Shrinkage Strategy In Stratified Random Sample Subject To Measurement Error

Séverien Nkurunziza*

Abstract

The empirical likelihood estimation approach has been used in statistical applications. In many situations, samples are subject to measurement errors. For example, samples may be obtained from one perfect and several imperfect instruments. The empirical likelihood estimation method successfully combines the perfect and imperfect measurements in the estimation process. With this framework, we propose shrinkage estimation strategy by using maximum empirical likelihood estimator (MELE) as the base estimator. We develop the asymptotic properties of the shrinkage estimators by using the notion of asymptotic distributional risk. Our asymptotic results clearly demonstrate the superiority of our proposed shrinkage strategy over the MELE. Monte Carlo simulation results show that such performance still holds in finite samples.

Keywords: ADR; Empirical likelihood; Measurement errors; Stratified random sampling; RMELE; Shrinkage methods; UMELE; Uncertain prior information.

*University of Windsor, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4. Email: severien@uwindsor.ca
1 Introduction

In this paper, we consider a statistical model where the observations are obtained by stratified sampling method similar to that in Wu and Zhang (2005). In particular, we consider the scenario where for each stratum, the sampling design is similar to that described in Zhong et al. (2000). Briefly, we consider $M$ strata and, for the $j^{th}$ stratum, $j = 1, 2, \ldots, M$, we have $H + 1$ independent random samples $s^{(j)}, s^{(j)}_1, \ldots, s^{(j)}_H$ of sizes $n^{(j)}_0, n^{(j)}_1, \ldots, n^{(j)}_H$ are drawn and measurements on $s^{(j)}_h$ are obtained through instrument $h$, and the samples are selected independently across the strata. Also, let $N$ and $N_j$ be the total population and the $j^{th}$ stratum sizes respectively. Further, let $n = \sum_{j=1}^{M} \sum_{h=0}^{H} n^{(j)}_h$. For each $j = 1, 2, \ldots, M$, the instrument 0 is considered as perfect and the rest as giving imperfect measurements. The actual measurements are denoted by $\left\{ s^{(j)}_{hi}, i \in s^{(j)}_h, h = 0, 1, \ldots, H \right\}$, $j = 1, 2, \ldots, M$ and the interest is to estimate the $p \times M$-matrix $\theta = \left( \theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(M)} \right)$ where for each $j = 1, 2, \ldots, M$, $\theta^{(j)}$ is a $p$-column vector with $p \leq H + 1$. Note that the sampling plan considered in Zhong et al. (2000) and Wu and Zhang (2007) become special cases with $M = 1$. Further, we extend the results in Wu and Zhang (2005) where stratified sampling in the presence of measurement error is also considered. In the quoted papers, the authors used the empirical likelihood methodology to construct an estimator that uses the perfect as well as the imperfect measurements in an effective way. To give other references where empirical likelihood approach is used under stratified sampling plan, we quote Zhong and Rao (2000) and for a recent research work, we refer to Fang et al. (2009).

In this paper, we consider to improve upon the maximum empirical likelihood estimator (MELE) of the parameter matrix. The proposed estimators outperformed over UMELE. The proposed shrinkage-based technique can be applied to a host of statistical problems. Following Wu and Zhang (2005, 2007) and Zhong et al. (2000), the empirical likelihood constitutes an effective way of combining
sample measurements from one perfect instrument and several imperfect instruments in the process of estimating unknown parameters of a population. This scenario is a common practical fact in survey data as well as in industrial quality control context.

We consider similar sampling plan as in Wu and Zhang (2005), which can be viewed as the multidimensional version Zhong et al. (2000) and Wu and Zhang (2007) frameworks. Namely, for each $j = 1, 2, \ldots, M$, let $\mathbf{g}^{(j)}(y, \theta^{(j)}) = \left( g_{0}^{(j)}(y, \theta^{(j)}), g_{1}^{(j)}(y, \theta^{(j)}), \ldots, g_{H}^{(j)}(y, \theta^{(j)}) \right)'$ be a $p \times H + 1$ matrix-valued, functionally independent and unbiased estimating equation, i.e. $E\left( \mathbf{g}^{(j)}(Y, \theta^{(j)}) \right) = 0$. For each $j = 1, 2, \ldots, M$, the profile log-empirical likelihood ratio (ELR) is then given by

$$r_{n}^{(j)}(\theta^{(j)}) = \sum_{h=0}^{H} \sum_{i \in s_{h}^{(j)}} \ln \left( 1 + \lambda_{h}^{(j)} \mathbf{g}_{h}^{(j)}(y_{hi}^{(j)}, \theta^{(j)}) \right), \quad (1.1)$$

where $\lambda_{h}^{(j)}$, is a vector of Lagrange multipliers satisfying

$$\sum_{i \in s_{h}^{(j)}} \frac{\mathbf{g}_{h}^{(j)}(y_{hi}^{(j)}, \theta^{(j)})}{1 + \lambda_{h}^{(j)} \mathbf{g}_{h}^{(j)}(y_{hi}^{(j)}, \theta^{(j)})} = 0, \quad h = 0, 1, \ldots, H + 1, \quad j = 1, 2, \ldots, M. \quad (1.2)$$

Thus, since the samples are independently selected across the strata, the joint profile log-empirical likelihood ratio (ELR) is then given by

$$r_{n}(\theta) = \sum_{j=1}^{M} \sum_{h=0}^{H} \sum_{i \in s_{h}^{(j)}} \ln \left( 1 + \lambda_{h}^{(j)} \mathbf{g}_{h}^{(j)}(y_{hi}^{(j)}, \theta^{(j)}) \right). \quad (1.3)$$

We consider a matrix estimation problem in measurement error models when there are many potential variables and some of them may not be of relevant. Such situation arises, for instance, in the context of regression when an investigator suspects a priori that several factors have zero effect on the outcome of interest. In such cases, a prior hypothesis is formed that the coefficients of the factors are zero. Relying completely on the prior information leads to restricted estimators whereas completely ignoring it leads to unrestricted estimators. We propose Shrinkage-type estimators that improve the performance of the maximum empirical likelihood estimator (MELE). The MELE underlying is derived
by using similar method as proposed by Zhong et al. (2000) in the context of perfect/imperfect instru-
ments. In particular, we consider two competing models, where one model includes all variables and
the other restricts parameters to a candidate linear subspace based on prior knowledge. With respect to
these two models, we investigate the relative performances of the Stein-type shrinkage estimators. We
develop large sample theory for the estimators including derivation of asymptotic distributional risk.
Asymptotic and a Monte Carlo simulation studies show that the shrinkage estimators perform better
than UMELE.

