ESAIM: Probability and Statistics www.esaim-ps.org

ESAIM: PS 16 (2012) 139–150 DOI: 10.1051/ps/2010015

SHRINKAGE STRATEGIES IN SOME MULTIPLE MULTI-FACTOR DYNAMICAL SYSTEMS

SÉVÉRIEN NKURUNZIZA¹

Abstract. In this paper, we are interested in estimation problem for the drift parameters matrices of m independent multivariate diffusion processes. More specifically, we consider the case where the m-parameters matrices are supposed to satisfy some uncertain constraints. Given such an uncertainty, we develop shrinkage estimators which improve over the performance of the maximum likelihood estimator (MLE). Under an asymptotic distributional quadratic risk criterion, we study the relative dominance of the established estimators. Further, we carry out simulation studies for observation periods of small and moderate lengths of time that corroborate the theoretical finding for which shrinkage estimators outperform over the MLE. The proposed method is useful in model assessment and variable selection.

Mathematics Subject Classification. 62M05, 58J65.

Received July 10, 2009. Revised February 27, 2010.

1. Introduction

The diffusion processes play a crucial role in modeling diverse phenomenon in different scientific fields such as quantitative finance [10], economics [3], biomedical sciences [7], physics [15] ecology [6]. In this paper, we consider m independent p-vectors of the parametric diffusion processes $\left\{\left(X_k^{(i)}(t), k=1,2,\ldots,p\right); 0 \leqslant t \leqslant T\right\}$, $i=1,2,\ldots,m$ which are solutions of multi-factors stochastic differential equation (SDE)

$$dX_k^{(i)}(t) = \sum_{j=1}^p \theta_{kj}^{(i)} V_j \left(\mathbf{X}^{(i)}(t) \right) dt + \sigma_k^{(i)} \left(\mathbf{X}^{(i)}(t) \right) dW_k^{(i)}(t), \quad 0 \le t \le T,$$
(1.1)

with for each $i=1,2,\ldots,m$, $\boldsymbol{X}^{(i)}(t)=\left(X_k^{(i)}(t),k=1,2,\ldots,p\right)$, and $\left\{W_k^{(i)}(t),0\leqslant t\leqslant T\right\}$, $k=1,2,\ldots,p$ are p-independent Wiener processes, and $\sigma_k^{(i)}\left(\boldsymbol{X}^{(i)}(t)\right)>0$, $k=1,2,\ldots,p$ and the real-valued functions $V_k\left(\boldsymbol{X}^{(i)}(t)\right)$, $k=1,2,\ldots,p$ are such that the processes in (1.1) exist. Thereafter, we denote

$$V\left(X^{(i)}(t)\right) = \left(V_1\left(X^{(i)}(t)\right), V_2\left(X^{(i)}(t)\right), \dots, V_p\left(X^{(i)}(t)\right)\right)',$$

Keywords and phrases. Asymptotic distributional risk, diffusion process, MLE, Shrinkage estimator, Wiener process.

¹ University of Windsor, 401 Sunset Avenue, Windsor, N9B 3P4, Ontario, Canada. severien@uwindsor.ca

