

Article

# Univariate Linear Normal Models: Optimal Equivariant Estimation

Gloria García <sup>1</sup>, Marta Cubedo <sup>2,\*</sup>  and Josep M. Oller <sup>2</sup> <sup>1</sup> Department of Education and Professional Training, Gencat, 08021 Barcelona, Spain; ggarcia76@xtec.cat<sup>2</sup> Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Avda. Diagonal 643, 08028 Barcelona, Spain; joller@ub.edu

\* Correspondence: mcubedo@ub.edu

## Abstract

In this paper, we establish the existence and uniqueness of the minimum intrinsic risk equivariant (MIRE) estimator for univariate linear normal models. The estimator is derived under the action of the subgroup of the affine group that preserves the column space of the design matrix, within the framework of intrinsic statistical analysis based on the squared Rao distance as the loss function. This approach provides a parametrization-free assessment of risk and bias, differing substantially from the classical quadratic loss, particularly in small-sample settings. The MIRE is compared with the maximum likelihood estimator (MLE) in terms of intrinsic risk and bias, and a simple approximate version (*a*-MIRE) is also proposed. Numerical evaluations show that the *a*-MIRE performs closely to the MIRE while significantly reducing the intrinsic bias and risk of the MLE for small samples. The proposed intrinsic methods could extend to other invariant frameworks and connect with recent developments in robust estimation procedures.

**Keywords:** equivariant estimator; information metric; riemannian risk; intrinsic bias**MSC:** 62F99; 60K35; 62A99

Academic Editor: Heng Lian

Received: 2 October 2025

Revised: 30 October 2025

Accepted: 11 November 2025

Published: 14 November 2025

**Citation:** García, G.; Cubedo, M.; Oller, J.M. Univariate Linear Normal Models: Optimal Equivariant Estimation. *Mathematics* **2025**, *13*, 3659. <https://doi.org/10.3390/math13223659>

**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Methodological developments concerning the point estimation of parameters in statistical models remain a subject of significant scholarly interest. This is particularly true in settings that demand robust estimation procedures or involve complex multiparameter structures, where conventional or historically established methods may be unsuitable or inefficient. Notable contributions in this field include those by [1–3], among others.

A classical topic in statistical inference is the study of parametric statistical models that remain invariant under the action of a transformation group  $\mathcal{G}$  acting on the sample space  $\chi$ . This action induces a corresponding group  $\bar{\mathcal{G}}$ , acting on the parameter space  $\Theta$ . In such a setting, it is logically necessary to restrict attention to equivariant estimators, that is, estimators  $U$  of  $\theta \in \Theta$ , satisfying

$$U(gX) = \bar{g}U(X), \quad g \in \mathcal{G}, \bar{g} \in \bar{\mathcal{G}}$$

here,  $X$  is a random map with domain  $\chi$ .

Only equivariant estimators yield coherent and consistent inference, free of paradoxes arising from reparametrization or alternative representations of the data. Moreover, if the loss function  $L$  is itself invariant under the group action,

$$L(U(X), \theta) = L(U(gX), \hat{g}\theta)$$

then the associated risk is constant along orbits of  $\bar{\mathcal{G}}$ . In particular, if the induced action is transitive on  $\Theta$ , the risk is constant throughout the entire parameter space, and a minimum-risk equivariant estimator is optimal in the strongest possible sense.

Furthermore, the Fisher information matrix of a regular statistical model endows the parameter space with a natural Riemannian structure [4–6]. This geometric viewpoint has led to intrinsic approaches to estimation [7–9] and has also revealed connections with physical laws in certain contexts [10]. Within this intrinsic framework, non-invariant criteria such as squared error loss are replaced with intrinsic loss functions, most notably the squared Rao distance, which is directly induced by the Fisher–Rao metric.

As a motivating example, consider the multivariate normal family  $Y \sim N_n(\mu, \Sigma)$ , which is invariant under the affine group. If  $Y \mapsto Z = PY + C$  for a non-singular matrix  $P$  and vector  $C$ , then  $Z \sim N_n(P\mu + C, P\Sigma P^T)$ , and the induced transformation in parameter space is  $(\mu, \Sigma) \mapsto (P\mu + C, P\Sigma P^T)$ .

In this work, we focus on the univariate linear normal model

$$Y \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$$

where  $X$  is a fixed  $n \times m$  design matrix of full rank, and  $(\boldsymbol{\beta}, \sigma) \in \Theta = \mathbb{R}^m \times \mathbb{R}^+$ . This family is invariant under a particular subgroup of the affine group, acting on  $\mathbb{R}^n$ , making it suitable for the application of equivariant decision theory.

We first characterize the class of equivariant estimators under this group action. Then, adopting the squared Rao distance as an intrinsic loss function, we establish the existence and uniqueness of the minimum-risk equivariant estimator and derive a closed-form expression for its intrinsic bias. We further compare its intrinsic risk and bias with those of the maximum likelihood estimator and introduce a simple approximation (*a*-MIRE) that retains the essential advantages of the optimal equivariant estimator while being easier to compute.

Although our perspective is classical, intrinsic Bayesian approaches based on noninformative priors are also possible [9,11,12] and represent a promising direction for future research. For broader background on statistical estimation and its intrinsic developments, see [13–16].

## 2. Equivariant Estimators for Linear Models

Let us consider the univariate linear normal model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{1}$$

where  $\mathbf{y}$  is an  $n \times 1$  random vector,  $\mathbf{X}$  is an  $n \times m$  matrix of known constants with  $0 < \text{rank}(\mathbf{X}) = m < n$ ,  $\boldsymbol{\beta}$  is an  $m \times 1$  vector of unknown parameters to be estimated, and  $\mathbf{e}$  is the fluctuation or error of  $\mathbf{y}$  about  $\mathbf{X}\boldsymbol{\beta}$ . We assume that the errors are unbiased, independent, with the same variance  $\sigma^2$ , and following an  $n$ -variate normal distribution, which is  $\mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ , where  $\mathbf{I}_n$  is the  $n \times n$  identity matrix. Therefore,  $\mathbf{y}$  follows a distribution belonging to the parametric family  $\mathcal{P} = \{N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \mid (\boldsymbol{\beta}, \sigma) \in \Theta\}$ , where the parameter space is  $\Theta = \mathbb{R}^m \times \mathbb{R}^+$ , a  $(m + 1)$ -dimensional simply connected real manifold. Hereafter, we identify elements of  $\mathbb{R}^m$  with  $m \times 1$  column vectors whenever necessary.

Denote by

$$O_E(n) = \{\mathbf{H} \in O(n) \mid \mathbf{y} \in E \Rightarrow \mathbf{H}\mathbf{y} \in E\}$$

where  $E$  is a subspace of  $\mathbb{R}^n$  and  $O(n)$  is the group of  $n \times n$  orthogonal matrices with entries from  $\mathbb{R}$ . Define  $F$  as the subspace of  $\mathbb{R}^n$  spanned by the columns of  $X$ , that is  $F = \text{Col}(X)$ . Observe that  $O_F(n) = O_{F^\perp}(n)$ ,  $I_n \in O_F(n)$ ,  $\text{Col}(X) = \text{Col}(HX) \quad \forall H \in O_F(n)$ , and if  $H \in O_F(n)$ , then  $H^t \in O_F(n)$ . Moreover, every  $H \in O_F(n)$  induces, in  $F$  and  $F^\perp$ , two isomorphisms preserving the Euclidean norm in each subspace.

$\mathcal{P}$  is invariant under the action of the subgroup of the affine group in  $\mathbb{R}^n$  given by the family of transformations

$$g_{(a,H,c)}(\mathbf{y}) = aH\mathbf{y} + \mathbf{c}, \mathbf{y} \in \mathbb{R}^n \tag{2}$$

where  $a > 0, H \in O_F(n), \mathbf{c} \in F$ .

Here,  $H$  belongs to the subgroup  $O_F(n) \subset O(n)$ , consisting of all orthogonal transformations that preserve the column space  $F = \text{Col}(X)$  and its orthogonal complement. In other words,  $\mathbb{R}^n$  decomposes as  $F \oplus F^\perp$ , and any  $H \in O_F(n)$  acts as an isometry on each of these subspaces,  $F$  and  $F^\perp$ , possibly including rotations or reflections, but it cannot mix these subspaces or alter lengths.

Observe that  $\mathcal{G} = \{g_{(a,H,c)} \mid a > 0, H \in O_F(n), \mathbf{c} \in F\}$  induces an action on the parameter space  $\theta$  given by

$$\overline{g_{(a,H,c)}}(\boldsymbol{\beta}, \sigma) = (a(X^tX)^{-1}X^tHX\boldsymbol{\beta} + (X^tX)^{-1}X^t\mathbf{c}, a\sigma) \tag{3}$$

where this result is obtained, taking into account that  $X(X^tX)^{-1}X^t$  is the projection matrix into  $F$ , and thus, if  $\mathbf{w} \in F$ , there exists a unique  $\boldsymbol{\eta}$  such that  $\mathbf{w} = X\boldsymbol{\eta}$  and  $X(X^tX)^{-1}X^t\mathbf{w} = \mathbf{w}$ .