Uncertain candidate subspace (UCS): As a motivating context, consider first the situation where the
parameter matrix $\theta$ is suspected to lie in the following subspace:

$$\text{UCS}_1 : L_0\theta^{(1)} = L_0\theta^{(2)} = \cdots = L_0\theta^{(M)}$$  \hspace{1cm} (1.4)

with $L_0$ is known $q \times p$-matrix full rank with $q \leq p$. Note that the restriction in (1.4) includes a special
case where the $M$ strata are suspected to be homogeneous. Indeed, by taking $L_0$ as the identity matrix,
the restriction in (1.4) becomes $\theta^{(1)} = \theta^{(2)} = \cdots = \theta^{(M)}$. Also, note that the restriction in (1.4) can be
rewritten as $L_0\theta L_1 = 0$ where $L_1$ is $M \times (M - 1)$-matrix given by

$$L_1 = \begin{pmatrix}
-1 & 0 & 0 & \cdots & 0 & 0 \\
1 & -1 & 0 & \cdots & 0 & 0 \\
0 & 1 & -1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 & -1 \\
0 & 0 & 0 & \cdots & 0 & 1
\end{pmatrix}.$$  \hspace{1cm} (1.5)

Thus, in the sequel and for the sake of generality, we consider the following candidate subspace

$$\text{UCS}_2 : F\theta L = d$$  \hspace{1cm} (1.6)
where $F$ and $L$ are known, respectively $q \times p$ and $M \times r$-matrices of full rank with $q \leq p$ and $r \leq M$; $d$ is known $q \times r$-matrix.

The maximization of (1.3) subject to (1.2) gives unrestricted maximum empirical likelihood estimator (UMELE) which completely ignores the prior knowledge (UCS) given in (1.6). On the other hand, the maximization of (1.3) subject to (1.2) and (1.6) yields restricted estimator (RMELE).

The reminder of this paper is organized as follows. Section 2 gives asymptotic properties of the RMELE and UMELE estimators. It also presents the shrinkage and positive shrinkage estimators for combining the RMELE and UMELE in an optimal way and we give the asymptotic properties of the shrinkage estimators. In Section 3, we present Monte Carlo simulation results. Section 4 presents concluding remarks whereas technical proofs and details are given in the Appendix.

2 Unrestricted and Restricted Estimation Strategies

Let $\hat{\theta}$ be the UMELE of the parameter matrix $\theta$, in the context of stratified sampling with measurement errors described in section 1. The $\hat{\theta} = \left(\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots, \hat{\theta}^{(M)}\right)$ is obtained by maximizing (1.3) subject to (1.2). More precisely, for each $j = 1, 2, \ldots, M$, $\hat{\theta}^{(j)}$ the solution to the following system of equations,

$$
\sum_{i \in s_h^{(j)}} \frac{g_h^{(j)}(y_{hi}, \theta^{(j)})}{1 + \lambda_h^{(j)\gamma} g_h^{(j)}(y_{hi}, \theta^{(j)})} = 0, \quad h = 0, 1, \ldots, H, \quad \sum_{h=0}^{H} \sum_{i \in s_h^{(j)}} \left[ \frac{\partial g_h^{(j)}(y_{hi}, \theta^{(j)})}{\partial \theta^{(j)}}\right]' \lambda_h^{(j)} = 0. \quad (2.1)
$$

Further, as mentioned above, the RMELE $\tilde{\theta}$ is obtained by maximizing (1.3) subject to (1.2) and (1.6). To this end, let $\mu$ be $q \times r$-matrix. In the similar way as in Qin and Lawless (1994), from (1.3)-(1.6) and Lagrangian multipliers, the restricted estimator RMELE $\tilde{\theta}$ is obtained from the following system
of equations,
\[
\sum_{i \in s_h^{(j)}} g_h^{(j)} \left( y_{hi}, \theta^{(j)} \right) = 0, \quad h = 0, 1, \ldots, H, \quad j = 1, 2, \ldots, M; \quad (2.2)
\]
\[
\left( \sum_{h=0}^{H} \sum_{i \in s_h^{(j)}} \left[ \frac{\partial g_h^{(j)} (y_{hi}, \theta^{(j)})}{\partial \theta^{(j)}} \right] \lambda_h^{(j)} \right)_{j=1, 2, \ldots, M} + F' \mu L' = 0, \quad F \theta L - d = 0. \quad (2.3)
\]

2.1 Shrinkage Estimation Strategies (SES)

We consider the following Stein-type shrinkage (SMELE) based on the maximum empirical likelihood estimators UMELE and RMELE of Section 1,
\[
\hat{\theta}_S \leftarrow \tilde{\theta} + \left\{ 1 - c \psi_n^{-1} \right\} (\hat{\theta} - \tilde{\theta}), \quad \text{where} \quad 0 \leq c < 2(qr - 2), \quad \text{with} \quad qr > 2. \quad (2.4)
\]
In the similar way as in Nkurunziza and Ahmed (2009), we take \( c = qr - 2 \). Also, as in the quoted paper, the function \( \psi_n \) in (2.4) represents the test statistic, testing the prior knowledge (the null hypothesis) given in (1.6). Here, the quantity \( \psi_n \) is nonnegative almost surely. Hence, \( \psi_n < c \iff 1 - c \psi_n^{-1} < 0 \), which may cause a phenomenon known as over-shrinkage. Because of that, we also propose the positive-part shrinkage estimator (PSMELE) derived from SMELE. Namely, PSMELE is given by
\[
\hat{\theta}_S^+ = \tilde{\theta} + \max \left\{ 0, (1 - c \psi_n^{-1}) \right\} (\hat{\theta} - \tilde{\theta}). \quad (2.5)
\]
In general, the UMELE, RMELE do not have a closed form and hence, it is impossible to obtain the finite sample risk of the above estimators. As in Nkurunziza and Ahmed (2009), Jurečková and Sen (1996), we overcome this difficulty by using asymptotic methods. In particular, we use the concept of asymptotic distributional risk (ADR) as defined in Sen and Saleh (1987), Ahmed et al. (2006, 2009). To this end, let \( \delta \) be nonzero \( q \times r \)-matrix linearly independent with the matrix \( d \) with \( \| \delta \| < \infty \) and consider the following sequence of alternatives to (1.6),
\[
K_n : F \theta L = d + \delta / \sqrt{n}, \quad n = 1, 2, 3, \ldots \quad (2.6)
\]
2.2 Asymptotic results

In this subsection, we present the ADR of the proposed estimators. For establishing the ADR, the estimating functions in (2.3) need to satisfy the following regularities conditions.

\( \mathcal{A}_1 \) For each \( j = 1, 2, \ldots, M \), \( \sum_{h=0}^{H} \sum_{i \in s_h^{(j)}} g_h^{(j)} (y_{hi}, \theta^{(j)}) g_h^{(j)' (y_{hi}, \theta^{(j)})} \) is a positive definite matrix with probability one;

\( \mathcal{A}_2 \) For each \( j = 1, 2, \ldots, M \), \( E \left[ \left\| g^{(j)} (X, \theta^{(j)}) g^{(j)' (X, \theta^{(j)})} \right\| \right] < \infty \) and for every \( h = 0, 1, \ldots, H \),
\[
E \left[ \sum_{i \in s_h^{(j)}} g_h (y_{hi}^{(j)}, \theta^{(j)}) g_h^{(j)' (y_{hi}, \theta^{(j)})} \right]
\]
is a positive definite matrix;

\( \mathcal{A}_3 \) For each \( j = 1, 2, \ldots, M \), the function \( g^{(j)} (x, \theta) \) is two-differentiable in \( \theta \) and
\[
\partial^2 g^{(j)} (x, \theta^{(j)}) / \partial \theta^{(j)} \partial \theta^{(j)'} \text{ is continuous in a neighborhood of the true value } \theta_0^{(j)}. \]
Further,
\[
\| \partial g^{(j)} (x, \theta^{(j)}) / \partial \theta^{(j)} \|, \| \partial^2 g^{(j)} (x, \theta^{(j)}) / \partial \theta^{(j)} \partial \theta^{(j)'} \| \text{ and } \| g^{(j)} (x, \theta^{(j)}) \|^3
\]
are bounded by some integrable functions \( G^{(j)} (x) \) in this neighborhood.