and in the sequel, we assume that the functions $\sigma_k^{(i)}\left(\boldsymbol{X}^{(i)}(t)\right)$ and $V_k\left(\boldsymbol{X}^{(i)}(t)\right)$ are such that the process in (1.1) is ergodic (see for example [11]). To give an example, note that if $\sigma_k^{(i)}\left(\boldsymbol{X}^{(i)}(t)\right) = \sigma_k^{(i)} > 0$ and $V_k\left(\boldsymbol{X}^{(i)}(t)\right) = -\boldsymbol{X}_k^{(i)}$, $k=1,2,\ldots,p,\ i=1,2,\ldots,m$, the process in (1.1) is ergodic if and only if $\boldsymbol{\theta}^{(i)}=\left(\theta_{kj}^{(i)}\right)_{1\leqslant k,j\leqslant p}$ is a positive definite matrix, $i=1,2,\ldots,m$. In this paper, we are interested in estimating the parameters matrices $\boldsymbol{\theta}^{(i)}\in\boldsymbol{\Theta},\ i=1,2,\ldots,m$ when $\boldsymbol{\theta}^{(i)}$ may be subjected to some uncertainties. For the simplicity sake, assume that $\sigma_k^{(i)}\left(\boldsymbol{X}^{(i)}(t)\right)=\sigma_k^{(i)}>0,\ k=1,2,\ldots,p,\ i=1,2,\ldots,m$. Also, to simplify notation, let $\boldsymbol{\Sigma}^{(i)}$ be diagonal matrix whose diagonal entrees are $\sigma_{1i}^2,\sigma_{2i}^2,\ldots,\sigma_{pi}^2,\ i.e.,\ \boldsymbol{\Sigma}_i=\mathrm{diag}\left(\sigma_{1i}^2,\sigma_{2i}^2,\ldots,\sigma_{pi}^2\right)$, and let $\boldsymbol{W}^{(i)}(t)=\left(W_1^{(i)}(t),W_2^{(i)}(t),\ldots,W_p^{(i)}(t)\right)'$, for $0\leqslant t\leqslant T$ and for each $i=1,2,\ldots,m$. Then, from relation (1.1), we have

$$dX^{(i)}(t) = \theta^{(i)}V(X^{(i)}(t)) dt + \Sigma_i^{\frac{1}{2}} dW^{(i)}(t), \quad 0 \le t \le T, i = 1, 2, \dots, m.$$
(1.2)

Further, the parameters matrices $\theta^{(i)}$ are assumed to lie or not in a restricted hyper-plan whose rank is q < p. For instance, suppose that

$$L_1^* \theta^{(1)} L_2^* = L_1^* \theta^{(2)} L_2^* = \dots = L_1^* \theta^{(m)} L_2^*, \tag{1.3}$$

where L_1^* is a given $q \times p$ -matrix of full rank with q < p, L_2^* is known $p \times r$ matrix. It is noticed that the constraint in (1.3) extends that in [8]. Thus, as in classical multivariate regression, the constraint in (1.3) is useful for example in model assessment and variable selection, and in profile analysis. Also, the constraint in (1.3) is useful in financial modeling where, for instance, different groups of countries decide to unify their economic policies, as is the case for the European Union countries. Thus, within each group of the united countries, the economic policy is supposed to has been harmonized and thus, one would expect homogeneity of the parameters of the process under consideration. Further, because of the globalization of the economy, different groups (or unions) need to negotiate on the international transactions rules and on some regulatory financial-economic rules. In this context, it is reasonable to assume that the parameter matrices $\theta^{(i)}$, $i = 1, 2, \ldots, m$ satisfy the constraint as in (1.3). Another interesting example is related to the current worldwide financial crisis where multinational institutions may be forced to form the same chain group in order to survive.

To simplify the notation, let $\boldsymbol{\theta} = \left(\boldsymbol{\theta}^{(1)'}, \boldsymbol{\theta}^{(2)'}, \dots, \boldsymbol{\theta}^{(m)'}\right)'$ and let \boldsymbol{I}_m denote the identity matrix with rank m. Then, the constraint in (1.3) can be rewritten as

$$(\boldsymbol{L}_{3}^{*} \otimes \boldsymbol{L}_{1}^{*}) \boldsymbol{\theta} \boldsymbol{L}_{2}^{*} = \boldsymbol{0}, \tag{1.4}$$

where L_3^* is $(m-1) \times m$ -matrix given by

$$L_3^* = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}. \tag{1.5}$$

Accordingly, one considers the following constraint that is more general than (1.4),

$$L_1 \theta L_2 = d \tag{1.6}$$

where L_1 is a given $q \times mp$ -matrix of full rank with q < mp, and L_2 and d are known $mp \times r$ and $q \times r$ -matrix of full rank with r < mp. It should be noticed that the statistical model in (1.2) with the constraint (1.3) is an extension of the model considered in [14] where V(X(t)) = X(t) and m = r = 1. An appropriate estimation

method for $\left(\boldsymbol{\theta}^{(1)'}, \boldsymbol{\theta}^{(2)'}, \dots, \boldsymbol{\theta}^{(m)'}\right)'$ needs to incorporate the constraint in (1.3). In particular, a test statistic is used to take care of the constraint (1.6).

