Multivariate Calibrations with Auxiliary Information

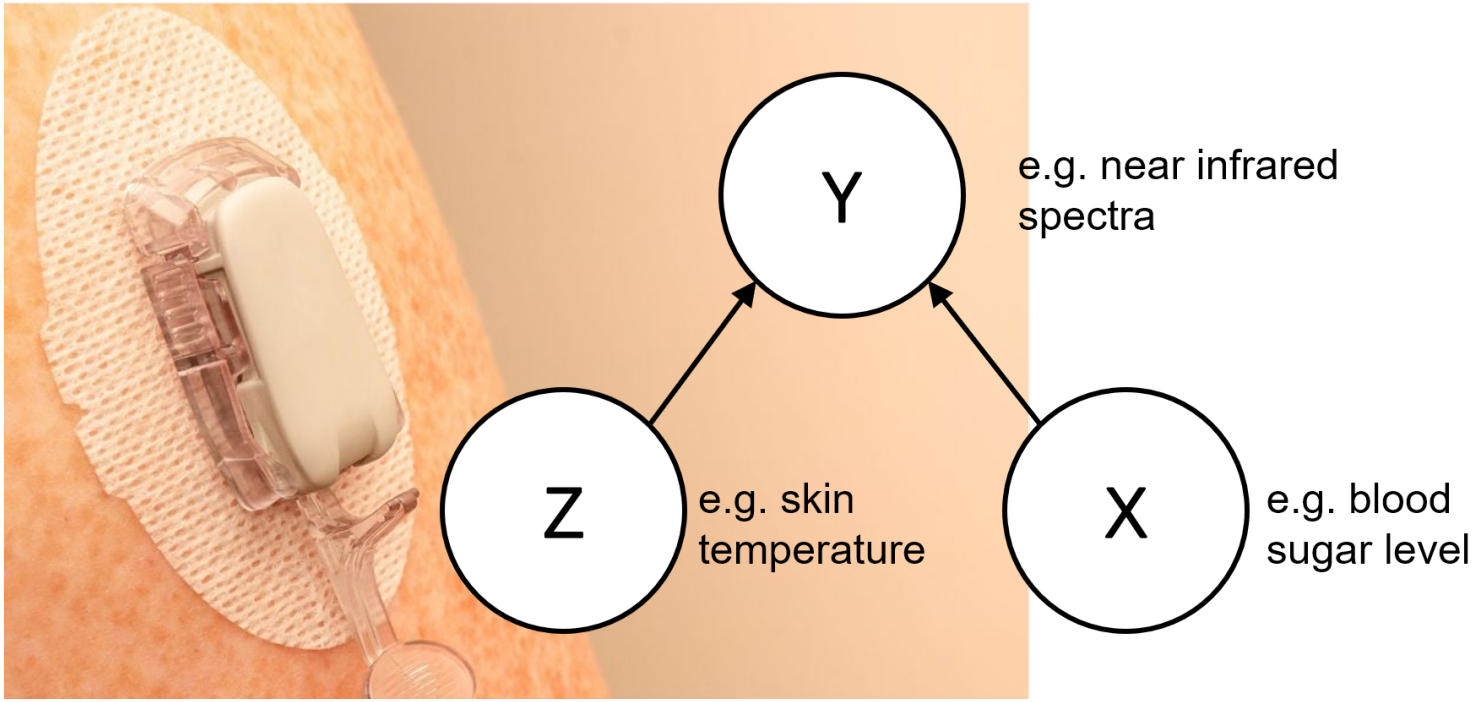
Tong Peifeng

Guanghua School of Management, Peking University

Supervised by Song Xi Chen and Cheng Yong Tang

2023/04/16

Introduction of inverse regression



Inverse prediction is a common situation in many areas such as economics, sociology and earth sciences. In such a prediction setting, we desire to use some cheap and quick but **error-prone** measurement Y, to predict the true amount of a constituent X, which in itself can be measured accurately in laboratory condition but **much greater effort or cost** (Brown, 2006).