Since the family  $\mathcal{P}$  is invariant under the action of  $\mathcal{G}$ , it is natural to restrict our attention to the class of equivariant estimators  $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$  of  $(\boldsymbol{\beta}, \sigma)$ , i.e., an estimator satisfying  $\mathbf{U}(g_{(a,H,c)}(\mathbf{y})) = \overline{g_{(a,H,c)}}(\mathbf{U}(\mathbf{y}))$  for all  $g_{(a,H,c)} \in \mathcal{G}$ .

**Proposition 1.** *Let  $\mathbf{U}$  be an equivariant estimator of  $(\boldsymbol{\beta}, \sigma)$ . Then  $\mathbf{U}$  belongs to the family  $\{\mathbf{U}_\lambda, \lambda \in \mathbb{R}_+\}$ , where*

$$\mathbf{U}_\lambda(\mathbf{y}) = \left( (X^tX)^{-1}X^t\mathbf{y}, \lambda \left\{ \mathbf{y}^t \left( I_n - X(X^tX)^{-1}X^t \right) \mathbf{y} \right\}^{1/2} \right)$$

**Proof.** Let  $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$ . The equivariance condition for  $\mathbf{U}$  involves

$$\begin{aligned} \mathbf{U}_1(aH\mathbf{y} + \mathbf{c}) &= a(X^tX)^{-1}X^tHX\mathbf{U}_1(\mathbf{y}) + (X^tX)^{-1}X^t\mathbf{c} \\ \mathbf{U}_2(aH\mathbf{y} + \mathbf{c}) &= a\mathbf{U}_2(\mathbf{y}) \end{aligned}$$

for any  $a > 0, H \in O_F(n), \mathbf{c} \in F$ .

Any  $\mathbf{y} \in \mathbb{R}^n$  can be written in a unique form as  $\mathbf{y} = \mathbf{y}_F + \mathbf{y}_{F^\perp}$ , where  $\mathbf{y}_F \in F$  and  $\mathbf{y}_{F^\perp} \in F^\perp$ . Specifically,

$$\mathbf{y}_F = X(X^tX)^{-1}X^t\mathbf{y} \quad \text{and} \quad \mathbf{y}_{F^\perp} = (I_n - X(X^tX)^{-1}X^t)\mathbf{y} \tag{4}$$

If we choose  $\mathbf{c} = -aH\mathbf{y}_F$  in the previous expressions, we obtain

$$\mathbf{U}_1(aH\mathbf{y}_{F^\perp}) = a(X^tX)^{-1}X^tHX\mathbf{U}_1(\mathbf{y}) - a(X^tX)^{-1}X^tH\mathbf{y}_F \tag{5}$$

$$\mathbf{U}_2(aH\mathbf{y}_{F^\perp}) = a\mathbf{U}_2(\mathbf{y}) \tag{6}$$

First, we focus on  $\mathbf{U}_1$ . If we let  $H^* = (I_n - 2X(X^tX)^{-1}X^t)$ , we have  $H^* \in O_F(n)$  with

$$H^*\mathbf{v} = -\mathbf{v} \quad \forall \mathbf{v} \in F \quad \text{and} \quad H^*\mathbf{v} = \mathbf{v} \quad \forall \mathbf{v} \in F^\perp$$

Now observe that (5) is satisfied for any  $a > 0$  and  $\mathbf{H}$  in  $O_F(n)$ , in particular for  $\mathbf{I}_n$  and  $\mathbf{H}^*$ . Therefore,

$$\begin{aligned} \mathbf{U}_1(a\mathbf{y}_{F^\perp}) &= a\mathbf{U}_1(\mathbf{y}) - a(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}_F \\ \mathbf{U}_1(a\mathbf{H}^*\mathbf{y}_{F^\perp}) &= -a\mathbf{U}_1(\mathbf{y}) + a(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}_F \end{aligned}$$

But  $a\mathbf{H}^*\mathbf{y}_{F^\perp} = a\mathbf{y}_{F^\perp}$ , which leads to

$$0 = \mathbf{U}_1(\mathbf{y}) - (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}_F$$

From (4),

$$\mathbf{U}_1(\mathbf{y}) = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}, \quad \mathbf{y} \in \mathbb{R}^n$$

Next, we consider  $U_2$ . Let  $a = 1$  and  $\mathbf{H} = \mathbf{I}_n$  in (6); it follows

$$U_2(\mathbf{y}_{F^\perp}) = U_2(\mathbf{y}) \tag{7}$$

Accordingly, it is enough to determine  $U_2$  on  $F^\perp$ . Since  $F^\perp$  is at least one-dimensional (recall that  $m < n$ ), we can take a unit vector  $\mathbf{z} \in F^\perp$  such that  $U_2(\mathbf{z}) = \lambda > 0$ . Then, by (6), we have

$$U_2(\mathbf{H}\mathbf{z}) = U_2(\mathbf{z})$$

for any  $\mathbf{H} \in O_F(n)$ . Observe that any arbitrary unit vector in  $F^\perp$  can be written as  $\mathbf{H}\mathbf{z}$  for a proper  $\mathbf{H} \in O_F(n)$ . Therefore, for any  $\mathbf{y}_{F^\perp} \in F^\perp, \mathbf{y}_{F^\perp} \neq \mathbf{0}$ , with probability one, we have

$$U_2(\mathbf{y}_{F^\perp}) = \|\mathbf{y}_{F^\perp}\| U_2\left(\frac{\mathbf{y}_{F^\perp}}{\|\mathbf{y}_{F^\perp}\|}\right) = \|\mathbf{y}_{F^\perp}\| U_2(\mathbf{H}\mathbf{z}) = \|\mathbf{y}_{F^\perp}\| U_2(\mathbf{z}) = \lambda \|\mathbf{y}_{F^\perp}\|$$

Finally, from (7) and (4), we have

$$U_2(\mathbf{y}) = \lambda \left\{ \mathbf{y}^t \left( \mathbf{I}_n - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \right) \mathbf{y} \right\}^{1/2}$$

□

Observe that the standard maximum likelihood estimator, MLE, for the present model, is an equivariant estimator, with  $\lambda = \frac{1}{\sqrt{n}}$ .

### 3. Minimum Riemannian Risk Estimators

In the framework of intrinsic analysis, where the loss function is the square of the Rao distance, the Riemannian distance is induced by the information metric in the parameter space  $\Theta$ , and once the class of the equivariant estimators has been determined, a natural question arises: which is the equivariant estimator that minimizes the risk?

First of all, we summarize the basic geometric results corresponding to the model (1) which are going to be used hereafter. We use a standardized version of the information metric, given by the usual information metric corresponding to this linear model divided by a constant factor  $n$ , i.e., the number of rows of matrix  $\mathbf{X}$ . This metric is given by the solution of the geodesic equations

$$ds^2 = \frac{1}{n\sigma^2} \left( d\boldsymbol{\beta}^t \mathbf{X}^t \mathbf{X} d\boldsymbol{\beta} + 2n (d\sigma)^2 \right), \tag{8}$$

which is, up to a linear coordinate change, the Poincaré hyperbolic metric of the upper half space  $\mathbb{R}^m \times \mathbb{R}^+$ , see [17]. The Riemannian curvature  $\kappa = -\frac{1}{2}$  is constant and negative

and the unique geodesic, parameterized by the arc length, which connects two points  $\theta_1 = (\beta_1, \sigma_1)$  and  $\theta_2 = (\beta_2, \sigma_2) \in \Theta$ , when  $\beta_1 \neq \beta_2$ , is given by

$$\begin{aligned} \beta(s) &= 2n K^{-2} \tanh\left(\frac{s}{\sqrt{2}} + \epsilon\right) (\mathbf{X}^t \mathbf{X})^{-1/2} \mathbf{C} + \mathbf{D} \\ \sigma(s) &= K^{-1} \operatorname{sech}\left(\frac{s}{\sqrt{2}} + \epsilon\right) \end{aligned} \tag{9}$$

where  $s$  is the arc length,  $\mathbf{C}$  and  $\mathbf{D}$  are  $m \times 1$  vectors whose components, and also  $\epsilon$ , are integration constants which appear when solving the differential equation geodesic system, such that  $\beta(0) = \beta_1$ ,  $\beta(\rho_{12}) = \beta_2$ , and  $\sigma(0) = \sigma_1$ ,  $\sigma(\rho_{12}) = \sigma_2$ , with  $\rho_{12}$  being the Riemannian distance between  $\theta_1$  and  $\theta_2$ . Finally,  $K$  is given by  $K^2 = 2n \mathbf{C}^t \mathbf{C}$ . When  $\beta_1 = \beta_2$ , the geodesic is given by

$$\beta(s) = \mathbf{D} \quad \sigma(s) = B e^{\pm \frac{s}{\sqrt{2}}} \tag{10}$$

where  $B$  is a positive integration constant.