The rest of this paper is organized as follows. Section 1 presents the restricted and unrestricted MLEs of θ as well as the asymptotic normality of these estimators. In Section 2, we present the shrinkage estimator and show that it dominates the MLE. Section 3 presents some simulation results, and Section 4 gives concluding remarks. Finally, technical results are given in the Appendix A.

2. The maximum likelihood estimator and asymptotic normality

In this section, we present the unrestricted maximum likelihood estimator (UMLE) and the restricted maximum likelihood estimator (RMLE) for the parameter matrix $\boldsymbol{\theta}$. Also, we present the results on the asymptotic properties of the UMLE and RMLE, which are used in deriving the test statistic for the constraint (1.6). As presented in Section 2, the UMLE and RMLE are combined in order to form a class of estimators which improve the performance of UMLE and RMLE. For the sake of simplicity, we assume that the distribution of the matrix-initial value $\boldsymbol{X}_0 = \left(\boldsymbol{X}_0^{(1)'}, \boldsymbol{X}_0^{(2)'}, \ldots, \boldsymbol{X}_0^{(m)'}\right)'$ does not depend on $\boldsymbol{\theta}$. Indeed, for the case where the distribution of \boldsymbol{X}_0 depends of $\boldsymbol{\theta}$, the obtained results are similar to that established conditionally to \boldsymbol{X}_0 . Let $\boldsymbol{X}(t) = \left(\boldsymbol{X}^{(1)'}(t), \boldsymbol{X}^{(2)'}(t), \ldots, \boldsymbol{X}^{(m)'}(t)\right)'$, $\boldsymbol{V}(\boldsymbol{X}(t)) = \left(\boldsymbol{V}\left(\boldsymbol{X}^{(1)}(t)\right), \boldsymbol{V}\left(\boldsymbol{X}^{(2)}(t)\right), \ldots, \boldsymbol{V}\left(\boldsymbol{X}^{(m)}(t)\right)\right)'$, and let

$$U_{i}(T) = \int_{0}^{T} dX^{(i)}(t)V'\left(X^{(i)}(t)\right), \quad D_{i}(T) = \int_{0}^{T} V\left(X^{(i)}(t)\right)V'\left(X^{(i)}(t)\right) dt, \quad i = 1, 2, \dots, m.$$
 (2.1)

Also, let

$$U_T = (U_1'(T), U_2'(T), \dots, U_m'(T))', \text{ and let } D_T = \text{diag}(D_1(T), D_2(T), \dots, D_m(T)).$$
 (2.2)

Further, let $\Sigma = \operatorname{diag}(\Sigma_1, \Sigma_2, \dots, \Sigma_m)$ and let $J = \Sigma L'_1 (L_1 \Sigma L'_1)^{-1}$. Moreover, let $\widehat{\theta}$ be the UMLE of θ , and let $\widetilde{\theta}$ be the RMLE of θ . It is noticed that, for the one-parameter case, the UMLE $\widehat{\theta}$ corresponds to that given in [12] (Chap. 17, pp. 206–207) or in [11] (p. 63). This paper gives an extension of the one-parameter inference problem to matrix-parameter inference problem. In addition, we generalize the inference strategies in [14].

The following proposition gives the UMLE $\hat{\theta}$ and the RMLE $\hat{\theta}$. The proof follows from standard stochastic calculus techniques. For the convenience of the reader, we outline it below. It is noticed that since the subsequent results are derived under continuous sampling plan, Σ_i , i = 1, 2, ..., m are assumed to be known. Indeed, it suffices to take the corresponding quadratic variation.

Proposition 2.1. We have
$$\widehat{\boldsymbol{\theta}} = \boldsymbol{U}_T \boldsymbol{D}_T^{-1}$$
, and $\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} - \boldsymbol{J} \left(\boldsymbol{L}_1 \widehat{\boldsymbol{\theta}} \boldsymbol{L}_2 - \boldsymbol{d} \right) \left(\boldsymbol{L}_2' \boldsymbol{D}_T^{-1} \boldsymbol{L}_2 \right)^{-1} \boldsymbol{L}_2' \boldsymbol{D}_T^{-1}$.