Bias due to conditioning on a collider

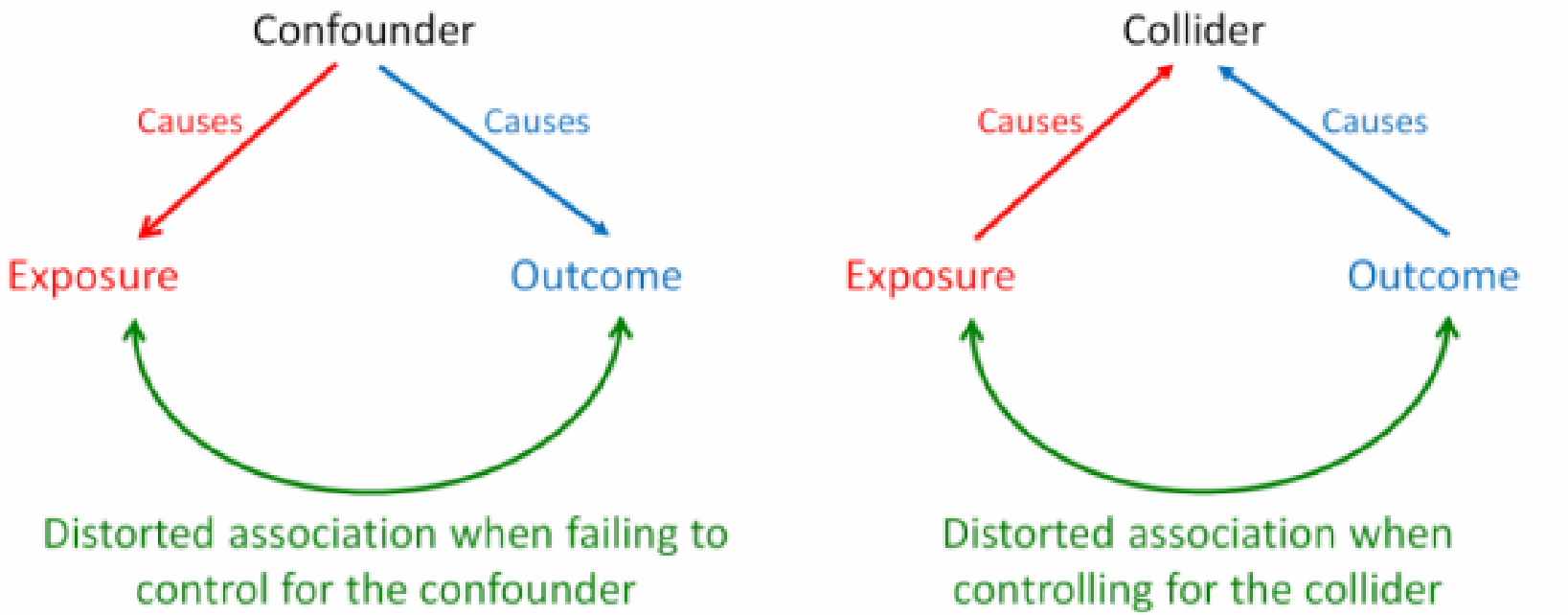


Figure 1: An illustration of the differences between confounder and collider.

The calibration model setting

Given a sample of n observations $\{(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\}_{i=1}^n$, where $\boldsymbol{y}_i \in \mathbb{R}^q$ is the multivariate response with covariates $\boldsymbol{x}_i \in \mathbb{R}^p$ and $\boldsymbol{z}_i \in \mathbb{R}^{p'}$. We consider the following two models:

Observation model:

$$\boldsymbol{y}_i = \boldsymbol{\alpha} + \boldsymbol{B}^T \boldsymbol{x}_i + \boldsymbol{D}^T \boldsymbol{z}_i + \boldsymbol{e}_i, \tag{1}$$

Calibration model:

$$\boldsymbol{y}_j' = \boldsymbol{\alpha} + \boldsymbol{B}^T \boldsymbol{\xi} + \boldsymbol{D}^T \boldsymbol{z}_j' + \boldsymbol{e}_j'. \tag{2}$$

In the calibration model (2), repeated measurements $\{\boldsymbol{y}_j'\}_{j=1}^l$ are available together with **only** $\{\boldsymbol{z}_j'\}_{j=1}^l$. The un-observable counterpart of $\boldsymbol{x}_j'$ is considered as shared by these l measurements, and is considered as a standalone parameter denoted by $\boldsymbol{\xi} \in \mathbb{R}^p$.

The calibration model setting

Collectively, we write the two models in their matrix forms as

$$\boldsymbol{Y} = \mathbf{1}_n \boldsymbol{\alpha}^T + \boldsymbol{X}\boldsymbol{B} + \boldsymbol{Z}\boldsymbol{D} + \boldsymbol{E}, \tag{3}$$

$$\boldsymbol{Y}' = \mathbf{1}_l \boldsymbol{\alpha}^T + \mathbf{1}_l \boldsymbol{\xi}^T \boldsymbol{B} + \boldsymbol{Z}'^T \boldsymbol{D} + \boldsymbol{E}', \tag{4}$$

where residuals $\boldsymbol{E}$ and $\boldsymbol{E}'$ are the exogenous residuals whose row vectors $\boldsymbol{e}_i^T$ are assumed to be independent and identically distributed satisfying

$$\mathbb{E}(\boldsymbol{e}_i) = 0, \mathbb{E}(\boldsymbol{e}_i \boldsymbol{e}_i^T) = \boldsymbol{\Gamma}. \tag{5}$$

We begin with this conventional cross-sectional setting and then extend the calibration problem to cover dependent $\boldsymbol{y}_j$s.