The Rao distance  $\rho$  between the points  $\theta_1$  and  $\theta_2$  is

$$\rho_{12} \equiv \rho(\theta_1, \theta_2) = \sqrt{2} \ln\left(\frac{1 + \delta(\theta_1, \theta_2)}{1 - \delta(\theta_1, \theta_2)}\right) = 2\sqrt{2} \operatorname{arctanh}(\delta(\theta_1, \theta_2)) \tag{11}$$

where

$$\delta(\theta_1, \theta_2) = \left( \frac{d_M^2(\beta_1, \beta_2) + 2 \frac{(\sigma_1 - \sigma_2)^2}{\sigma_1 \sigma_2}}{d_M^2(\beta_1, \beta_2) + 2 \frac{(\sigma_1 + \sigma_2)^2}{\sigma_1 \sigma_2}} \right)^{1/2} \tag{12}$$

and

$$d_M^2(\theta_1, \theta_2) = \frac{1}{n \sigma_1 \sigma_2} (\beta_1 - \beta_2)^t \mathbf{X}^t \mathbf{X} (\beta_1 - \beta_2) \tag{13}$$

or, equivalently,

$$\rho(\theta_1, \theta_2) = \sqrt{2} \operatorname{arccosh}\left(\frac{1}{4} d_M^2(\theta_1, \theta_2) + \frac{\sigma_1^2 + \sigma_2^2}{2 \sigma_1 \sigma_2}\right) \tag{14}$$

Let  $\exp_{\theta_1}^{-1}(\theta_2)$  be the inverse of the exponential map corresponding to Levi-Civita connection and  $W^1, \dots, W^m, W^{m+1}$ , with its components corresponding to the basis field  $(\partial/\partial\beta^1)_{\theta_1}, \dots, (\partial/\partial\beta^m)_{\theta_1}, (\partial/\partial\sigma)_{\theta_1}$ . Then, we have

$$\begin{aligned} W^i &= \frac{\rho_{12}/\sqrt{2}}{\sinh(\rho_{12}/\sqrt{2})} (\beta_2^i - \beta_1^i) \frac{\sigma_1}{\sigma_2}, \quad i = 1, \dots, m \\ W^{m+1} &= \frac{\rho_{12}/\sqrt{2}}{\sinh(\rho_{12}/\sqrt{2})} \left( \cosh(\rho_{12}/\sqrt{2}) - \frac{\sigma_1}{\sigma_2} \right) \sigma_1 \end{aligned} \tag{15}$$

It is well known that the Riemannian distance induced by the information metric is invariant under equivariant estimator transformations. We shall supply direct and alternative proof for the linear model setting.

**Proposition 2.** The Rao distance  $\rho$  given by (11) is invariant under the action of the induced group by  $\mathcal{G}$  on the parameter space,  $\overline{\mathcal{G}}$ . In other words,

$$\rho(\overline{\mathcal{G}(a, H, c)} \theta_1, \overline{\mathcal{G}(a, H, c)} \theta_2) = \rho(\theta_1, \theta_2)$$

**Proof.** Observe that

$$HX(\beta_1 - \beta_2) \in F, \forall H \in O_F(n)$$

and taking into account that  $X(X^tX)^{-1}X^t$  is the projection matrix into  $F$ , we have

$$X(X^tX)^{-1}X^tHX(\beta_1 - \beta_2) = HX(\beta_1 - \beta_2)$$

Therefore,

$$(\beta_1 - \beta_2)^t X^t H^t X (X^t X)^{-1} X^t H X (\beta_1 - \beta_2) = (\beta_1 - \beta_2)^t X^t X (\beta_1 - \beta_2)$$

and the invariance of  $\delta$  and  $\rho$  trivially follows. □

**Proposition 3.**  $\overline{\mathcal{G}}$  acts transitively on  $\Theta$ .

**Proof.** The transitivity follows observing that  $a$  is an arbitrary positive real number and  $X(X^tX)^{-1}X^t$  is the projection matrix into  $F$  with  $\text{rank } X = m$ , the dimension of  $F$ . □

Since  $\rho$ , and thus  $\rho^2$ , is invariant under the action of  $\overline{\mathcal{G}}$ , and  $\overline{\mathcal{G}}$  acts transitively on  $\Theta$ , the distribution of  $\rho^2(\mathbf{U}_\lambda(\mathbf{y}), \theta)$  does not depend on  $\theta$ , and therefore, the risk of any equivariant estimator remains constant and independent of the target parameter provided that this risk is finite. More precisely, observe that if we let

$$\mathbf{z} = \frac{1}{\sigma} (\mathbf{y}_F - X\boldsymbol{\beta}), \quad \mathcal{V} = \|\mathbf{z}\|^2 \quad \text{and} \quad \mathcal{W} = \frac{1}{\sigma} \|\mathbf{y}_{F^\perp}\|^2 \tag{16}$$

from (1) and (4), we clearly have  $\mathbf{z} \sim N_n(\mathbf{0}, X(X^tX)^{-1}X^t)$  with a rank  $m$  idempotent covariance matrix, and  $\mathcal{V}^2$  and  $\mathcal{W}^2$  are independent random variables following a chi-square distribution with  $m$  and  $(n - m)$  degrees of freedom, equal to the dimensions of  $F$  and  $F^\perp$  since  $\mathcal{V}^2$  and  $\mathcal{W}^2$  are quadratic forms based on the projection matrices on these subspaces of  $\mathbb{R}^n$ , and  $\mathbf{y}_F$  (or  $\mathbf{z}$ ) and  $\mathbf{y}_{F^\perp}$  are independent random vectors. Therefore, since  $X\mathbf{U}_{\lambda 1}(\mathbf{y}) = \mathbf{y}_F$  and  $U_{\lambda 2}(\mathbf{y}) = \lambda \|\mathbf{y}_{F^\perp}\|^2$ , we have

$$d_M^2(\mathbf{U}_\lambda(\mathbf{y}), \theta) = \frac{1}{n U_{\lambda 2}(\mathbf{y}) \sigma} (\mathbf{U}_{\lambda 1}(\mathbf{y}) - \boldsymbol{\beta})^t X^t X (\mathbf{U}_{\lambda 1}(\mathbf{y}) - \boldsymbol{\beta}) = \frac{\mathcal{V}^2}{n \lambda \mathcal{W}} \tag{17}$$

$$\delta(\mathbf{U}_\lambda(\mathbf{y}), \theta) = \left( \frac{\frac{\mathcal{V}^2}{n \lambda \mathcal{W}} + 2 \frac{(\lambda \mathcal{W} - 1)^2}{\lambda \mathcal{W}}}{\frac{\mathcal{V}^2}{n \lambda \mathcal{W}} + 2 \frac{(\lambda \mathcal{W} + 1)^2}{\lambda \mathcal{W}}} \right)^{1/2} \tag{18}$$

and

$$\rho(\mathbf{U}_\lambda(\mathbf{y}), \theta) = 2\sqrt{2} \operatorname{arctanh} \left( \left( \frac{\mathcal{V}^2 + 2n(\lambda \mathcal{W} - 1)^2}{\mathcal{V}^2 + 2n(\lambda \mathcal{W} + 1)^2} \right)^{1/2} \right) \tag{19}$$

or

$$\rho(\mathbf{U}_\lambda(\mathbf{y}), \theta) = \sqrt{2} \operatorname{arccosh} \left( \frac{1}{4} \frac{\mathcal{V}^2}{n \lambda \mathcal{W}} + \frac{1}{2} \left( \lambda \mathcal{W} + \frac{1}{\lambda \mathcal{W}} \right) \right) \tag{20}$$

which have a distribution depending only on  $\mathcal{V}^2$  and  $\mathcal{W}^2$ , with independent random variables with fixed distribution, whatever the value of  $\theta$ .

Since the risk of any equivariant estimator remains constant on the parameter space, it is enough to examine it at one point, for instance, at the point  $(\mathbf{0}, 1) \in \Theta$ . Let us denote the expectation with respect to the  $n$ -variate linear normal model  $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$  by  $E_{(\boldsymbol{\beta}, \sigma)}$  and by  $E$  the  $E_{(\mathbf{0}, 1)}$ . We can prove the following propositions.

**Proposition 4.** *Since  $n \geq m + 1$  we have*

$$E_{(\boldsymbol{\beta}, \sigma)} \left( \rho^2(\mathbf{U}_\lambda(\mathbf{y}), (\boldsymbol{\beta}, \sigma)) \right) = E \left( \rho^2(\mathbf{U}_\lambda(\mathbf{y}), (\mathbf{0}, 1)) \right) < \infty, \forall \lambda > 0$$

for any  $(\boldsymbol{\beta}, \sigma) \in \Theta$ .

**Proof.** From (14) and (13), since

$$1 + \frac{1}{2!} \frac{\rho^2}{2} + \frac{1}{4!} \frac{\rho^4}{4} \leq \cosh\left(\frac{\rho}{2}\right) = \frac{1}{4} d_M^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) + \frac{\sigma_1^2 + \sigma_2^2}{2 \sigma_1 \sigma_2} \tag{21}$$

we have

$$0 \leq \rho^2(\theta_1, \theta_2) \leq 12 \left( \sqrt{1 + \frac{d_M^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{6} + \frac{1}{3} \frac{(\sigma_1 - \sigma_2)^2}{\sigma_1 \sigma_2}} - 1 \right) \tag{22}$$

developing the square of the difference and taking into account that the standard Euclidean norm of a vector is less or equal to the absolute value of the sum of its components, we obtain

$$0 \leq \rho^2(\theta_1, \theta_2) \leq 2\sqrt{6} d_M(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) + 4\sqrt{3} \left( \sqrt{\frac{\sigma_1}{\sigma_2}} + \sqrt{\frac{\sigma_2}{\sigma_1}} \right) + 4\sqrt{3} - 12 \tag{23}$$

Notice that both bounds (22) and (23) are invariant under the action of the induced group on the parameter space.