\( \mathcal{A}_4 \) For each \( h = 0, 1, \ldots, H, j = 1, 2, \ldots, M \), the matrices
\[
E \left( \partial g_h^{(j)} (X, \theta_0^{(j)}) / \partial \theta^{(j)} \right), \sum_{i \in s_h^{(j)}} E \left( \partial g_h^{(j)} (y_{hi}, \theta_0^{(j)}) / \partial \theta^{(j)} \right)
\]
are full rank.

\( \mathcal{A}_5 \) As \( n \to \infty, N_j, n_h^{(j)} \to \infty, n_h^{(j)}/n_0^{(j)} \to k_h^{(j)} > 0 \) and \( n_h^{(j)}/n \to \alpha_j > 0 \) for each \( j = 1, 2, \ldots, M \).

\( \mathcal{A}_6 \) As a function of the index \( j = 1, 2, \ldots, M \), \( \sum_{h=0}^{H} k_h^{(j)} E \left( \frac{\partial g_h^{(j)}}{\partial \theta^{(j)}} \right) \left[ E \left( g_h^{(j)} g_h^{(j)'} \right) \right]^{-1} E \left( \frac{\partial g_h^{(j)}}{\partial \theta^{(j)}} \right)' \)
is a constant.

It is noticed that the conditions in (\( \mathcal{A}_1 \))-(\( \mathcal{A}_4 \)) correspond to that in Qin and Lawless (1994, Lemma 1, Theorem 1). Also, Assumptions (\( \mathcal{A}_2 \)), (\( \mathcal{A}_3 \)) and (\( \mathcal{A}_5 \)) is similar to that in Wu and Zhang (2005, Theorem 1), and Assumption (\( \mathcal{A}_6 \)) corresponds to the homogeneity of variance-covariance in multivariate
multiple linear model. Now, let \( \varrho_n = \sqrt{n} (\hat{\theta} - \theta) L \), \( \xi_n = \sqrt{n} (\hat{\theta} - \tilde{\theta}) L \) and let \( \zeta_n = \sqrt{n} (\tilde{\theta} - \theta) L \). Further, let

\[
V = \sum_{h=0}^{H} k_h^{(j)} \mathbb{E} \left( \frac{\partial g_h^{(1)}}{\partial \theta^{(1)}} \right) \left[ \mathbb{E} \left( g_h^{(1)} g_h^{(1)'} \right) \right]^{-1} \mathbb{E} \left( \frac{\partial g_h^{(1)}}{\partial \theta^{(1)}} \right) ^{'}^{-1} ; \Omega = \text{Diag}(\alpha_1, \alpha_2, \ldots, \alpha_M) . \tag{2.7}
\]

\( J_0 = VF (FV F')^{-1} ; \delta^* = -J_0 \delta ; \) and let \( \Sigma^* = J_0 FV \).

Under the above notations and regularity conditions, the UMELE and RMELE are consistent estimators of \( \theta \). Proposition 2.1 given below shows that the UMELE is asymptotically normal under the local alternatives (2.6). Furthermore, Proposition 2.2 shows that the UMELE and RMELE are jointly asymptotically normal under the local alternatives (2.6). Thus, we extend the result in Qin and Lawless (1994) and Zhong et al. (2000) since we consider the multidimensional model and the sequence of local alternatives in (2.6) includes the null hypothesis. Furthermore, we consider the joint asymptotic distribution of the restricted and unrestricted estimators.

**Proposition 2.1** If (\( \mathcal{A}_1 \))-(\( \mathcal{A}_6 \)) hold, then under the local alternative \( K_n \) in (2.6), we have

\[
\sqrt{n} (\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}_{p \times M} (0, V \otimes \Omega) .
\]

□

The proof follows directly from the results which are proved in Zhong et al. (2000), Wu and Zhang (2005, Theorem 1; 2007, Lemma 2). Further, for our paper to be self-contained, we outlined the proof in the Appendix.

**Proposition 2.2** If (\( \mathcal{A}_1 \))-(\( \mathcal{A}_6 \)) hold, then under the local alternative \( K_n \) in (2.6), we have

\[
\begin{pmatrix}
\varrho_n \\
\xi_n
\end{pmatrix}
\xrightarrow{n \to \infty}
\begin{pmatrix}
\varrho \\
\xi
\end{pmatrix}
\sim \mathcal{N}_{2p \times 2r}
\begin{pmatrix}
0 \\
-\delta^*
\end{pmatrix}
\left(
\begin{pmatrix}
V \otimes (L' \Omega L) & (V - \Sigma^*) \otimes (L' \Omega L) \\
(V - \Sigma^*) \otimes (L' \Omega L) & (V - \Sigma^*) \otimes (L' \Omega L)
\end{pmatrix}
\right).
\]
\[
\begin{pmatrix}
\xi_n \\
\zeta_n
\end{pmatrix}
\xrightarrow{n \to \infty}
\begin{pmatrix}
\xi \\
\zeta
\end{pmatrix}
\sim \mathcal{N}_{2p \times 2r}
\begin{pmatrix}
\delta^* \\
-\delta^*
\end{pmatrix},
\begin{pmatrix}
\Sigma^* \otimes (L'\Omega L) & 0 \\
0 & (V - \Sigma^*) \otimes (L'\Omega L)
\end{pmatrix}.
\]
\[ \square \]

**Corollary 2.1** Let \( \Delta = \text{trace} \left( \delta' (FV F')^{-1} \delta (L' \Omega L)^{-1} \right) \) and let \( \chi_m^2 (\Delta) \) denote a chi-square random variate with \( m \) degrees of freedom and noncentrality parameter \( \Delta \). If Proposition 2.2 holds, then
\[
\text{trace} \left( \xi_n' F (FV F')^{-1} F \xi_n (L' \Omega L)^{-1} \right) \xrightarrow{n \to \infty} \psi \sim \chi_{qr}^2 (\Delta).
\]
\[ \square \]

Corollary 2.1 follows from Proposition 2.2 which is established by using similar techniques as used in deriving Proposition 2.2. For completeness, the proofs Proposition 2.1 and Corollary 2.1 are outlined in the Appendix.