Two traditional ways in solving ξ

The so-called generalized least squares estimator (GLS) and the inverse regression (IR) estimator are generalized to our new setting with auxiliary variables $\boldsymbol{Z}$ in the calibration problem.

GLS:

$$\hat{\boldsymbol{\xi}}_{gls} = (\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}'), \tag{6}$$

where $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{B}}$, and $\hat{\boldsymbol{D}}$ are the OLS estimators of the observation model (3), $\bar{\boldsymbol{z}}' = (l^{-1}\mathbf{1}_l^T\boldsymbol{Z}')^T$, $\boldsymbol{S} = (\boldsymbol{Y}-\hat{\boldsymbol{Y}})^T(\boldsymbol{Y}-\hat{\boldsymbol{Y}})$, and $\hat{\boldsymbol{Y}} = \mathbf{1}_n\hat{\boldsymbol{\alpha}}^T + \boldsymbol{X}\hat{\boldsymbol{B}} + \boldsymbol{Z}\hat{\boldsymbol{D}}$ is the fitted/predicted responses of the model (3).

IR:

$$\hat{\boldsymbol{\xi}}_{ir} = \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\Phi}}^T\bar{\boldsymbol{y}}' + \hat{\boldsymbol{\Psi}}^T\bar{\boldsymbol{z}}', \tag{7}$$

where $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\Phi}}$ and $\hat{\boldsymbol{\Psi}}$ are the OLS estimators of a working model $\boldsymbol{X} = \mathbf{1}_n\boldsymbol{\theta}^T + \boldsymbol{Y}\boldsymbol{\Phi} + \boldsymbol{Z}\boldsymbol{\Psi} + \tilde{\boldsymbol{E}}$.

A Bayesian view for the IR estimator

We denote by $\mathbb{P}_A = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ and $\mathbb{M}_A = \boldsymbol{I} - \mathbb{P}_A$ be the projection matrices to its column space and the orthogonal complement, respectively.

Let $\boldsymbol{H}_{ir} = (\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}(\boldsymbol{X}^T\mathbb{M}_Z\boldsymbol{X})^{-1}$ and $\boldsymbol{\zeta} = \boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'$. Our main results on the properties of the GLS and inverse regression estimators, and a new **shrinkage estimator** are given in the following theorems.

Theorem

(i) Under some regularity conditions,

$$\hat{\boldsymbol{\xi}}_{ir} = (\boldsymbol{I} + \boldsymbol{H}_{ir})^{-1}(\hat{\boldsymbol{\xi}}_{gls} - \boldsymbol{\zeta}) + \boldsymbol{\zeta}. \tag{8}$$

(ii) Under the normality assumption and a proper prior distribution, $\hat{\boldsymbol{\xi}}_{ir}$ is the **Bayesian estimator**, *namely the posterior mean that minimizes the Bayesian risk.*

The prior distribution of $\boldsymbol{\xi}$ given $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{Z}'$ is the multivariate t-distribution $T_{\nu-p}\big(\boldsymbol{\zeta}, \frac{1}{\nu-p}(l^{-1} + n^{-1} + c_3 - \frac{1}{4}\boldsymbol{c}_2^T\boldsymbol{C}_1^{-1}\boldsymbol{c}_2)\boldsymbol{C}_1^{-1}\big)$, where $\nu = n - p - p' - q$, $\boldsymbol{C}_1 = (\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1}$, $\boldsymbol{c}_2 = ((\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\mathbb{M}_x\boldsymbol{Z})^{-1})\bar{\boldsymbol{z}}'$ and $c_3 = \bar{\boldsymbol{z}}'^T(\boldsymbol{Z}^T\mathbb{M}_x\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'$.

A better shrinkage estimator

We define the limiting predictive mean square error of the $\hat{\boldsymbol{\xi}}$ as

$$G(\boldsymbol{\xi},\boldsymbol{\zeta}) = \|\{(\boldsymbol{I}+\boldsymbol{H})^{-1}-\boldsymbol{I}\}(\boldsymbol{\xi}-\boldsymbol{\zeta})\|_2^2 + l^{-1}\mathrm{tr}\{(\boldsymbol{I}+\boldsymbol{H})^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}(\boldsymbol{I}+\boldsymbol{H}^T)^{-1}\}.$$

Upon treating the design as random, we define the averaged mean squared error as $\mathrm{AMSE} = \mathbb{E}\{G(\boldsymbol{\xi},\boldsymbol{\zeta})\}$.
Then, for the class of estimators $\hat{\boldsymbol{\xi}} = (\boldsymbol{I}+\boldsymbol{H})^{-1}(\hat{\boldsymbol{\xi}}_{gls}-\boldsymbol{\zeta})+\boldsymbol{\zeta}$ with a positive definite $\boldsymbol{H} \in \mathbb{R}^{p\times p}$, we propose to find the **optimal shrinkage estimator** that minimizes the AMSE; the result is given in the next theorem.