As we mentioned before, from [18], it is enough to prove that the risk is finite at  $(\mathbf{0}, 1) \in \Theta$ . Taking into account (16), it follows from (23) that

$$0 \leq \rho^2(\mathbf{U}_\lambda(\mathbf{y}), (\mathbf{0}, 1)) \leq 2\sqrt{6} \sqrt{\frac{\mathcal{V}^2}{\lambda \mathcal{W} n}} + 4\sqrt{3} \left( \sqrt{\lambda \mathcal{W}} + \frac{1}{\sqrt{\lambda \mathcal{W}}} \right) + 4\sqrt{3} - 12 \tag{24}$$

Observe that if  $Q$  has a chi-square distribution with  $k$  degrees of freedom,

$$E(Q^\alpha) = 2^\alpha \frac{\Gamma(\frac{k}{2} + \alpha)}{\Gamma(\frac{k}{2})} \quad \text{provided that} \quad \frac{k}{2} + \alpha > 0$$

Therefore, since  $\mathcal{V}^2$  and  $\mathcal{W}^2$  are independent random variables following a central chi-square distribution with  $m$  and  $(n - m)$  degrees of freedom, we have

$$E(\mathcal{V}) = \sqrt{2} \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})}, \quad E(\sqrt{\mathcal{W}}) = \sqrt[4]{2} \frac{\Gamma(\frac{2(n-m)+1}{4})}{\Gamma(\frac{n-m}{2})} \quad \text{provided that} \quad n - m > -1/2$$

and

$$E\left(\frac{1}{\sqrt{\mathcal{W}}}\right) = \frac{1}{\sqrt[4]{2}} \frac{\Gamma(\frac{2(n-m)-1}{4})}{\Gamma(\frac{n-m}{2})} \quad \text{provided that} \quad n - m > 1/2$$

Then, taking the average, it follows that

$$\begin{aligned}
 E\left(\rho^2(\mathbf{U}_\lambda(\mathbf{y}), (0, 1))\right) &< \frac{2\sqrt{6}}{\sqrt{\lambda} n} E(\mathcal{V}) E\left(\frac{1}{\sqrt{\mathcal{W}}}\right) + \\
 &+ 4\sqrt{3}\left(\sqrt{\lambda} E(\sqrt{\mathcal{W}}) + \frac{1}{\sqrt{\lambda}} E\left(\frac{1}{\sqrt{\mathcal{W}}}\right)\right) + 4\sqrt{3} - 12 \\
 &< +\infty, \quad \forall \lambda > 0 \text{ and } n - m > 1/2
 \end{aligned}$$

Since  $n$  and  $m$  are positive integers, with  $n > m$ , we conclude that the risk is finite if  $n \geq m + 1$ . □

Proposition 4 is a sufficient condition for the existence of the Riemannian risk of the equivariant estimator  $\mathbf{U}_\lambda$ , thus

$$\Phi(\lambda) = E\left(\rho^2(\mathbf{U}_\lambda(\mathbf{y}), (0, 1))\right), \quad \lambda > 0, \tag{25}$$

is well defined.

**Proposition 5.** *There exists a unique minimizer  $\lambda_{nm} > 0$  of the Riemannian risk given by  $\Phi$ .*

**Proof.** Let us consider the Riemannian risk at  $\lambda = e^{\frac{s}{2\sqrt{2}}}$  as a function of  $s$ , that is,

$$F(s) = \Phi\left(e^{\frac{s}{2\sqrt{2}}}\right), \quad s \in \mathbb{R}$$

The particular selection of  $\lambda$ , from which  $F$  follows, relies on the Riemannian structure of  $\Theta$  induced by the information metric. The Riemannian curvature is constant and equal to  $-\frac{1}{2}$ , and taking into account (10), we have that

$$\gamma(s) = \left( (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}, e^{\frac{s}{\sqrt{2}}} \left\{ \mathbf{y}^t \left( \mathbf{I}_n - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \right) \mathbf{y} \right\}^{1/2} \right)$$

is a geodesic in  $\Theta$ ; precisely, a geodesic parameterized by the arc length; see [17] for further details.

Then, following [19], the real valued function

$$s \mapsto \rho^2(\gamma(s), (0, 1))$$

is strictly convex. Since almost surely convexity of a stochastic process carries over the mean of a process, the map  $F$  is strictly convex as well.

On the other hand, from Fatou’s Lemma  $F(s) \rightarrow +\infty$  as  $s \rightarrow -\infty$  or  $s \rightarrow \infty$ . This, together with the strict convexity of  $F$ , yields the existence of a unique minimizer  $s^*$  of the function  $F$ , which depends on  $n$  and  $m$ .

The optimal value of  $\lambda$  is therefore obtained by expressing the unique minimizer  $s^*$  of  $F(s)$  in the original parameterization  $\lambda = e^{s/(2\sqrt{2})}$ , which provides the correspondence between the minimization problems in  $s$  and in  $\lambda$ .

Finally, since the map  $s \mapsto e^{\frac{s}{2\sqrt{2}}}$  is a strictly monotonous function; a unique  $\lambda_{nm}$  must exist, namely  $\lambda_{nm} = e^{\frac{s^*}{2\sqrt{2}}}$ , such that  $\Phi(\lambda_{nm}) = \min_{\lambda > 0} \Phi(\lambda)$ . □

In fact, this result guarantees the unicity of the MIRE, although a numerical analysis is required to obtain it explicitly (see next section). It could be useful to develop a simple

approximate estimator, that shall be referred to hereafter as *a*-MIRE, obtained, luckily, by minimizing a convenient upper bound of  $\Phi(\lambda)$ . Since

$$1 + \frac{1}{2!} \frac{\rho^2}{2} \leq \cosh\left(\frac{\rho}{2}\right) = \frac{1}{4} d_M^2(\theta_1, \theta_2) + \frac{\sigma_1^2 + \sigma_2^2}{2 \sigma_1 \sigma_2} \tag{26}$$

we shall have

$$0 \leq \rho^2(\mathbf{U}_\lambda(\mathbf{y}), (\mathbf{0}, 1)) \leq \frac{1}{n} \frac{\mathcal{V}^2}{\lambda \mathcal{W}} + 2 \left( \sqrt{\lambda \mathcal{W}} - \frac{1}{\sqrt{\lambda \mathcal{W}}} \right)^2 \tag{27}$$

and therefore, taking expectations, we have

$$0 \leq \Phi(\lambda) \leq H(\lambda) = \frac{1}{\lambda} \frac{\Gamma(\frac{n-m}{2} - \frac{1}{2})}{\sqrt{2} \Gamma(\frac{n-m}{2})} \left( \frac{m}{n} + 2 \right) + 2\lambda\sqrt{2} \frac{\Gamma(\frac{n-m}{2} + \frac{1}{2})}{\Gamma(\frac{n-m}{2})} - 4 \tag{28}$$

the upper bound  $H(\lambda)$  is clearly a convex function with an absolute minimum attained when  $\lambda$  satisfies

$$\widetilde{\lambda}_{nm} = \left( \sqrt{1 + \frac{m}{2n}} \right) \frac{1}{\sqrt{n - m - 1}} \tag{29}$$

Furthermore, given an arbitrary  $m$ , we have

$$\lim_{n \rightarrow \infty} \left( \widetilde{\lambda}_{nm} \right)^2 n = 1 \tag{30}$$

and, therefore, *a*-MIRE is very close to MLE for large values of  $n$ . Observe also that it is possible to compute *a*-MIRE for  $n > m + 1$ , a condition which is slightly stronger than the result required for the existence of MIRE in Proposition 4.

A further aspect is the intrinsic bias of the equivariant estimators. In fact, connections between minimum risk, bias, and invariance have been established; see [18]. Since the action of the group  $G$  is not commutative, we cannot guarantee the unbiasedness of the MIRE, and an additional analysis must be performed. First of all, we compute the vector bias, see [8], a quantitative measure of the bias which is compatible with the Lehmann results.