It should be noticed that the above proposition and corollary still hold under the null hypothesis in (1.6) by taking \( \delta^* = 0 \) and \( \Delta = 0 \). For practical aspects, let
\[
\hat{V}_j = \left[ \sum_{h=0}^{H} \hat{\zeta}^{(j)}_h \left\{ \sum_{i=1}^{n_h} \hat{p}_{hi}^{(j)} \partial g_h^{(j)} (y_{hi}, \hat{\theta}^{(j)}) / \partial \theta^{(j)} \right\}' \left\{ \sum_{i=1}^{n_h} \hat{p}_{hi}^{(j)} g_h^{(j)} (y_{hi}, \hat{\theta}^{(j)}) g_h^{(j)} (y_{hi}, \hat{\theta}^{(j)})' \right\}^{-1} \cdot \left\{ \sum_{i=1}^{n_h} \hat{p}_{hi}^{(j)} \partial g_h^{(j)} (y_{hi}, \hat{\theta}^{(j)}) / \partial \theta^{(j)} \right\} \right]^{-1}. \tag{2.8}
\]
where \( \hat{p}_{hi}^{(j)} = \hat{p}_{hi}^{(j)} (\hat{\theta}^{(j)}) = \left[ n_h^{(j)} \left( 1 + \lambda_h^{(j)} \frac{g_h^{(j)} (y_{hi}, \theta^{(j)})}{\partial \theta^{(j)}} \right) \right]^{-1} \), \( \partial g_h^{(j)} (y_{hi}, \theta^{(j)}) / \partial \theta^{(j)} \) denote the derivative of \( \partial g_h^{(j)} (y_{hi}, \theta^{(j)}) / \partial \theta^{(j)} \) evaluated at \( \theta^{(j)} = \hat{\theta}^{(j)} \) and \( \hat{k}_{hi}^{(j)} = n_h^{(j)} / n_0^{(j)} \). Following Wu and Zhang (2005) and Zhong et al. (2000) among others, one concludes that, for each \( j = 1, 2, \ldots, M \), the estimator \( \hat{V}_j \) given in (2.8) is consistent estimator of \( V \). So, in this paper, a consistent estimator of the covariance-variance matrix \( V \) is \( \hat{V} = \sum_{j=1}^{M} n^{(j)} \hat{V}_j / n \) with \( n^{(j)} = \sum_{h=0}^{H} n_h^{(j)} \). Further, let
\[ \psi_n = \text{trace} \left( \xi_n F'(F\hat{V} F')^{-1} F \xi_n (L'\hat{\Omega} L)^{-1} \right) \]

where \( \hat{\Omega} = \text{Diag}(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_m) \) with \( \hat{\alpha}_j = n_0^{(j)}/n \). By combining Corollary 2.1 and Slusky’s, we have \( \psi_n \xrightarrow{L} \psi \sim \chi^2_{qr}(\Delta) \). Thus, one can establish a Wald-type \( \alpha \)-level test based on the statistic \( \psi_n \).

### 2.3 Asymptotic Distributional Risk

In this subsection, we give the ADR of the proposed estimators, and this optimality criterion requires the introduction of loss function. In particular, for an estimator \( \hat{\theta}^* \) of \( \theta \), we consider a quadratic loss function of the form

\[ \mathcal{L}(\hat{\theta}^*, \theta; W) = \text{trace} \left\{ \left[ \sqrt{n} L' (\hat{\theta}^* - \theta) \right] W \left[ \sqrt{n} (\hat{\theta}^* - \theta) L \right] \right\}, \] (2.9)

where \( W \) is a \( M \times M \)-nonnegative definite matrix. Taking the expected value both sides of (2.9), we get the expected loss that would be called the quadratic risk \( R^o_n(\hat{\theta}^*, \theta; W) \). Also, suppose that \( \sqrt{n} (\hat{\theta}^* - \theta) L \) converges in law to \( \rho^* \), as \( n \) tends to infinity. Then, the ADR of \( \hat{\theta}^* \) is defined as

\[ \text{ADR}(\hat{\theta}^*, \theta; W) = \mathbb{E}[\text{trace} (\rho^* W (\rho^*)')] \] .

From Proposition 2.2 and Corollary 2.1, we establish the following lemma which is useful in deriving the ADR of the proposed estimators. Let \( A \) be a matrix and let \( \| A \|^2_{\Xi_1, \Xi_2} = \text{trace} (A' \Xi_1 A \Xi_2) \) with \( \Xi_1 \) a known nonnegative definite matrix, and \( \Xi_2 \) a known positive definite matrix. Further, let \( h \) be known Borel measurable and real-valued integrable function. Also, consider the following class (in \( h \)) of quantity \( \hat{\theta}(h) = \hat{\theta}^* + h \left( \| \xi \|^2_{\Xi_1, \Xi_2} \right) \xi / \sqrt{n} \), where \( \xi \) is defined in Proposition 2.2 and \( \hat{\theta}^* \) is a matrix variate \( \sqrt{n} (\hat{\theta}^* - \theta) L \) that has the same distribution as \( \zeta \), given in Proposition 2.2.
Lemma 2.1 Let \( \mathcal{L} \left( \hat{\theta}(h), \theta; W \right) \) be the loss function in (2.9). Then,
\[
E \left( \mathcal{L} \left( \hat{\theta}(h), \theta; W \right) \right) = \text{trace} \left( W (V - \Sigma^*) \right) \text{trace} \left( \Xi_2^{-1} \right) + \text{trace} \left( \delta' W \delta^* \right)
\]
\[
- 2E \left[ h \left( \chi_{qr+2}^2 \left( \text{trace} \left( \Xi_2 \delta'' \Xi_1 \delta^* \right) \right) \right) \right] \text{trace} \left( \delta'' W \delta^* \right)
\]
\[
+ E \left[ h^2 \left( \chi_{qr+2}^2 \left( \text{trace} \left( \Xi_2 \delta'' \Xi_1 \delta^* \right) \right) \right) \right] \text{trace} \left( W \Sigma^* \right) \text{trace} \left( \Xi_2^{-1} \right)
\]
\[
+ E \left[ h^2 \left( \chi_{qr+4}^2 \left( \text{trace} \left( \Xi_2 \delta'' \Xi_1 \delta^* \right) \right) \right) \right] \text{trace} \left( \delta'' W \delta^* \right).
\]
\]

The proof of Lemma 2.1 is outlined in the Appendix. Further, from Lemma 2.1, we derive below Corollary 2.2 that gives more explicit expressions of the ADR of the proposed estimators. Let \( H_v(x; \Delta) = P \{ \chi_v^2(\Delta) \leq x \}, x \geq 0. \)