Theorem

If $\boldsymbol{\xi}$ and $\{\boldsymbol{x}_i\}_{i=1}^n$ are independent and identically distributed (IID) from a distribution having finite $(2+\delta)$-th moment for a $\delta > 0$, let the covariance matrices of $\boldsymbol{X}$ and $\boldsymbol{Z}$ be $\boldsymbol{\Gamma}_x$, $\boldsymbol{\Gamma}_z$ and $\boldsymbol{\Gamma}_{xz}$, the AMSE is minimized at

$$\tilde{\boldsymbol{H}}_{opt} = l^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}(\boldsymbol{\Gamma}_x - \boldsymbol{\Gamma}_{xz}\boldsymbol{\Gamma}_z^{-1}\boldsymbol{\Gamma}_{xz}^T)^{-1}.$$

Furthermore, as $n \to \infty$,

$$\boldsymbol{H}_{opt} = l^{-1}(\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}(\boldsymbol{X}^T\mathbb{M}_Z\boldsymbol{X})^{-1}$$

is consistent to $\tilde{\boldsymbol{H}}_{opt}$.

The relationship between the proposed estimators

Here $\tilde{\boldsymbol{H}}_{opt}$ in Theorem 2 is an unknown hyper-parameter; and we propose to implement its empirical counterpart $\boldsymbol{H}_{opt}$ for empirically obtaining the optimal shrinkage estimator:

$$\hat{\boldsymbol{\xi}}_{opt} = (\boldsymbol{I} + \boldsymbol{H}_{opt})^{-1}(\hat{\boldsymbol{\xi}}_{gls} - \boldsymbol{\zeta}) + \boldsymbol{\zeta}. \tag{9}$$

In summary, the relationship between these estimators is depicted in Figure 2, which shows the spectrum of trade-offs between bias and variance in the class of shrinkage estimators.

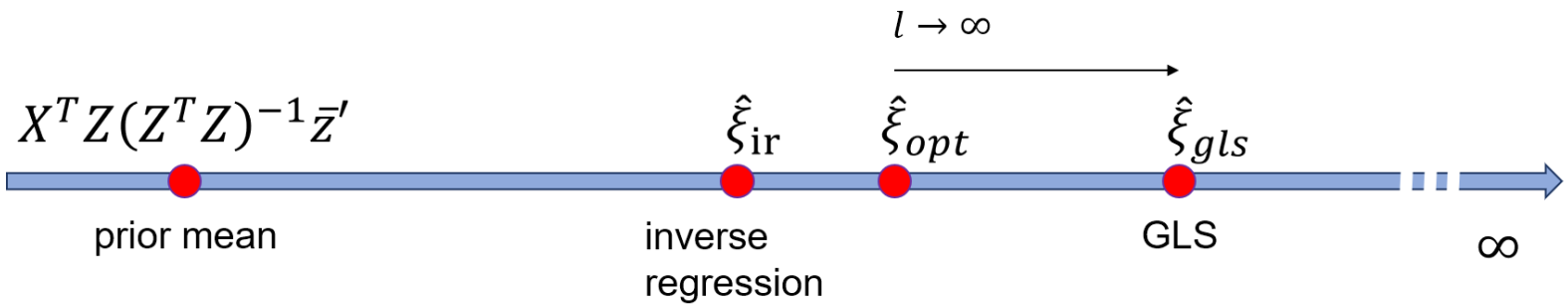


Figure 2: The relationship among GLS, inverse regression and optimal shrinkage estimators. When $l = 1$, the optimal shrinkage estimator coincides with the inverse regression estimator. The optimal shrinkage estimator converges to the GLS estimator as $l \to \infty$.

An objective function interpretation

Remarkably, as shown in the next theorem, the optimal shrinkage estimator, the GLS and inverse regression estimators link to a class of the **penalized estimations**:

$$\min_{\boldsymbol{\xi}} \operatorname{tr}[\{\boldsymbol{Y}' - \hat{\boldsymbol{Y}}'(\boldsymbol{\xi})\}\boldsymbol{S}^{-1}\{\boldsymbol{Y}' - \hat{\boldsymbol{Y}}'(\boldsymbol{\xi})\}^T] + \lambda\sigma^2(\boldsymbol{\xi}), \tag{10}$$

where $\hat{\boldsymbol{Y}}'(\boldsymbol{\xi}) = \mathbf{1}_l\hat{\boldsymbol{\alpha}}^T + \mathbf{1}_l\boldsymbol{\xi}^T\hat{\boldsymbol{B}} + \boldsymbol{Z}'\hat{\boldsymbol{D}}$, $\sigma^2(\boldsymbol{\xi}) = 1/l + 1/n + \boldsymbol{\xi}^T\boldsymbol{C}_1\boldsymbol{\xi} - \boldsymbol{c}_2^T\boldsymbol{\xi} + c_3$, with $\boldsymbol{C}_1$, $\boldsymbol{c}_2$ and c_3 defined in Assumption 2 (b).

