

REFERENCES

Blackwell, M., Honaker, J., and King, G., (2017). **A unified approach to measurement error and missing data: Overview and applications.** *Sociological Methods and Research,* 46(3):303-341.

Goldstein, H., Browne, W.J., and Charlton, C., (2018). **A Bayesian model for measurement and misclassification errors alongside missing data, with an application to higher education participation in Australia**. *Journal of Applied Statistics*, 45(5):918-931.

Rue, H., Martino, S., and Chopin, N. (2009). **Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Muff, S., Riebler, A., Held, L., Rue, H., and Saner, P. (2015). **Bayesian analysis of measurement error models using integrated nested Laplace approximations**. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(2):231-252.

NTNU