Corollary 2.2 If Proposition 2.2 holds, then, the ADR of the estimators are given as follows:
\[
\text{ADR} \left( \hat{\theta}, \theta; W \right) = \text{trace}(L' \Omega L) \text{trace}(W'V),
\]
\[
\text{ADR} \left( \hat{\theta}, \theta; W \right) = \text{trace}(L' \Omega L) \text{trace}(W'V) - \text{trace}(L' \Omega L) \text{trace}(W \Sigma^*) + \text{trace}(\delta'' W \delta^*),
\]
\[
\text{ADR} \left( \hat{\theta}^S, \theta; W \right) = \text{ADR} \left( \hat{\theta}, \theta; W \right) + \text{trace}(\delta'' W \delta^*) ( (qr)^2 - 4 ) E(\chi_{qr+4}^{-4}(\Delta))
\]
\[
- (qr - 2) \text{trace} (L' \Omega L) \text{trace}(W \Sigma^*) \left\{ 2E(\chi_{qr+2}^{-2}(\Delta)) - (qr - 2) E(\chi_{qr+2}^{-4}(\Delta)) \right\},
\]
\[
\text{ADR} \left( \hat{\theta}^{S+}, \theta; W \right) = \text{ADR} \left( \hat{\theta}^S, \theta; W \right) + (qr - 2) \text{trace}(L' \Omega L) \text{trace}(W \Sigma^*)
\]
\[
\times \left[ 2E \left\{ \chi_{qr+2}^{-2}(\Delta) I(\chi_{qr+2}^2(\Delta) \leq (qr - 2)) \right\} - (qr - 2) E \left\{ \chi_{qr+2}^{-4}(\Delta) I(\chi_{qr+2}^2(\Delta) \leq (qr - 2)) \right\} \right]
\]
\[
- \text{trace} (L' \Omega L) \text{trace}(W V) H_{qr+2}(qr - 2; \Delta)
\]
\[
+ \text{trace}(\delta'' W \delta^*) \left\{ 2H_{qr+2}(qr - 2; \Delta) - H_{qr+4}(qr - 2; \Delta) \right\}
\]
\[
- (qr - 2) \text{trace}(\delta'' W \delta^*) \left[ 2E(\chi_{qr+2}^{-2}(\Delta)) I(\chi_{qr+2}^2(\Delta) \leq (qr - 2)) \right]
\]
\[
- 2E \left\{ \chi_{qr+4}^{-2}(\Delta) I(\chi_{qr+4}^2(\Delta) \leq (qr - 2)) \right\} + (qr - 2) E \left\{ \chi_{qr+4}^{-4}(\Delta) I(\chi_{qr+4}^2(\Delta) \leq (qr - 2)) \right\}.
\]
\]
Proof The derivation of $\text{ADR} \left( \tilde{\theta}, \theta; W \right)$ and $\text{ADR} \left( \tilde{\theta}, \theta; W \right)$ follow directly from Lemma 2.1 by taking $h = 1$ (with $\Xi = (L'\Omega L)^{-1}$) and $h = 0$ respectively. To derive $\text{ADR} \left( \tilde{\theta}^S, \theta; W \right)$, we apply Lemma 2.1 by taking $\Xi_1 = F'(FV F')^{-1} F$, $\Xi_2 = (L'\Omega L)^{-1}$, $h(x) = 1 - (qr - 2)/x$, $x > 0$ and the rest follows from some algebraic computations as well as the identity $E \left( \chi^{-2}_{qr+4} (\Delta) \right) = E \left( \chi^{-2}_{qr+2} (\Delta) \right) - 2E \left( \chi^{-4}_{qr+4} (\Delta) \right)$. Further, the derivation of $\text{ADR} \left( \tilde{\theta}^{S+}, \theta; W \right)$ follows directly from Lemma 2.1 by taking $h(x) = \max(1 - (qr - 2)/x, 0)$ along with some algebraic computations, that completes the proof.

Remark 2.1 Let $\text{ch}_{\text{max}}(A)$ denote the largest eigenvalue of the matrix $A$. Following a method similar to that in Ahmed and Saleh (1999), for the choice of the matrix

$$W \in \{ W : (qr + 2)\text{ch}_{\text{max}}(WV) \leq 2\text{trace}(WV)\text{trace}(L'\Omega L) \},$$

we have

$$\text{ADR} \left( \tilde{\theta}^S, \theta; W \right) < \text{ADR} \left( \tilde{\theta}, \theta; W \right) \text{ for all } \Delta \geq 0. \text{ Hence, } \tilde{\theta}^S \text{ outperforms over } \tilde{\theta}. \text{ Further, by using Corollary 2.2 along with some algebraic computations, we conclude that } \text{ADR} \left( \tilde{\theta}^{S+}, \theta; W \right) \leq \text{ADR} \left( \tilde{\theta}^S, \theta; W \right) \text{ for all } \Delta \geq 0, \text{ and this proves } \tilde{\theta}^{S+} \text{ is also superior to } \tilde{\theta} \text{ for } W \text{ selected from the set above.}$$

3 Numerical examples

3.1 Simulations studies

In this simulation study, we used two strata (i.e. $M = 2$) and for each strata, we generated a $p$- dimensional multivariate normal model $N(\theta, \Sigma)$, with $p = 6$, $\Sigma = \sigma^2 I_p$, where $I_p$ denote a $p \times p$-identity matrix and $\sigma^2$ the component variances. Also, for each strata, we considered, one sample from perfect machine and one from imperfect machine (i.e. $H = 1$) with equal sample sizes $n_0 = n_1 = 50, 100, 150, 200$
and we generated 500 samples. Further, we assumed that the variance of the perfect machine measurements is \( \sigma_0^2 = 1 \) and that of the imperfect machine is \( \sigma_1^2 = 4 \). As uncertain prior information, we consider \( F \theta L = 0 \); with \( L = I_2 \), \( F = L_1' \) where \( L_1 \) is defined in (1.5); i.e. \( r = 2 \), \( q = p - 1 \). In other words, the uncertain prior knowledge is that the two strata are homogeneous and inside each stratum, the mean components are identical. From Figure 1, it is noticed that the shrinkage estimators dominate UMELE. Further, Figure 1 shows that beyond the small interval near the null hypothesis, SMELE and PSMELE dominate also RMELE.

4 Concluding remarks

In this paper, we considered a stratified sampling subject to measurement errors within each stratum. Further, we proposed empirical likelihood shrinkage estimators that combine both uncertain prior knowledge and information from samples. Also, we established a lemma which is useful in deriving the ADR of a class of shrinkage estimators for parameter matrix. Theoretically, it was established that the proposed shrinkage estimators dominate the UMELE. Further, Monte Carlo simulation studies corroborate our theoretical findings.

Appendix

A Technical results and proofs

For the sake of simplicity, we denote the left-hand of the system of equations (2.1) by

\[
Q_{1n}^{(j)} \left( \theta^{(j)}, \lambda^{(j)} \right), \ h = 0,1,\ldots, H \quad \text{and} \quad Q_{2n}^{(j)} \left( \theta^{(j)}, \lambda^{(j)} \right), \ \text{respectively, and further we shall omit some}
\]
indexes and arguments of the functions.

Also, let \( \Psi_j = \sum \frac{\partial g^{(j)}}{\partial \theta^{(j)}} \), \( B_j = \sum g^{(j)}g^{(j)'} \), \( j = 1, 2, \ldots, M \).

**Proof of Proposition 2.1:** Similar to Wu and Zhang (2005, 2007), Zhong et al. (2000), Qin and Lawless (1994, 1995), we use the Taylor series expansions of \( 0 = Q_{1n}^{(j)}(\theta^{(j)}, \lambda_j), h = 0, 1, \ldots, H \) and \( 0 = Q_{2n}^{(j)}(\theta^{(j)}, \lambda_j) \) around \((\theta^{(j)}, 0)\), about \( \theta^{(j)} \). We have

\[
\hat{\theta}^{(j)} - \theta^{(j)} = - \left[ \sum_h \Psi_j B_j^{-1} \Psi_j' \right]^{-1} \sum_h \Psi_j B_j^{-1} \mathbf{g}(j) + o_p \left( (n_0^j)^{-\frac{1}{2}} \right), \quad j = 1, 2, \ldots, M. \tag{A.1}
\]

Also, by the central limit theorem and since \( E \left( g^{(j)} \right) = 0 \), we have

\[
\sum \mathbf{g}(j) / \sqrt{n_0^j} \xrightarrow{\mathcal{L}} \mathcal{N}_p \left( 0, k_h \mathbf{I} \right) \mathbf{E} \left( \mathbf{g}(j) \mathbf{g}(j)' \right), \tag{A.2}
\]

and by using the strong law of large numbers, we get,

\[
\Psi_j / n_0^j \xrightarrow{a.s.} k_h \mathbf{I} \mathbf{E} \left( \frac{\partial g^{(j)}}{\partial \theta^{(j)}} \right), \quad \text{and} \quad B_j / n_0^j \xrightarrow{a.s.} k_h \mathbf{I} \mathbf{E} \left( \mathbf{g}(j) \mathbf{g}(j)' \right), \quad j = 1, 2, \ldots, M. \tag{A.3}
\]