Theorem

Under some regularity conditions,

1. $\hat{\boldsymbol{\xi}}_{ir}$ *is the solution for* $\lambda = l$;
2. $\hat{\boldsymbol{\xi}}_{opt}$ *is the solution for* $\lambda = 1$;
3. $\hat{\boldsymbol{\xi}}_{gls}$ *is the solution for* $\lambda = 0$.

Calibration with random effect

So far, our method treats the n measurements in $\boldsymbol{Y}$ of model (3) as independent. For incorporating the **correlated scenarios**, we extend the framework to accommodate the mixed or random effects in handling broader situations. Specifically, we consider an extended observation model

$$\boldsymbol{Y} = \mathbf{1}_n \boldsymbol{\alpha}^T + \boldsymbol{XB} + \boldsymbol{ZD} + \boldsymbol{\Psi A} + \boldsymbol{E}, \tag{11}$$

where $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)^T \in \mathbb{R}^{k\times q}$, $\boldsymbol{a}_j \in \mathbb{R}^q$ $(j = 1, \ldots, k)$ are k random effects, and $\boldsymbol{\Psi} \in \mathbb{R}^{n\times k}$ is a **known model matrix**. We assume $\boldsymbol{a}_k \sim \mathcal{N}(0, \boldsymbol{\Gamma}_1)$ and $\boldsymbol{e}_i \sim \mathcal{N}(0, \boldsymbol{\Gamma}_2)$. We develop an expectation-maximization (EM) algorithm that is computationally highly efficient for the estimation of $\boldsymbol{\alpha}$, $\boldsymbol{B}$, $\boldsymbol{D}$, $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$. The vector form of (11) is

$$\mathrm{vec}(\boldsymbol{Y}^T) = \mathrm{vec}(\boldsymbol{\alpha}\mathbf{1}_n^T) + \mathrm{vec}(\boldsymbol{B}^T\boldsymbol{X}^T) + \mathrm{vec}(\boldsymbol{D}^T\boldsymbol{Z}^T) + (\boldsymbol{\Psi}\otimes\boldsymbol{I})\mathrm{vec}(\boldsymbol{A}^T) + \mathrm{vec}(\boldsymbol{E}^T). \tag{12}$$

Model (11) implies that

$$\mathrm{var}\{\mathrm{vec}(\boldsymbol{Y}^T)\} = \boldsymbol{\Psi}\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1 + \boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2.$$

Calibration with random effect

Corresponding to (2), the calibration model with random effect is

$$\boldsymbol{y}_j' = \boldsymbol{\alpha} + \boldsymbol{B}^T\boldsymbol{\xi} + \boldsymbol{D}^T\boldsymbol{z}_j' + \boldsymbol{A}^T\boldsymbol{\psi}' + \boldsymbol{e}_j' \quad \text{for } j = 1, 2, \ldots l, \tag{13}$$

where the extra $\boldsymbol{A}^T\boldsymbol{\psi}'$ reflects the reality that some of the random effects may be shared as reflected by the design $\boldsymbol{\psi}'$.
Let $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{D}}, \hat{\boldsymbol{\Gamma}}_1$, and $\hat{\boldsymbol{\Gamma}}_2$ be the estimators using the EM algorithm. Then, the **GLS estimator** incorporating the random effects is

$$\hat{\boldsymbol{\xi}}_{gls} = (\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}' - \hat{\boldsymbol{A}}^T\boldsymbol{\psi}'), \tag{14}$$

where $\hat{\boldsymbol{A}} = \mathbb{E}(\boldsymbol{A}^T|\boldsymbol{Y}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{D}}, \hat{\boldsymbol{\Gamma}}_1, \hat{\boldsymbol{\Gamma}}_2)$ and

$$\boldsymbol{S} = l^{-1}\hat{\boldsymbol{\Gamma}}_2 + \boldsymbol{\psi}'^T\boldsymbol{\psi}' \otimes \hat{\boldsymbol{\Gamma}}_1 - (\boldsymbol{\psi}'^T\boldsymbol{\Psi}^T \otimes \hat{\boldsymbol{\Gamma}}_1)(\boldsymbol{\Psi}\boldsymbol{\Psi}^T \otimes \hat{\boldsymbol{\Gamma}}_1 + \boldsymbol{I}_n \otimes \hat{\boldsymbol{\Gamma}}_2)^{-1}(\boldsymbol{\Psi}\boldsymbol{\psi}' \otimes \hat{\boldsymbol{\Gamma}}_1)$$

is the conditional variance of $\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}' - \hat{\boldsymbol{A}}\boldsymbol{\psi}'$ given $(\boldsymbol{Y}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{D}}, \hat{\boldsymbol{\Gamma}}_1, \hat{\boldsymbol{\Gamma}}_2)$.
Inspired by (9), we can develop an **optimal shrinkage estimator** as

$$\hat{\boldsymbol{\xi}}_{opt} = (\boldsymbol{I} + \boldsymbol{H})^{-1}\hat{\boldsymbol{\xi}}_{gls}, \tag{15}$$

where $\boldsymbol{H} = (\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}(\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