Let  $A_\theta(\lambda) = \exp_\theta^{-1}(\mathbf{U}_\lambda(\mathbf{y}))$  and  $A_\theta^1(\lambda), \dots, A_\theta^m(\lambda), A_\theta^{m+1}(\lambda)$  be the components of  $A_\theta(\lambda)$  corresponding to the basis field  $(\partial/\partial\beta^1)_\theta, \dots, (\partial/\partial\beta^m)_\theta, (\partial/\partial\sigma)_\theta$ . With matrix notation,  $A_{\theta 1}(\lambda) = (A_\theta^1(\lambda), \dots, A_\theta^m(\lambda))^t$ . Furthermore, let us define  $h(x) \equiv x/\sinh x$  for  $x \neq 0$  and  $h(0) \equiv 0$ ; taking into account (16) and (19), from (11) and (15), we have

$$\begin{aligned} \mathbf{X}A_{\theta 1}(\lambda) &= f_\lambda(\mathcal{V}, \mathcal{W}) \mathbf{z} \frac{\sigma}{\lambda \mathcal{W}} \\ A_\theta^{m+1}(\lambda) &= f_\lambda(\mathcal{V}, \mathcal{W}) g_\lambda(\mathcal{V}, \mathcal{W}) \sigma \end{aligned} \tag{31}$$

where

$$f_\lambda(\mathcal{V}, \mathcal{W}) \equiv h(\rho(\mathbf{U}_\lambda(\mathbf{y}), \theta)/\sqrt{2})$$

and

$$g_\lambda(\mathcal{V}, \mathcal{W}) \equiv \cosh(\rho(\mathbf{U}_\lambda(\mathbf{y}), \theta)/\sqrt{2}) - \frac{1}{\lambda \mathcal{W}}$$

Let  $\mathbf{B}_\theta(\lambda) = E_\theta(\exp_\theta^{-1}(\mathbf{U}_\lambda(\mathbf{y})))$  be the intrinsic bias vector corresponding to an equivariant estimator  $\mathbf{U}_\lambda(\mathbf{y}) = (U_{\lambda 1}(\mathbf{y}), U_{\lambda 2}(\mathbf{y}))$  evaluated at the point  $\theta = (\beta, \sigma)$  and let  $B_\theta^1(\lambda), \dots, B_\theta^m(\lambda), B_\theta^{m+1}(\lambda)$  be their components. In matrix notation,  $\mathbf{B}_{\theta 1}(\lambda) = (B_\theta^1(\lambda), \dots, B_\theta^m(\lambda))^t$ . We have the following.

**Proposition 6.** *If  $n \geq m + 1$ , the bias vector is finite and*

$$\begin{aligned} \mathbf{B}_{\theta 1}(\lambda) &= \mathbf{0} \\ B_{\theta}^{m+1}(\lambda) &= E(f_{\lambda}(\mathcal{V}, \mathcal{W}) \ g_{\lambda}(\mathcal{V}, \mathcal{W})) \sigma \end{aligned} \tag{32}$$

where  $\mathcal{V}^2$  and  $\mathcal{W}^2$  are independent random variables following a chi-square distribution with  $m$  and  $n - m$  degrees of freedom, respectively.

Moreover, the square of the norm of the bias vector is constant and given by

$$\|\mathbf{B}_{\theta}(\lambda)\|_{\theta}^2 = 2 ( E( f_{\lambda}(\mathcal{V}, \mathcal{W}) \ g_{\lambda}(\mathcal{V}, \mathcal{W}) ) )^2 \tag{33}$$

**Proof.** Observe that if  $n \geq m + 1$ , we have

$$\begin{aligned} \|\mathbf{B}_{\theta}(\lambda)\|_{\theta}^2 &\leq E_{\theta}(\|\exp_{\theta}^{-1}(\mathbf{U}_{\lambda}(\mathbf{y}))\|_{\theta})^2 \leq E_{\theta}(\|\exp_{\theta}^{-1}(\mathbf{U}_{\lambda}(\mathbf{y}))\|_{\theta}^2) \\ &= E_{\theta}(\rho^2(\mathbf{U}_{\lambda}(\mathbf{y}), \theta)) < \infty \end{aligned}$$

where  $\|\cdot\|_{\theta}$  denotes the Riemannian norm at the tangent space at  $\theta$ .

On the other hand, taking into account (31) and defining  $\mathbf{z}$  as in (16), observe that  $\mathcal{V} = \|\mathbf{z}\| = \|\mathbf{-z}\|$ ,  $\mathbf{z}$  is independent of  $\mathcal{W}$ , and  $\mathbf{z}$  and  $\mathbf{-z}$  have the same distribution. Then we have

$$\begin{aligned} \mathbf{X}\mathbf{B}_{\theta 1}(\lambda) &= E_{\theta}\left(f_{\lambda}(\mathcal{V}, \mathcal{W}) \ \mathbf{z} \frac{\sigma}{\lambda\mathcal{W}}\right) \\ &= E_{\theta}\left(f_{\lambda}(\mathcal{V}, \mathcal{W}) \ (\mathbf{-z}) \frac{\sigma}{\lambda\mathcal{W}}\right) = -E_{\theta}\left(f_{\lambda}(\mathcal{V}, \mathcal{W}) \ \mathbf{z} \frac{\sigma}{\lambda\mathcal{W}}\right) = \mathbf{0} \end{aligned}$$

and since  $\text{rank}(\mathbf{X}) = m$ , it follows that  $\mathbf{B}_{\theta 1}(\lambda) = \mathbf{0}$ .

$B_{\theta}^{m+1}(\lambda)$  is obtained directly from (31). The distribution of  $\mathbf{z}$  and  $\mathcal{W}^2$  follow from the basic properties of multivariate normal distribution. Finally, the norm of the bias vector field follows from (32) and (8). □

We may remark, finally, that the norm of the bias vector field of any equivariant estimator,  $\|\mathbf{B}_{\theta}(\lambda)\|_{\theta}$ , is invariant under the action of the induced group,  $\overline{\mathcal{G}}$ , on the parameter space, and since this group acts transitively on  $\Theta$ , this quantity must be constant, which is clear from (33).

### 4. Numerical Evaluation

In this section, we compare, numerically, the MIRE estimator with the standard MLE. Observe that both estimators only differ in the estimation of the parameter  $\sigma$  (or  $\sigma^2$ ). Precisely, the MIRE and the MLE of  $\sigma^2$  are, respectively,

$$\widehat{\sigma}^2 = \lambda_{nm}^2 \mathbf{y}^t \left( \mathbf{I}_n - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \right) \mathbf{y} \quad \sigma^{*2} = \frac{1}{n} \mathbf{y}^t \left( \mathbf{I}_n - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \right) \mathbf{y} \tag{34}$$

namely, they only differ by the factor  $\zeta_{nm} = \lambda_{nm}^2 n$ , since  $\widehat{\sigma}^2 = \zeta_{nm} \sigma^{*2}$ .

In order to compare MLE with the MIRE, we computed the factor  $\zeta_{nm}$ , the intrinsic risk, and the square of the norm of the bias vector for each estimator. All the computations were performed using Mathematica 10.2, specifically, the main procedure used was NIntegrate, with the parameter AccuracyGoal  $\rightarrow 16$  [20].

Moreover, we suggested a rather simple approximation for  $\zeta_{nm}$ , or  $\lambda_{nm}$ , which allows us to approximate the MIRE estimator through (29), i.e.,

$$\widetilde{\zeta}_{nm} = \left( 1 + \frac{m}{2n} \right) \frac{n}{n - m - 1} \tag{35}$$

The corresponding estimator, which shall be referred to hereafter as *a*-MIRE, was also compared with MLE and MIRE in terms of the intrinsic risk.

The results are summarized in the following figures which summarize the tables given in Appendix A; see [21] for a convincing argument for the use of plots over tables. In Figure 1, numerical results of  $\zeta_{nm}$  (left) and  $\Phi(\lambda_{nm}) = \Phi(\sqrt{\zeta_{nm}/n})$  (right) are displayed graphically. Observe that, for *m* fixed, as *n* increases,  $\Phi(\sqrt{\zeta_{nm}/n})$  goes to zero and  $\zeta_{nm}$  goes to one. The exact numerical values are given in Appendix A, Table A1.

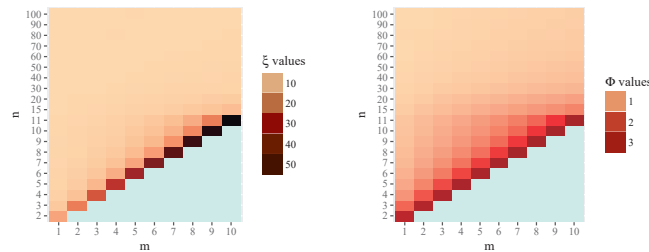


Figure 1. Numerical results for  $\zeta_{nm}$  (left) and  $\Phi(\lambda_{nm})$  (right), with  $\zeta_{nm} = \lambda_{nm}^2 n$ .

Figure 2 graphically shows the percentage of the intrinsic risk increment for MLE (left) and *a*-MIRE (right), that is

$$100 \frac{\left(\Phi\left(\frac{1}{\sqrt{n}}\right) - \Phi(\lambda_{nm})\right)}{\Phi(\lambda_{nm})} \quad \text{and} \quad 100 \frac{\left(\Phi(\tilde{\lambda}_{nm}) - \Phi(\lambda_{nm})\right)}{\Phi(\lambda_{nm})} \tag{36}$$

respectively, for some values of *n* and *m*. The exact numerical results are given in Appendix A, Table A2.