Then, combining (2.7), (A.5), (A.2) and (A.3) \( \sqrt{n_0^j} \left( \hat{\theta}^{(j)} - \theta^{(j)} \right) \xrightarrow{\mathcal{L}} \mathcal{N}_p \left( 0, V \right) \), for each \( j = 1, 2, \ldots, M \). Therefore, since \( \hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots, \hat{\theta}^{(M)} \) are independent, we get

\[
\left( \sqrt{n_0^1} \left( \hat{\theta}^{(1)} - \theta^{(1)} \right), \sqrt{n_0^2} \left( \hat{\theta}^{(2)} - \theta^{(2)} \right), \ldots, \sqrt{n_0^M} \left( \hat{\theta}^{(M)} - \theta^{(M)} \right) \right) \xrightarrow{\mathcal{L}} \mathcal{N}_{p \times M} \left( 0, V \otimes I_M \right),
\]

and since

\[
\sqrt{n} \left( \hat{\theta} - \theta \right) = \left( \sqrt{n_0^1} \left( \hat{\theta}^{(1)} - \theta^{(1)} \right), \sqrt{n_0^2} \left( \hat{\theta}^{(2)} - \theta^{(2)} \right), \ldots, \sqrt{n_0^M} \left( \hat{\theta}^{(M)} - \theta^{(M)} \right) \right) \hat{\Omega}^{rac{1}{2}},
\]

\[
\sqrt{n} \left( \hat{\theta} - \theta \right) \xrightarrow{\mathcal{L}} \mathcal{N}_{p \times M} \left( 0, V \otimes \Omega \right), \tag{A.4}
\]

that completes the proof.
Proof of Proposition 2.2: In the similar way as in Qin and Lawless (1994, 1995) and as for the proof of Proposition 2.1, by using the Taylor series expansions of the left-hand of the system of equations (2.2)-(2.3), along with some algebraic computations, we get

\[(\hat{\theta} - \theta) L = (-I + J_0 F) (\hat{\theta} - \theta) L + J_0 \delta / \sqrt{n} + o_p \left( n^{-\frac{1}{2}} \right), \tag{A.5} \]

with \( J_0 = VF'(FVF')^{-1} \). Therefore,

\[
\begin{pmatrix} \rho_n \\ \zeta_n \end{pmatrix} = \begin{pmatrix} I_p \\ -I_p + J_0 F \end{pmatrix} \rho_n + \begin{pmatrix} 0 \\ J_0 \delta \end{pmatrix} + \begin{pmatrix} o_p(1) \\ o_p(1) \end{pmatrix}. \tag{A.6} \]

Further, by using (A.4) and Slutsky Theorem, we have

\[
\rho_n \xrightarrow{\mathcal{L}} \mathcal{N}_{p \times r} \left( 0, V \otimes L' \Omega L \right), \tag{A.7} \]

Hence, combining (A.7), (A.6) and Slutsky Theorem, we get

\[
\begin{pmatrix} \rho_n \\ \zeta_n \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{Q}^* = \begin{pmatrix} I_p \\ -I_p + J_0 F \end{pmatrix} \rho + \begin{pmatrix} 0 \\ J_0 \delta \end{pmatrix}, \tag{A.8} \]

and then, combining (A.7), (A.8) and algebraic computation, we get

\[
\mathcal{Q}^* \sim \mathcal{N}_{2p \times 2r} \left( 0, \begin{pmatrix} V \otimes (L' \Omega L) & (V - \Sigma^*) \otimes (L' \Omega L) \\ (V - \Sigma^*) \otimes (L' \Omega L) & (V - \Sigma^*) \otimes (L' \Omega L) \end{pmatrix} \right), \tag{A.9} \]

that proves the first statement of the proposition. Further, we have

\[
\begin{pmatrix} \xi_n \\ \zeta_n \end{pmatrix} = \begin{pmatrix} I_p & -I_p \\ 0 & I_p \end{pmatrix} \begin{pmatrix} \rho_n \\ \zeta_n \end{pmatrix}, \tag{A.9} \]

and then, combining (A.9) and the first statement of the proposition, we prove the second statement of the proposition and that completes the proof.
Proof of Corollary 2.1: From Proposition 2.2, under the local alternative in (2.6), we have $\xi_n \xrightarrow{\mathcal{L}} \xi \sim \mathcal{N}_{p \times r} (\delta^*, \Sigma^* \otimes L'\Omega L)$, and then,

$$(L'\Omega L)^{-1/2} \xi_n (FV F')^{-1} F \xi_n (L'\Omega L)^{-1/2} \xrightarrow{\mathcal{L}} (L'\Omega L)^{-1/2} \xi' (FV F')^{-1} F \xi (L'\Omega L)^{-1/2}.$$ 

Further, one can verify that

$$\left( \Sigma^* (FV F')^{-1} F \right)^2 = \Sigma^* (FV F')^{-1} F,$$

so that

$$\text{rank} \left[ \Sigma^* (FV F')^{-1} F \right] = q, \quad \delta^* \left( \Sigma^* (FV F')^{-1} F \right)^2 = \delta^* \Sigma^* (FV F')^{-1} F,$$

and

$$\delta^* (FV F')^{-1} F \Sigma^* (FV F')^{-1} F \delta^* = \delta^* (FV F')^{-1} F \delta^*.$$

Hence,

$$(L'\Omega L)^{-1/2} \xi' L' (LV L')^{-1} L \xi (L'\Omega L)^{-1/2} \sim W_r \left( q, I_r; (L'\Omega L)^{-1/2} \delta^* L' (LV L')^{-1} L \delta^* (L'\Omega L)^{-1/2} \right).$$

Therefore,

$$\text{trace} \left( (L'\Omega L)^{-1/2} \xi' L' (LV L')^{-1} L \xi (L'\Omega L)^{-1/2} \right) \sim \chi^2_{qr} \left( \text{trace} \left( \delta^* L' (LV L')^{-1} L \delta^* (L'\Omega L)^{-1} \right) \right),$$

as stated.

Below, we establish Proposition A.1, Theorems A.1-A.2 which are useful in deriving Lemma 2.1. Let $A$ denote $M \times p$-matrix and let $\text{Vec}(A)$ denote the $np$ column vector obtained by stacking together the columns of $A$ one underneath the other.

**Proposition A.1** Let $X \sim \mathcal{N}_{q \times k} (M, \Upsilon \otimes \Lambda)$ where $\Lambda$ is a positive definite matrix, and $\Upsilon$ is a non-negative definite matrix with rank $r \leq k$. Also, let $\Xi$ be a symmetric and positive definite matrix which satisfies the two following conditions.
(i) $\Xi \Xi$ is idempotent matrix;  \quad (ii) $\Xi \Xi M = \Xi M$.