A baseline setting

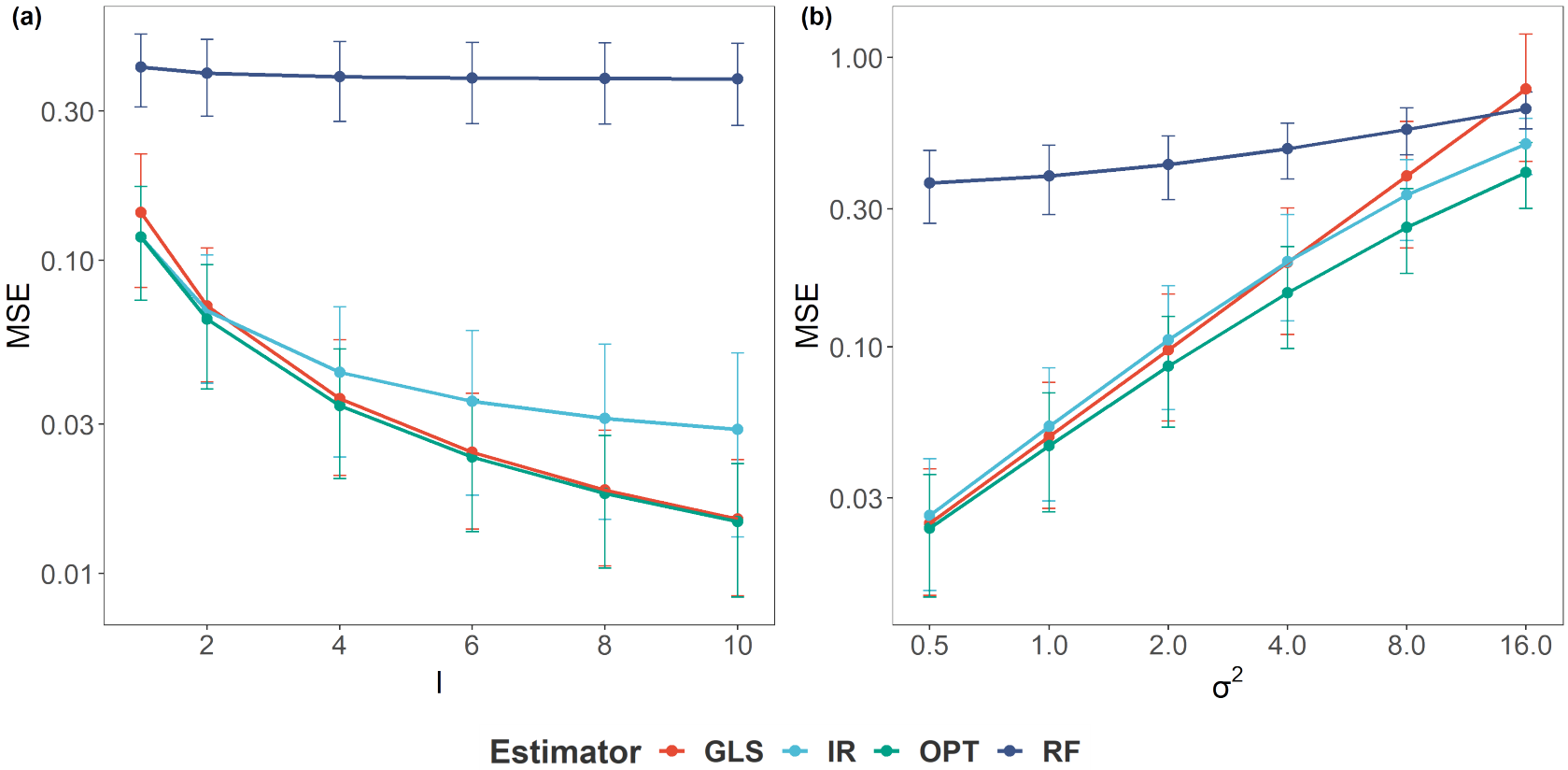


Figure 3: Empirical means and the 10th, 90th quantiles of the MSEs (in log-scale) calculated from 1000 simulation replications for the GLS, the inverse regression (IR), the optimal shrinkage (OPT) and the inverse regression with the random forest (RF) estimators with respect to numbers of repeated measurements l (a) and the variances of the noises σ^2 (b).