Proof. The log-likelihood function is given by $l(\boldsymbol{\theta}) = \operatorname{trace} \left(\boldsymbol{\theta}' \Sigma^{-1} \boldsymbol{U}_{T}\right) - \frac{1}{2} \operatorname{trace} \left(\boldsymbol{\theta}' \Sigma^{-1} \boldsymbol{\theta} \boldsymbol{D}_{T}\right)$, and then, by algebraic computations, we get the first statement of the proposition. Further, by the Lagrangian method, the RMLE $\tilde{\boldsymbol{\theta}}$ satisfies the system of equations $\Sigma^{-1}\boldsymbol{U}_{T} - \Sigma^{-1}\boldsymbol{D}_{T}\tilde{\boldsymbol{\theta}} + \boldsymbol{L}'_{1}\boldsymbol{\lambda}\boldsymbol{L}'_{2} = \boldsymbol{0}$, $\left(\boldsymbol{L}_{1}\tilde{\boldsymbol{\theta}}\boldsymbol{L}_{2} - \boldsymbol{d}\right) = 0$, where the matrices \boldsymbol{L}_{1} , \boldsymbol{L}_{2} and \boldsymbol{d} are given in (1.6). The rest of proof follows from algebraic computations.

Note that the MLEs given in Proposition 2.1 presuppose that all sample paths of the process in (1.2) are observed and that this sampling plan has a major impact on the diffusion term. Indeed, while in continuous time it is possible to estimate it without error and therefore it makes perfect sense to assume it known, in discrete time this is certainly not the case. Also, we assume that the drift coefficient is such that, almost surely,

$$\lim_{T \to \infty} \|\boldsymbol{D}_T\| = \infty, \quad \lim_{T \to \infty} T^{-1} \boldsymbol{D}_T = \mathbb{E}\left(\boldsymbol{X}_0 \boldsymbol{X}_0'\right) = \boldsymbol{\Sigma}_0, \tag{2.3}$$

with Σ_0 a positive definite matrix. Note that the second assumption in relation (2.3) holds for ergodic processes. Under some conditions, the following proposition shows that the UMLE and RMLE are strongly consistent.

Proposition 2.2. Assume that the model (1.2) holds along with relations (1.6) and (2.3). Then,

$$\widehat{\boldsymbol{\theta}} \xrightarrow[T \to \infty]{a.s.} \boldsymbol{\theta}, \quad and \quad \widetilde{\boldsymbol{\theta}} \xrightarrow[T \to \infty]{a.s.} \boldsymbol{\theta}.$$
 (2.4)

Proof. From equations (1.2), we have

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \boldsymbol{\Sigma}^{\frac{1}{2}} \int_0^T d\boldsymbol{W}(t) \, \boldsymbol{V}'(\boldsymbol{X}(t)) \, \boldsymbol{D}_T^{-1}, \tag{2.5}$$

and by the martingale Strong Law of Large Numbers, we get the first statement of the proposition. Further,

$$\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta} = (\boldsymbol{I}_{pm} - \boldsymbol{J}\boldsymbol{L}_1) \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) - \boldsymbol{J} \left(\boldsymbol{L}_1 \boldsymbol{\theta} \boldsymbol{L}_2 - \boldsymbol{d}\right),$$
 (2.6)

and then, under (1.6), we get $\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta} = (\boldsymbol{I}_{pm} - \boldsymbol{J}\boldsymbol{L}_1) \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)$. Therefore, from the first statement of the proposition, we get $\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \xrightarrow[T \to \infty]{a.s.} \boldsymbol{0}$, and that completes the proof.

Under Proposition 2.2, the following corollary shows that UMLE is asymptotically normal.