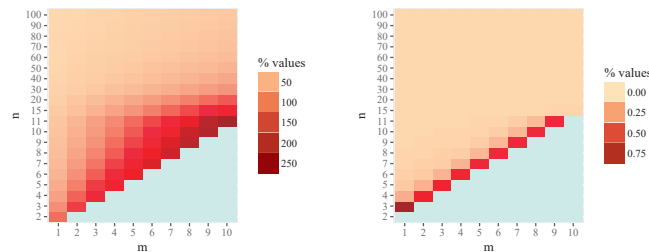


Figure 2. Percentage of intrinsic risk increment for MLE (left) and *a*-MIRE (right).

Observe that if we approximate the MIRE by the MLE for a certain value of *m*, the relative difference of risk decreases as *n* increases. Notice that these relative differences of risks are rather moderate or small (about 10–15%) only if  $n > 10m$ . Let us remark also the non-appropriate behavior of the MLE for small values of  $n - m$ . This is in regard to the intrinsic risk increment: when we use MLE instead of MIRE, this increment oscillates between 80.5% and 251.1% when  $n - m = 1$ , with *m* from 1 to 10. On the other hand, the behavior of the *a*-MIRE is reasonably good, with its intrinsic risk very similar to the risk of the MIRE estimator: here, the percentage on risk increment is less than 1% for all studied cases, and also, this percentage decreases as *n* increases. In fact, this percentage is lower than 1‰ when  $n - m \geq 4$ , which indicates the extraordinary degree of approximation; therefore, the *a*-MIRE is a reasonable and useful approximation of MIRE. In particular, as an example, for a two-way analysis of variance, with *a* and *b* levels for factors *A* and *B*, respectively, with a single replicate for treatment, we shall have

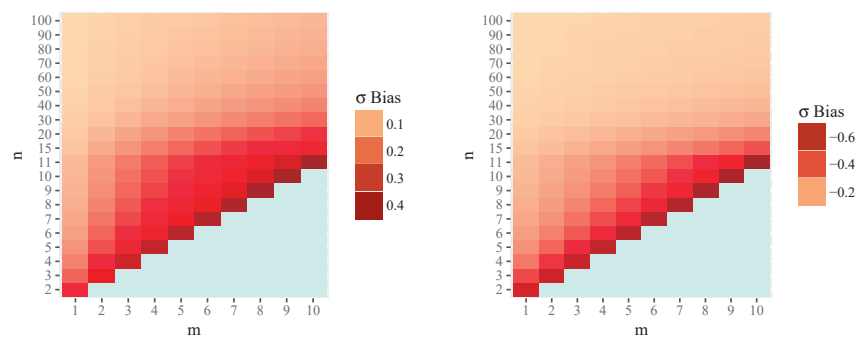
$$\tilde{\lambda} = \left( \sqrt{1 + \frac{a + b - 1}{2ab}} \right) \frac{1}{\sqrt{ab - a - b}} \tag{37}$$

with  $a > 1$  and  $b > 1$ , while the corresponding quantity for MLE is  $\lambda^* = \frac{1}{\sqrt{ab}}$ .  
 In particular, if  $a = 3$  and  $b = 5$ , we shall have

$$\tilde{\lambda} = \left( \sqrt{1 + \frac{7}{30}} \right) \frac{1}{\sqrt{7}} = 0.419750$$

and  $\lambda^* = \frac{1}{\sqrt{15}} = 0.258199$ , which is sensibly different. At this moment, it could be useful to recall that the quadratic loss and the squared Riemannian distance behave very differently; see [9].

Figure 3 graphically displays the numerical results of the  $m + 1$  component of the bias vector divided by  $\sigma$ , i.e., the unique non-zero physical component of this vector field, for MIRE (left) and MLE (right),  $\frac{1}{\sigma} B_{\theta}^{m+1}(\lambda_{nm})$  and  $\frac{1}{\sqrt{n}\sigma} B_{\theta}^{m+1}\left(\frac{1}{\sqrt{n}}\right)$ , respectively, for some values of  $n$  and  $m$ . Observe the sign differences, meaning that MIRE overestimates, on average,  $\sigma$ , while MLE underestimates this quantity on average as well. This follows from the equation of the geodesics of the present model (9). The exact numerical values are given in Appendix A, Table A3.

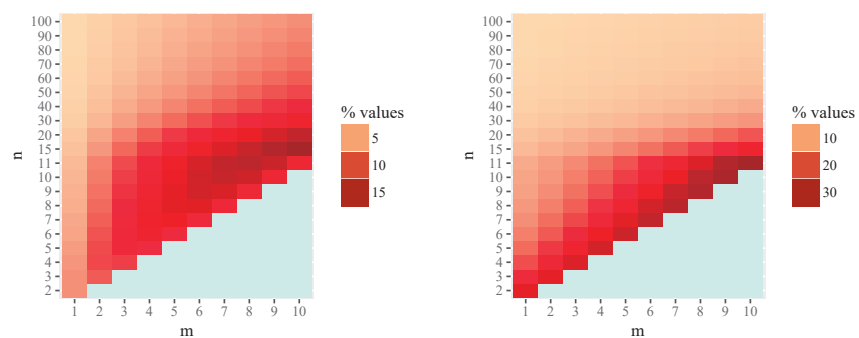


**Figure 3.** Numerical results for the unique non-null physical component of bias vector field  $\frac{1}{\sigma} B_{\theta}^{m+1}(\lambda)$ , for MIRE ((left), with  $\lambda = \lambda_{nm}$ ) and MLE ((right), with  $\lambda = 1/\sqrt{n}$ ).

Figure 4 graphically shows the numerical results of the percentage of intrinsic risk due to bias for MIRE and MLE, that is

$$100 \frac{\|B\|_{\theta}^2(\lambda_{nm})}{\Phi(\lambda_{nm})} \quad \text{and} \quad 100 \frac{\|B\|_{\theta}^2\left(\frac{1}{\sqrt{n}}\right)}{\Phi(\lambda_{nm})} \quad (38)$$

respectively, for some values of  $n$  and  $m$ . Observe that the bias is moderate, with respect to the intrinsic risk, in both estimators. The bias of the MIRE estimator is smaller than the bias of MLE for small values of  $m$  and the opposite for large values. The exact numerical results are given in Appendix A, Table A4.



**Figure 4.** Percentage of intrinsic risk due to bias for MIRE (left) and MLE (right), referred to as the risk of MIRE.

## 5. Conclusions

In this work, we characterized the class of equivariant estimators for the univariate linear normal model under the largest subgroup of the affine group that preserves the column space of the design matrix. Within this class, and using the squared Rao distance as an intrinsic loss function, we established the existence and uniqueness of the minimum-risk equivariant estimator (MIRE). We also derived an explicit expression for the intrinsic bias of any equivariant estimator and compared the intrinsic risk and bias of the MIRE with those of the maximum likelihood estimator.

The numerical evaluation shows that

- The intrinsic risk and bias of the MLE can be substantially larger than those of the MIRE in small-sample settings.
- The difference between the two estimators decreases as the sample size increases.
- The approximate estimator (*a*-MIRE) provides a practical and computationally efficient alternative, achieving intrinsic risk values nearly identical to those of the MIRE for moderate sample sizes.

The results highlight the importance of incorporating the geometric structure of the parameter space in statistical estimation. When the parameter space is endowed with the Fisher–Rao metric, the squared Rao distance provides a natural intrinsic loss function that captures aspects of performance not reflected by classical criteria such as quadratic loss.

### *Future Research Directions*

This work suggests several avenues for further development:

- Extension to heteroscedastic or correlated error structures, where the geometry of the model may differ significantly.
- Intrinsic Bayesian formulations based on noninformative or reference priors, where posterior summaries may be interpreted in geometric terms.
- Computational schemes for approximating the Rao distance efficiently in high-dimensional or large-sample contexts.

We hope that these contributions will encourage further exploration of intrinsic methods in statistical inference and strengthen the connection between information geometry and classical estimation theory.

**Author Contributions:** Formal analysis, J.M.O.; Investigation, G.G. and M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Acknowledgments:** We have to thank the referees and the editor comments and suggestions for the improvement of this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

**Table A1.** Numerical results for  $\zeta_{nm}$  (above) and  $\Phi(\lambda_{nm})$  (below), with  $\zeta_{nm} = \lambda_{nm}^2 n$ .