Sensitivity analysis

Table 1: Empirical means and standard deviations of the MSEs from 1000 simulations for the GLS, the inverse regression (IR), the optimal shrinkage (OPT) and the inverse regression with the random forest (RF) estimators with respect to (a) the number of dependence variables q^* used for training, (b) the number of covariates p'^* used for training,

(a) with respect to number of dependence variables

q^*	$0.2q$	$0.4q$	$0.6q$	$0.8q$	q
GLS	3922 ± 60948	0.298 ± 0.568	0.117 ± 0.207	0.069 ± 0.045	0.049 ± 0.022
IR	0.438 ± 0.166	0.199 ± 0.112	0.112 ± 0.066	0.074 ± 0.040	0.053 ± 0.024
OPT	0.385 ± 0.165	0.159 ± 0.094	0.090 ± 0.052	0.061 ± 0.030	0.045 ± 0.018
RF	0.746 ± 0.120	0.599 ± 0.114	0.489 ± 0.103	0.432 ± 0.094	0.391 ± 0.085

(b) with respect to number of covariates

p'^*	$0.2p'$	$0.4p'$	$0.6p'$	$0.8p'$	p'
GLS	0.315 ± 0.255	0.187 ± 0.171	0.110 ± 0.086	0.070 ± 0.044	0.049 ± 0.024
IR	0.211 ± 0.085	0.149 ± 0.069	0.103 ± 0.052	0.072 ± 0.036	0.054 ± 0.025
OPT	0.214 ± 0.097	0.141 ± 0.072	0.092 ± 0.049	0.063 ± 0.030	0.046 ± 0.019
RF	0.413 ± 0.092	0.410 ± 0.089	0.389 ± 0.087	0.388 ± 0.084	0.389 ± 0.083

Sensitivity analysis

Table 2: Empirical means and standard deviations of the MSEs from 1000 simulations for the GLS, the inverse regression (IR), the optimal shrinkage (OPT) and the inverse regression with the random forest (RF) estimators with respect to (c) the sample size N and (d) the explanation power of $\boldsymbol{X}$ as represented by shrinking $\boldsymbol{B}^*$ to $\gamma\boldsymbol{B}^*$.

(c) with respect to sample size

N	30	100	300	1000	2500
GLS	0.461 ± 0.314	0.079 ± 0.038	0.056 ± 0.026	0.050 ± 0.024	0.049 ± 0.023
IR	0.418 ± 0.245	0.079 ± 0.033	0.059 ± 0.026	0.054 ± 0.025	0.053 ± 0.024
OPT	0.441 ± 0.279	0.074 ± 0.030	0.052 ± 0.021	0.047 ± 0.019	0.045 ± 0.018
RF	0.843 ± 0.163	0.666 ± 0.099	0.545 ± 0.086	0.447 ± 0.083	0.390 ± 0.081

(d) with respect to explaining power of X

γ	0.1	0.2	0.33	0.5	1
GLS	4.471 ± 1.832	1.157 ± 0.471	0.420 ± 0.171	0.187 ± 0.076	0.047 ± 0.019
IR	0.863 ± 0.040	0.608 ± 0.076	0.356 ± 0.080	0.193 ± 0.061	0.051 ± 0.021
OPT	0.792 ± 0.054	0.498 ± 0.078	0.276 ± 0.067	0.150 ± 0.046	0.044 ± 0.016
RF	0.999 ± 0.022	0.934 ± 0.034	0.819 ± 0.058	0.676 ± 0.077	0.383 ± 0.082