Then, for any $h$ Borel measurable and real-valued integrable function, we have

$$E\left[h\left(\text{trace}\left(\Lambda^{-1}X'\Xi\Xi X\right)\right)X\right] = E\left[h\left(\chi_{q}\right)^2 \left(\text{trace}\left(\Lambda^{-1}M'\Xi\Xi M\right)\right)\right] M.$$

\[\square\]

**Proof**  Since $\Xi \Xi \Xi \Xi$ is a symmetric idempotent matrix, there exists an orthogonal matrix $Q$ such that $Q\Xi \Xi \Xi \Xi Q' = \text{Diag} (I_r, 0)$. Moreover, let $V = Q\Xi \Xi \Xi \Xi \Lambda^{-\frac{1}{2}}$, then,

$$\text{Vec} (V) = \left( Q\Xi \Xi \Xi \Xi \otimes \Lambda^{-\frac{1}{2}} \right) \text{Vec} (X) \quad \text{and hence},$$

$$\text{Vec} (V) = \left( V_1', V_2' \right)' \sim N_{q \times k} \left( (\mu_1', 0)' \otimes \text{Diag} (I_r, 0) \otimes I_q \right), \quad (A.10)$$

$$\mu_1 = [I_{rq}, 0] E(\text{Vec} (V)) = ([I_r, 0] \otimes I_q) \left( Q\Xi \Xi \Xi \Xi \otimes \Lambda^{-\frac{1}{2}} \right) \text{Vec} (M). \quad (A.11)$$

Therefore, from (A.10), $\text{trace}\left(\Lambda^{-1}X'\Xi\Xi X\right) = \text{trace}\left(V'Q\Xi \Xi \Xi \Xi Q'V\right) = V_1'V_1$. Hence,

$$\text{Vec} \left( E\left[h\left(\text{trace}\left(\Lambda^{-1}X'\Xi\Xi X\right)\right)X\right]\right) = E\left[h\left(V_1'V_1\right)\left(\Xi \Xi \Xi \Xi \otimes \Lambda^{-\frac{1}{2}}\right)\text{Vec} (V)\right] \quad \text{and then},$$

$$\text{Vec} \left( E\left[h\left(\text{trace}\left(\Lambda^{-1}X'\Xi\Xi X\right)\right)X\right]\right) = \left( E\left[h\left(V_1'V_1\right)\right] V_1, E\left[h\left(V_1'V_1\right)V_2\right]\right). \quad (A.12)$$

From (A.10), $V_2$ is zero with probability one and hence, $E\left[h\left(V_1'V_1\right)V_2\right] = 0$. By Theorem 1 in Judge and Bock (1978), we get

$$E\left[h\left(V_1'V_1\right)V_1\right] = \mu_1 E\left(h\left(\chi_{q}\right)^2 (\mu_1'\mu_1)\right), \quad (A.13)$$

where $\mu_1$ is given by (A.11). Using (A.11),

$$\mu_1'\mu_1 = \text{trace} \left( M\Lambda^{-1}M'\Xi\Xi\Xi\Xi\Xi \right) = \text{trace} \left( M'\Xi\Xi\Xi\Xi\Xi M\Lambda^{-1} \right). \quad (A.14)$$
Further, combining (A.12) and (A.13), we have

\[
\text{Vec} \left( \mathbb{E} \left[ h \left( \text{trace} \left( \Lambda^{-1} X' \Xi \Xi X \right) \right) X \right] \right) = \mathbb{E} \left( h \left( \chi_{pq+2}^2 (\mu_1') \right) \right) \left( \Xi^{-\frac{1}{2}} Q \otimes \Lambda^{\frac{1}{2}} \right) \times \left[ \text{Diag} (I_r, 0) \otimes I_q \right] \left( Q \Xi^{\frac{1}{2}} \otimes \Lambda^{-\frac{1}{2}} \right) \text{Vec} (M),
\]

this gives \( \text{Vec} \left( \mathbb{E} \left[ h \left( \text{trace} \left( \Lambda^{-1} X' \Xi \Xi X \right) \right) X \right] \right) = \mathbb{E} \left( h \left( \chi_{qr+2}^2 (\mu_1') \right) \right) \text{Vec} (\Xi \Xi M). \) Then, since \( \Xi \) is nonsingular matrix, we have \( \Xi \Xi M = M. \) Hence,

\[
\text{Vec} \left( \mathbb{E} \left[ h \left( \text{trace} \left( \Lambda^{-1} X' \Xi \Xi X \right) \right) X \right] \right) = \text{Vec} \left( \mathbb{E} \left( h \left( \chi_{qr+2}^2 (\mu_1') \right) M \right) \right),
\]

this completes the proof.

\[\square\]

**Theorem A.1** Let \((X', Y')' \sim \mathcal{N}_{2q \times 2k} \left( (M_1, M_2), \begin{pmatrix} \Xi_{11} \otimes \Lambda_{11} & 0 \\ 0 & \Xi_{22} \otimes \Lambda_{22} \end{pmatrix} \right) \)

where \( \Lambda_{11} \) is a positive definite matrix, and \( \Xi_{11}, \Lambda_{22} \) are nonnegative definite matrices with rank \( r \leq k. \) Also, let \( \Xi \) be a symmetric and positive definite matrix which satisfies the two following conditions.

(i) \( \Xi_{11} \Xi \) is idempotent matrix;  
(ii) \( \Xi \Xi_{11} \Xi M_1 = \Xi M_1. \) Then, for any \( h \) Borel measurable and real-valued integrable function, and any nonnegative definite matrix \( A, \) we have

\[
\mathbb{E} \left[ h \left( \text{trace} \left( \Lambda_{11}^{-1} X' \Xi \Xi Y_{11} \Xi X \right) \right) Y' A X \right] = \mathbb{E} \left( h \left( \chi_{qr+2}^2 \left( \text{trace} \left( \Lambda_{11}^{-1} M_1' \Xi Y_{11} \Xi M_1 \right) \right) \right) \right) M_2' A M_1.
\]

\[\square\]

**Proof** Since \( X \) and \( Y \) are independent, we have

\[
\mathbb{E} \left[ h \left( \text{trace} \left( \Lambda_{11}^{-1} X' \Xi Y_{11} \Xi X \right) \right) Y' A X \right] = (\mathbb{E} (Y))' A \mathbb{E} \left[ h \left( \text{trace} \left( \Lambda_{11}^{-1} X' \Xi Y_{11} \Xi X \right) \right) X \right],
\]

and then, the rest of the proof follows directly from Proposition A.1 given above.

\[\square\]
Theorem A.2 Let \( X \sim \mathcal{N}_{q \times k}(M, \Upsilon \otimes \Lambda) \) where \( \Lambda \) is a positive definite matrix, and \( \Upsilon \) is a nonnegative definite matrix with rank \( r \leq k \). Also, let \( A \) and \( \Xi \) be positive definite symmetric matrices and assume that \( \Xi \) satisfies the two following conditions.

(i) \( \Upsilon \Xi \) is idempotent matrix;  
(ii) \( \Xi \Upsilon \Xi M = \Xi M \).