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
2	8.66992 3.03876									
3	3.07654 1.16322	13.85950 3.22626								
4	2.15828 0.71941	4.40442 1.32989	19.07110 3.31958							
5	1.79608 0.52257	2.87441 0.86369	5.72625 1.42579	24.29080 3.37544						
6	1.60525 0.41144	2.28365 0.65040	3.59243 0.95721	7.05039 1.48877	29.51720 3.41264					
7	1.48787 0.33979	1.97404 0.52570	2.77223 0.73918	4.31144 1.02277	8.37564 1.53330	34.74390 3.43918				
8	1.40849 0.28964	1.78417 0.44287	2.34344 0.60912	3.26141 0.80445	5.03105 1.07128	9.70161 1.56646	39.97520 3.45908			
9	1.35126 0.25252	1.65604 0.38342	2.08088 0.52102	2.71323 0.67272	3.75096 0.85446	5.75096 1.10863	11.02800 1.59210	45.20590 3.47455		
10	1.30807 0.22391	1.56381 0.33847	1.90388 0.45665	2.37787 0.58233	3.08328 0.72281	4.24078 0.89402	6.47135 1.13827	12.35490 1.61252	50.43540 3.48692	
11	1.27433 0.20116	1.49427 0.30320	1.77656 0.40721	2.15190 0.51544	2.67503 0.63172	3.45350 0.76331	4.73076 0.92608	7.19181 1.16237	13.68190 1.62918	55.65840 3.49704
15	1.19071 0.14318	1.33089 0.21469	1.49647 0.28618	1.69505 0.35810	1.93754 0.43107	2.24025 0.50594	2.62870 0.58406	3.14513 0.66756	3.86473 0.76019	4.93543 0.86917
20	1.13807 0.10534	1.23410 0.15768	1.34212 0.20973	1.46452 0.26160	1.60436 0.31342	1.76568 0.36539	1.95380 0.41771	2.17599 0.47070	2.44246 0.52476	2.76780 0.58046
30	1.08895 0.06895	1.14768 0.10318	1.21093 0.13718	1.27923 0.17097	1.35322 0.20457	1.43364 0.23801	1.52137 0.27132	1.61744 0.30453	1.72312 0.33769	1.83991 0.37084
40	1.06561 0.05126	1.10786 0.07673	1.15245 0.10205	1.19960 0.12721	1.24952 0.15224	1.30246 0.17713	1.35872 0.20191	1.41860 0.22657	1.48247 0.25114	1.55075 0.27561
50	1.05197 0.04080	1.08495 0.06108	1.11936 0.08126	1.15532 0.10134	1.19289 0.12131	1.23222 0.14119	1.27342 0.16098	1.31663 0.18068	1.36200 0.20029	1.40970 0.21983
60	1.04303 0.03388	1.07007 0.05074	1.09808 0.06752	1.12711 0.08423	1.15722 0.10086	1.18846 0.11742	1.22090 0.13391	1.25461 0.15033	1.28967 0.16669	1.32617 0.18298

**Table A1.** *Cont.*

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
70	1.03671 0.02897	1.05963 0.04340	1.08324 0.05776	1.10757 0.07207	1.13268 0.08632	1.15857 0.10051	1.18530 0.11466	1.21291 0.12874	1.24144 0.14278	1.27093 0.15676
80	1.03201 0.02531	1.05189 0.03791	1.07230 0.05047	1.09324 0.06298	1.11476 0.07545	1.13687 0.08787	1.15959 0.10025	1.18295 0.11259	1.20697 0.12489	1.23170 0.13715
90	1.02838 0.02246	1.04593 0.03366	1.06390 0.04481	1.08228 0.05593	1.10111 0.06701	1.12039 0.07806	1.14014 0.08907	1.16038 0.10005	1.18112 0.11099	1.20239 0.12190
100	1.02548 0.02020	1.04120 0.03026	1.05725 0.04029	1.07363 0.05030	1.09036 0.06027	1.10745 0.07022	1.12492 0.08014	1.14276 0.09002	1.16101 0.09988	1.17966 0.10972

**Table A2.** Percentage of intrinsic risk increment for MLE (above) and  $a$ -MIRE (below).

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
2	80.54 –									
3	57.08 0.95	114.18 –								
4	42.51 0.12	87.63 0.64	140.61 –							
5	33.73 0.04	67.93 0.11	114.44 0.61	162.61 –						
6	27.88 0.02	54.83 0.04	90.96 0.10	138.22 0.58	181.58 –					
7	23.72 0.01	45.78 0.02	74.42 0.03	112.04 0.10	159.65 0.56	198.31 –				
8	20.63 0.01	39.22 0.01	62.67 0.02	92.70 0.03	131.51 0.09	179.21 0.55	213.33 –			
9	18.25 0.00	34.27 0.01	54.00 0.01	78.62 0.01	109.84 0.03	149.62 0.09	197.26 0.54	226.98 –		
10	16.35 0.00	30.42 0.00	47.38 0.01	68.08 0.01	93.73 0.01	125.99 0.03	166.59 0.09	214.02 0.53	239.51 –	
11	14.81 0.00	27.33 0.00	42.19 0.00	59.96 0.00	81.53 0.01	108.10 0.01	141.27 0.03	182.58 0.08	229.71 0.53	251.10 –
15	10.75 0.00	19.42 0.00	29.26 0.00	40.45 0.00	53.25 0.00	68.04 0.00	85.30 0.00	105.68 0.00	130.06 0.01	159.58 0.01
20	8.00 0.00	14.25 0.00	21.13 0.00	28.68 0.00	37.01 0.00	46.24 0.00	56.51 0.00	68.00 0.00	80.96 0.00	95.69 0.00
30	5.29 0.00	9.29 0.00	13.56 0.00	18.11 0.00	22.95 0.00	28.11 0.00	33.63 0.00	39.52 0.00	45.85 0.00	52.66 0.00

**Table A2.** *Cont.*

<i>n</i> \ <i>m</i>	1	2	3	4	5	6	7	8	9	10
40	3.96 0.00	6.89 0.00	9.98 0.00	13.23 0.00	16.62 0.00	20.18 0.00	23.91 0.00	27.83 0.00	31.95 0.00	36.28 0.00
50	3.16 0.00	5.48 0.00	7.90 0.00	10.42 0.00	13.03 0.00	15.74 0.00	18.55 0.00	21.47 0.00	24.51 0.00	27.67 0.00
60	2.63 0.00	4.54 0.00	6.54 0.00	8.59 0.00	10.71 0.00	12.90 0.00	15.15 0.00	17.48 0.00	19.88 0.00	22.35 0.00
70	2.25 0.00	3.88 0.00	5.57 0.00	7.31 0.00	9.09 0.00	10.93 0.00	12.81 0.00	14.74 0.00	16.72 0.00	18.75 0.00
80	1.97 0.00	3.39 0.00	4.86 0.00	6.36 0.00	7.90 0.00	9.48 0.00	11.09 0.00	12.74 0.00	14.42 0.00	16.15 0.00
90	1.75 0.00	3.01 0.00	4.30 0.00	5.63 0.00	6.99 0.00	8.37 0.00	9.78 0.00	11.22 0.00	12.68 0.00	14.18 0.00
100	1.57 0.00	2.70 0.00	3.86 0.00	5.05 0.00	6.26 0.00	7.49 0.00	8.74 0.00	10.02 0.00	11.32 0.00	12.64 0.00

**Table A3.** Numerical results for the unique non-null physical component of bias vector field  $\frac{1}{\sigma} B_{\theta}^{m+1}(\lambda)$ , for MIRE (above, with  $\lambda = \lambda_{nm}$ ) and MLE (below, with  $\lambda = 1/\sqrt{n}$ ).

<i>n</i> \ <i>m</i>	1	2	3	4	5	6	7	8	9	10
2	0.25273 −0.64316									
3	0.15729 −0.34089	0.33352 −0.66450								
4	0.11879 −0.23333	0.23107 −0.38021	0.37337 −0.68608							
5	0.09542 −0.17763	0.18380 −0.27128	0.27237 −0.41182	0.39710 −0.70605						
6	0.07989 −0.14348	0.15393 −0.21196	0.22483 −0.30232	0.29910 −0.43827	0.41288 −0.72422					
7	0.06877 −0.12037	0.13280 −0.17426	0.19348 −0.24073	0.25307 −0.32857	0.31782 −0.46104	0.42409 −0.74076				
8	0.06039 −0.10369	0.11693 −0.14809	0.17044 −0.20063	0.22201 −0.26549	0.27371 −0.35132	0.33165 −0.48102	0.43251 −0.75589			
9	0.05384 −0.09107	0.10451 −0.12881	0.15255 −0.17224	0.19859 −0.22366	0.24357 −0.28719	0.28944 −0.37137	0.34228 −0.49884	0.43904 −0.76980		