Spatial-temporal settings

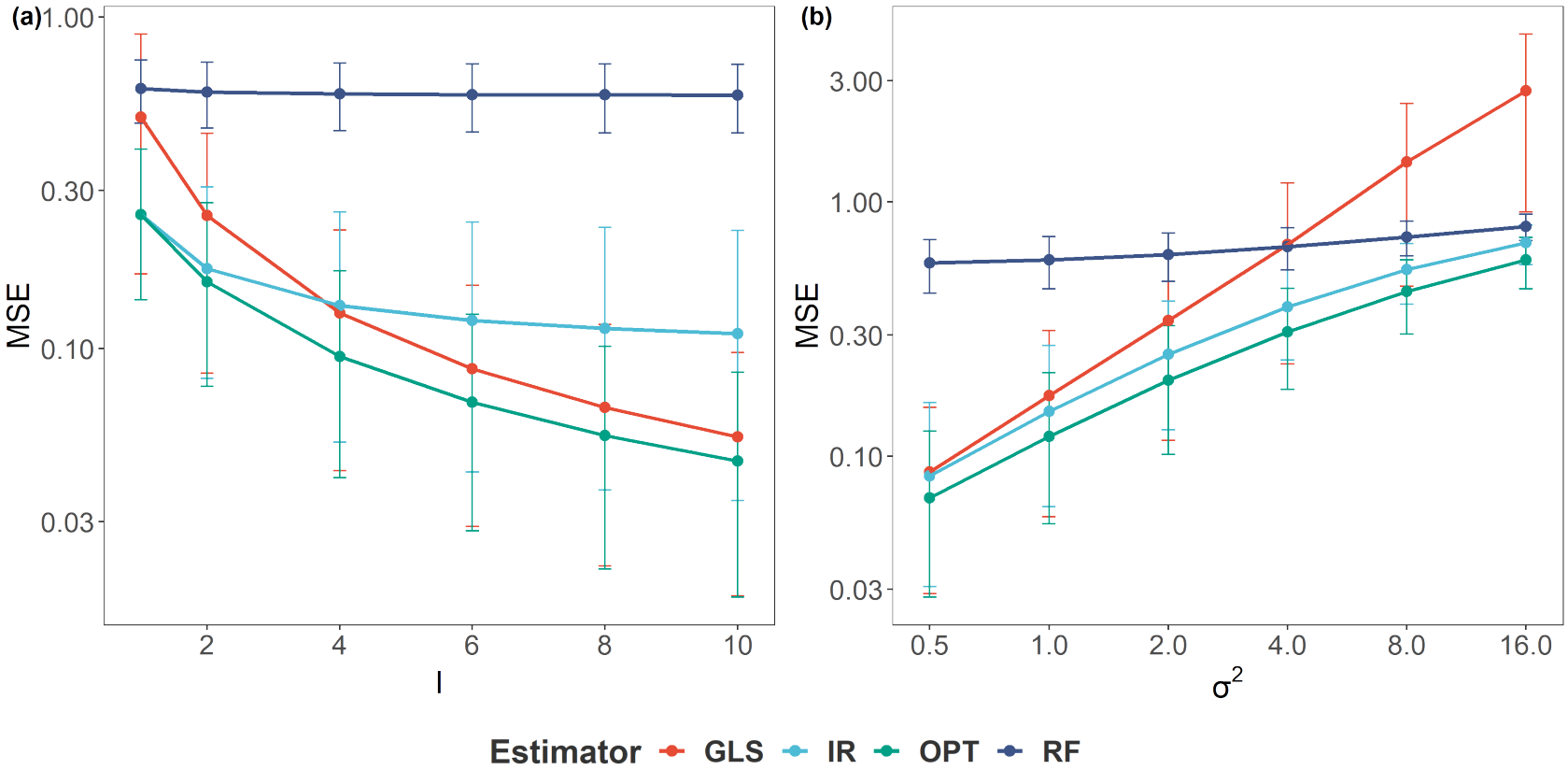


Figure 4: Empirical means and the 10th, 90th quantiles of the MSEs (in the log-scale), in the presence of the random effects, calculated from 1000 simulation replications for the GLS, the inverse regression (IR), the optimal shrinkage (OPT) and the inverse regression with the random forest (RF) estimators with respect to the numbers of repeated measurements l (a) and different scales of noises σ^2.

The aqueous glucose data

Diabetes Mellitus (DM) is one of the common chronic diseases throughout the globe. A continuous blood glucose monitoring is important for DM patients to understand their disease progression and adjust their lifestyle management. However, such blood glucose measurements are often invasive and expensive, thus not proper for the regular use. Nowadays, near infrared spectroscopy (NIR) is a promising technique for continuous blood glucose monitoring, since it is cheap, non-invasive and easy to deploy on some wearable devices.

The aqueous glucose data

In this case study, we applied our calibration method in a open sourced NIR spectroscopy data of aqueous glucose (Fuglerud, 2021). This dataset contained 127 samples measured under laboratory conditions. Five covariates (lactates, ethanol, caffeine, acetaminophen and temperature) were introduced to the dextrose water, alone with 4200 NIR wavelength channels ranged from 400nm-2500nm. The NIR spectral were collected in triplicate in each setting, sampled every 0.5nm frequency resolution with a bandwidth of 8.75nm.

According to Fuglerud et al. (2021), we removed those NIR bands which were highly absorbing water peaks (1900-2100nm, 2300-2500nm), the overlap of the two spectrometer detectors (1090-1110nm), and the fringes of the detectors ($<$500 and $>$2300 nm). After that, we selected the remaining NIR bands every 25nm, made the dimension of $\boldsymbol{Y}$ to be 127×60. All the data were pretreated to have mean 0 and variance 1.

The aqueous glucose data

Table 3: MSE and Bias for glucose prediction under different calibration methods

	MSE		Bias	
	l=1	l=3	l=1	l=3
GLS	0.0912	0.0683	0.2038	0.1563
inverse regression	0.0891	0.0714	0.2008	0.1604
optimal shrinkage	0.0891	0.0676	0.2008	0.1555
RF	0.4411	0.4690	0.5511	0.5668

Brown, P. J. (2006). Inverse prediction. *Encyclopedia of Environmetrics*.

Fuglerud, S. S. (2021). Aqueous glucose measured by NIR spectroscopy.

Fuglerud, S. S., Ellingsen, R., Aksnes, A., and Hjelme, D. R. (2021). Investigation of the effect of clinically relevant interferents on glucose monitoring using near-infrared spectroscopy. *Journal of Biophotonics*, 14(5).

Thanks!