Then, for any \( h \) Borel measurable and real-valued integrable function, we have

\[
\mathbb{E} \left[ h \left( \text{trace} \left( \Lambda^{-1} X' \Upsilon \Xi X \right) \right) \right] = \mathbb{E} \left[ h \left( \chi_{qr}^2 \left( \text{trace} \left( \Lambda^{-1} M' \Upsilon \Xi M \right) \right) \right) \right] \times \text{trace} (A \Upsilon) \text{trace} (\Lambda) + \mathbb{E} \left[ h \left( \chi_{qr+2}^2 \left( \text{trace} \left( \Lambda^{-1} M' \Upsilon \Xi M \right) \right) \right) \right] \text{trace} (M' A M). 
\]

\[\square\]

Proof As given in proof of Proposition A.1, we have \( h \left( \text{trace} \left( \Lambda^{-1} X' \Upsilon \Xi X \right) \right) = h(V_1' V_1), \) with \( V = Q\Xi^{-\frac{1}{2}} X \Lambda^{-\frac{1}{2}} \). We have \( X = \Xi^{-\frac{1}{2}} Q' V \Lambda^{\frac{1}{2}} \) and hence,

\[ X' A X = \Lambda^{\frac{1}{2}} V' Q \Xi^{-\frac{1}{2}} A \Xi^{-\frac{1}{2}} Q' V \Lambda^{\frac{1}{2}}, \]

and then,

\[ \mathbb{E} \left[ h \left( \text{trace} \left( \Lambda^{-1} X' \Upsilon \Xi X \right) \right) \right] \text{trace} (X' A X) = \mathbb{E} \left[ h(V_1' V_1) \text{trace} \left( \Lambda^{\frac{1}{2}} V' Q \Xi^{-\frac{1}{2}} A \Xi^{-\frac{1}{2}} Q' V \Lambda^{\frac{1}{2}} \right) \right]. \]

This leads to,

\[ \mathbb{E} \left[ h \left( \text{trace} \left( \Lambda^{-1} X' \Upsilon \Xi X \right) \right) \right] \text{trace} (X' A X) = \mathbb{E} \left[ h(V_1' V_1) \text{trace} \left( Q \Xi^{-\frac{1}{2}} A \Xi^{-\frac{1}{2}} Q' V \Lambda V' \right) \right]. \]

Also,

\[ \text{trace} \left( Q \Xi^{-\frac{1}{2}} A \Xi^{-\frac{1}{2}} Q' V \Lambda V' \right) = (\text{Vec}(V))' \left( Q \Xi^{-\frac{1}{2}} A \Xi^{-\frac{1}{2}} Q' \otimes \Lambda \right) \text{Vec}(V), \]

and

\[ 
\left( Q \Xi^{-\frac{1}{2}} A \Xi^{-\frac{1}{2}} Q \otimes \Lambda \right) = G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}. \tag{A.15} 
\]

Hence, since \( \text{Vec}(V) = (V_1', 0)' \), we have \( (\text{Vec}(V))' \left( \Lambda \otimes Q \Xi^{-\frac{1}{2}} A \Xi^{-\frac{1}{2}} Q \right) \text{Vec}(V) = V_1' G_{11} V_1 \).

Therefore, \( \mathbb{E} \left[ h(V_1' V_1) \left( \text{Vec}(V) \right)' \left( \Lambda \otimes Q \Xi^{-\frac{1}{2}} A \Xi^{-\frac{1}{2}} Q \right) \text{Vec}(V) \right] = \mathbb{E} [h(V_1' V_1) V_1' G_{11} V_1], \) and ap-
plying Theorem 2 in Judge and Bock (1978), we get
\[
E \left[ h(V_1'V_1) V_1'G_{11}V_1 \right] = E \left[ h \left( \chi_{q r+2}^2 \left( \text{trace} \left( \mu_1' \mu_1 \right) \right) \right) \right] \text{trace} (G_{11})
+ E \left[ h \left( \chi_{q r+4}^2 \left( \text{trace} \left( \mu_1' \mu_1 \right) \right) \right) \right] \left( \mu_1'G_{11}\mu_1 \right).
\]
(A.16)

where \( \mu_1 \) is given by (A.11). Further, since
\[
G_{11} = ([I_r, 0] \otimes I_q) G \left[ \text{Diag} (I_r, 0) \otimes I_q \right] \quad \text{(A.17)}
\]
by using (A.17) and (A.11), we have \( \mu_1'G_{11}\mu_1 = \text{trace} (MM'\Xi\Upsilon A) \). Further, under Assumption (ii), \( M'\Xi\Upsilon = M' \), and then
\[
\mu_1'G_{11}\mu_1 = \text{trace} (MM'A) = \text{trace} (M'AM). \quad \text{(A.18)}
\]

Now, by combining (A.15) and (A.17), we obtain
\[
\text{trace} (G_{11}) = \text{trace} \left[ \left( Q \Xi^{-\frac{1}{2}} A \Xi^{-\frac{1}{2}} Q \otimes A \right) \left( Q \Xi^{-\frac{1}{2}} \Upsilon \Xi^{\frac{1}{2}} Q' \otimes I_q \right) \right] = \text{trace} (A\Upsilon) \text{trace} (A). \quad \text{(A.19)}
\]

Finally, the proof is completed by combining (A.14), (A.16), (A.18), and (A.19).

\[ \square \]

**Proof of Lemma 2.1** We have
\[
E \left( \mathcal{L} \left( \tilde{\theta}(h), \theta; W \right) \right) = E \left\{ \text{trace} \left[ \left( \eta + h \left( \| \xi \|_{\Xi_1, \Xi_2}^2 \right) \right) \xi' W \left( \eta + h \left( \| \xi \|_{\Xi_1, \Xi_2}^2 \right) \xi \right] \right\}
\]
and since \( h \) is a real-valued function, we get
\[
E \left( \mathcal{L} \left( \tilde{\theta}(h), \theta; W \right) \right) = E \{ \text{trace} [\eta'W\eta] \} + 2E \{ h \left( \| \xi \|_{\Xi_1, \Xi_2}^2 \right) \text{trace} [\xi'W\eta] \}
+ E \{ h^2 \left( \| \xi \|_{\Xi_1, \Xi_2}^2 \right) \text{trace} [\xi'W\xi] \}.
\]
Then, combining Theorem A.1, Theorem A.2 and Proposition 2.2, we get

$$E\{\text{trace} [\eta' W \eta]\} = \text{trace} (W (V - \Sigma^*)) \text{trace} (\Xi_2^{-1}) + \text{trace} (\delta^* W \delta^*),$$

$$E\left\{ h \left( \|\xi\|_{\Xi_1^2, \Xi_2^2} \right) \text{trace} [\xi' W \eta] \right\} = -E \left[ h \left( \chi^2_{qr+2} \left( \text{trace} \left( \Xi_2 \delta^{*'} \Xi_1 \delta^* \right) \right) \right] \text{trace} (\delta^* W \delta^*),$$

$$E\left\{ h^2 \left( \|\xi\|^2_{\Xi_1^2, \Xi_2^2} \right) \text{trace} [\xi' W \xi] \right\} = E \left[ h^2 \left( \chi^2_{qr+4} \left( \text{trace} \left( \Xi_2 \delta^{*'} \Xi_1 \delta^* \right) \right) \right] \text{trace} (\delta^* W \delta^*)$$

$$+ E \left[ h^2 \left( \chi^2_{qr+2} \left( \text{trace} \left( \Xi_2 \delta^{*'} \Xi_1 \delta^* \right) \right) \right] \text{trace} (W \Sigma^*) \text{trace} (\Xi_2^{-1}),$$

this completes the proof.

□

References


Figure 1: Relative efficiency of SMELE, PSMELE and RMELE with respect to UMELE.