**Table A3.** *Cont.*

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
10	0.04858 −0.08120	0.09451 −0.11401	0.13817 −0.15101	0.17997 −0.19362	0.22044 −0.24409	0.26044 −0.30651	0.30184 −0.38931	0.35073 −0.51492	0.44424 −0.78265	
11	0.04426 −0.07326	0.08628 −0.10227	0.12633 −0.13451	0.16471 −0.17089	0.20175 −0.21278	0.23791 −0.26244	0.27399 −0.32390	0.31186 −0.40554	0.35758 −0.52956	0.44844 −0.79460
15	0.03267 −0.05266	0.06405 −0.07250	0.09426 −0.09377	0.12339 −0.11673	0.15156 −0.14172	0.17887 −0.16918	0.20545 −0.19972	0.23144 −0.23423	0.25707 −0.27404	0.28272 −0.32133
20	0.02461 −0.03897	0.04848 −0.05318	0.07164 −0.06810	0.09414 −0.08383	0.11602 −0.10047	0.13732 −0.11814	0.15809 −0.13700	0.17836 −0.15724	0.19818 −0.17912	0.21761 −0.20294
30	0.01649 −0.02564	0.03264 −0.03470	0.04845 −0.04405	0.06394 −0.05370	0.07913 −0.06368	0.09403 −0.07402	0.10864 −0.08474	0.12298 −0.09588	0.13707 −0.10748	0.15090 −0.11958
40	0.01240 −0.01911	0.02460 −0.02576	0.03661 −0.03256	0.04843 −0.03953	0.06008 −0.04666	0.07154 −0.05397	0.08284 −0.06146	0.09397 −0.06916	0.10495 −0.07707	0.11576 −0.08520
50	0.00994 −0.01523	0.01974 −0.02048	0.02942 −0.02583	0.03899 −0.03128	0.04843 −0.03682	0.05775 −0.04248	0.06696 −0.04824	0.07606 −0.05413	0.08506 −0.06013	0.09394 −0.06626
60	0.00829 −0.01266	0.01649 −0.01700	0.02460 −0.02141	0.03262 −0.02588	0.04056 −0.03042	0.04842 −0.03503	0.05620 −0.03971	0.06390 −0.04447	0.07152 −0.04931	0.07906 −0.05423
70	0.00711 −0.01083	0.01415 −0.01453	0.02113 −0.01827	0.02805 −0.02207	0.03490 −0.02591	0.04169 −0.02980	0.04842 −0.03375	0.05509 −0.03775	0.06170 −0.04180	0.06825 −0.04591
80	0.00622 −0.00946	0.01240 −0.01269	0.01852 −0.01594	0.02460 −0.01924	0.03062 −0.02257	0.03660 −0.02594	0.04253 −0.02934	0.04842 −0.03279	0.05425 −0.03628	0.06005 −0.03981
90	0.00554 −0.00840	0.01103 −0.01126	0.01649 −0.01414	0.02190 −0.01705	0.02728 −0.01999	0.03262 −0.02296	0.03792 −0.02595	0.04319 −0.02898	0.04841 −0.03204	0.05360 −0.03514
100	0.00498 −0.00756	0.00993 −0.01012	0.01485 −0.01270	0.01974 −0.01531	0.02459 −0.01794	0.02942 −0.02059	0.03421 −0.02327	0.03898 −0.02597	0.04371 −0.02870	0.04841 −0.03145

**Table A4.** Percentage of intrinsic risk due to bias for MIRE (above) and MLE (below), referred to as the risk of MIRE.

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
2	4.20 27.23									
3	4.25 19.98	6.90 27.37								
4	3.92 15.14	8.03 21.74	8.40 28.36							

Table A4. Cont.

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
5	3.48 12.08	7.82 17.04	10.41 23.79	9.34 29.54						
6	3.10 10.01	7.29 13.81	10.56 19.10	12.02 25.80	9.99 30.74					
7	2.78 8.53	6.71 11.55	10.13 15.68	12.52 21.11	13.18 27.73	10.46 31.91				
8	2.52 7.42	6.17 9.90	9.54 13.22	12.25 17.52	13.99 23.04	14.04 29.54	10.82 33.04			
9	2.30 6.57	5.70 8.66	8.93 11.39	11.72 14.87	13.89 19.31	15.11 24.88	14.72 31.26	11.10 34.11		
10	2.11 5.89	5.28 7.68	8.36 9.99	11.12 12.88	13.45 16.49	15.17 21.02	16.01 26.63	15.26 32.88	11.32 35.13	
11	1.95 5.34	4.91 6.90	7.84 8.89	10.53 11.33	12.89 14.33	14.83 18.05	16.21 22.66	16.73 28.30	15.70 34.43	11.50 36.11
15	1.49 3.87	3.82 4.90	6.21 6.14	8.50 7.61	10.66 9.32	12.65 11.31	14.45 13.66	16.05 16.44	17.39 19.76	18.39 23.76
20	1.15 2.88	2.98 3.59	4.89 4.42	6.78 5.37	8.59 6.44	10.32 7.64	11.97 8.99	13.52 10.51	14.97 12.23	16.32 14.19
30	0.79 1.91	2.06 2.33	3.42 2.83	4.78 3.37	6.12 3.96	7.43 4.60	8.70 5.29	9.93 6.04	11.13 6.84	12.28 7.71
40	0.60 1.42	1.58 1.73	2.63 2.08	3.69 2.46	4.74 2.86	5.78 3.29	6.80 3.74	7.80 4.22	8.77 4.73	9.72 5.27
50	0.48 1.14	1.28 1.37	2.13 1.64	3.00 1.93	3.87 2.24	4.72 2.56	5.57 2.89	6.40 3.24	7.22 3.61	8.03 3.99
60	0.41 0.95	1.07 1.14	1.79 1.36	2.53 1.59	3.26 1.83	3.99 2.09	4.72 2.36	5.43 2.63	6.14 2.92	6.83 3.21
70	0.35 0.81	0.92 0.97	1.55 1.16	2.18 1.35	2.82 1.56	3.46 1.77	4.09 1.99	4.71 2.21	5.33 2.45	5.94 2.69
80	0.31 0.71	0.81 0.85	1.36 1.01	1.92 1.18	2.49 1.35	3.05 1.53	3.61 1.72	4.16 1.91	4.71 2.11	5.26 2.31
90	0.27 0.63	0.72 0.75	1.21 0.89	1.72 1.04	2.22 1.19	2.73 1.35	3.23 1.51	3.73 1.68	4.22 1.85	4.71 2.03
100	0.25 0.57	0.65 0.68	1.10 0.80	1.55 0.93	2.01 1.07	2.47 1.21	2.92 1.35	3.37 1.50	3.83 1.65	4.27 1.80

## References

1. Park, C.; Gao, X.; Wang, M. Robust explicit estimators using the power-weighted repeated medians. *J. Appl. Stat.* **2023**, *51*, 1590–1608. [[CrossRef](#)] [[PubMed](#)]
2. Morikawa, K.; Terada, Y.; Kim, J.K. Semiparametric adaptive estimation under informative sampling. *Ann. Stat.* **2025**, *53*, 1347–1369. [[CrossRef](#)]
3. Jiang, J. Asymptotic distribution of maximum likelihood estimator in generalized linear mixed models with crossed random effects. *Ann. Stat.* **2025**, *53*, 1298–1318. [[CrossRef](#)]
4. Rao, C. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–91.
5. Burbea, J.; Rao, C. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *J. Multivar. Anal.* **1982**, *12*, 575–596. [[CrossRef](#)]
6. Burbea, J. Informative geometry of probability spaces. *Expo. Math.* **1986**, *4*, 347–378.
7. Amari, S.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society; Oxford University Press: Oxford, UK, 2000.
8. Oller, J.; Corcuera, J. Intrinsic Analysis of the Statistical Estimation. *Ann. Stat.* **1995**, *23*, 1562–1581. [[CrossRef](#)]
9. García, G.; Oller, J. What does intrinsic mean in statistical estimation? *Sort* **2006**, *2*, 125–146.
10. Bernal-Casas, D.; Oller, J. Variational Information Principles to Unveil Physical Laws. *Mathematics* **2024**, *12*, 3941. [[CrossRef](#)]
11. Bernardo, J.; Juárez, M. Intrinsic estimation. In *Bayesian Statistics 7*; Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Hackerman, W., Smith, A., West, M., Eds.; Oxford University Press: Berlin, Germany, 2003; pp. 465–476.
12. Oller, J.M.; Corcuera, J.M. Intrinsic Bayesian Estimation Using the Kullback–Leibler Divergence. *Ann. Inst. Stat. Math.* **2003**, *55*, 355–365.
13. Lehmann, E.; Casella, G. *Theory of Point Estimation*; Springer Science & Business Media: New York, NY, USA, 2006.
14. Amari, S.I. *Information Geometry and Its Applications*; Applied Mathematical Sciences; Springer Japan: Tokyo, Japan, 2016; Volume 194.
15. Ay, N.; Jost, J.; Lê, H.V.; Schwachhöfer, L. *Information Geometry*; Springer: Cham, Switzerland, 2017.
16. Nielsen, F. An Elementary Introduction to Information Geometry. *Entropy* **2020**, *22*, 1100. [[CrossRef](#)] [[PubMed](#)]
17. Burbea, J.; Oller, J.M. The information metric for univariate linear elliptic models. *Stat. Risk Model.* **1988**, *6*, 209–222. [[CrossRef](#)]
18. Lehmann, E. A general concept of unbiasedness. *Ann. Math. Stat.* **1951**, *22*, 587–592. [[CrossRef](#)]
19. Karcher, H. Riemannian Center of Mass and Mollifier Smoothing. *Commun. Pure Appl. Math.* **1977**, *30*, 509–541. [[CrossRef](#)]
20. Wolfram Research, Inc. *Mathematica, Version 10.2*; Wolfram Research, Inc.: Champaign, IL, USA, 2015. Available online: <https://www.wolfram.com/mathematica/> (accessed on 1 July 2016).
21. Gelman, A.; Pasarica, C.; Dodhia, R. Let’s practice what we preach: Turning tables into graphs. *Am. Stat.* **2002**, *56*, 121–130. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.