Corollary 2.3. Assume that the hypotheses of Proposition 2.2 hold. Then,

$$\sqrt{T} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \xrightarrow[T \to \infty]{\mathcal{L}} \mathcal{N}_{mp \times mp} \left(\mathbf{0}, \ \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}_{0}^{-1} \right) \quad and \quad \sqrt{T} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \boldsymbol{L}_{2} \xrightarrow[T \to \infty]{\mathcal{L}} \mathcal{N}_{mp \times r} \left(\mathbf{0}, \ \boldsymbol{\Sigma} \otimes \boldsymbol{L}_{2}' \boldsymbol{\Sigma}_{0}^{-1} \boldsymbol{L}_{2} \right).$$

Proof. The first statement follows directly from equation (2.5), and by applying the martingale central limit Theorem. Further, combing the first statement and Slutsky's theorem, we get the second statement of the proposition, and that completes the proof.

From Corollary 2.3, we establish the joint asymptotic normality of UMLE and RMLE under the following class of alternative constraints

$$K_T: \mathbf{L}_1 \boldsymbol{\theta} \mathbf{L}_2 = \mathbf{d} + \frac{\boldsymbol{\delta}}{\sqrt{T}}, \quad T > 0$$
(2.7)

where $\boldsymbol{\delta}$ is a nonzero $q \times r$ -matrix no linearly dependent with \boldsymbol{d} . Also, we assume that $\|\boldsymbol{\delta}\| < \infty$, and let $\boldsymbol{\varrho}_T = \sqrt{T} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \boldsymbol{L}_2$, $\boldsymbol{\xi}_T = \sqrt{T} \left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}} \right) \boldsymbol{L}_2$. Further, let $\boldsymbol{\Sigma}^* = \boldsymbol{J} \boldsymbol{L}_1 \boldsymbol{\Sigma}$. In the sequel, we simplify the computations by assuming that the initial value is chosen such that $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}$.

Proposition 2.4. If the hypotheses of Proposition 2.2 and the constraint in (2.7) hold, then

$$\left(egin{array}{c} oldsymbol{arrho}_T \ oldsymbol{\xi}_T \end{array}
ight) \xrightarrow{\mathcal{L}} \left(egin{array}{c} oldsymbol{arrho} \ oldsymbol{\xi} \end{array}
ight) \sim \mathcal{N}_{2mp imes 2r} \left(\left(egin{array}{c} oldsymbol{0} \ oldsymbol{J} oldsymbol{\delta} \end{array}
ight), \left(egin{array}{c} oldsymbol{\Sigma} \otimes \left(oldsymbol{L}_2'oldsymbol{\Sigma}^{-1}oldsymbol{L}_2
ight) \ oldsymbol{\Sigma}^* \otimes \left(oldsymbol{L}_2'oldsymbol{\Sigma}^{-1}oldsymbol{L}_2
ight) \end{array}
ight).$$

Furthermore,

$$egin{pmatrix} \left(egin{aligned} oldsymbol{\zeta}_T \ oldsymbol{\xi}_T \end{matrix}
ight) & \stackrel{\mathcal{L}}{\longrightarrow} \left(egin{aligned} oldsymbol{\zeta} \ oldsymbol{\xi} \end{matrix}
ight) \sim \mathcal{N}_{2mp imes 2r} \left(\left(egin{aligned} -J oldsymbol{\delta} \ J oldsymbol{\delta} \end{matrix}
ight), \left(egin{aligned} oldsymbol{\Sigma} \otimes \left(oldsymbol{L}_2' oldsymbol{\Sigma}^{-1} oldsymbol{L}_2
ight) - oldsymbol{\Sigma}^* \otimes \left(oldsymbol{L}_2' oldsymbol{\Sigma}^{-1} oldsymbol{L}_2
ight) \\ 0 & oldsymbol{\Sigma}^* \otimes \left(oldsymbol{L}_2' oldsymbol{\Sigma}^{-1} oldsymbol{L}_2
ight) \end{array}
ight).$$

The proof of this proposition is outlined in the Appendix. Note that, under some regularities conditions, as given for example in [11] (p. 212), the convergence in law given in Corollary 2.3 and Proposition 2.4 hold uniformly in θ . Also, from Proposition 2.4, we derive the following corollary that is useful in establishing the test statistic. We denote $W_r(n, \Sigma, \Upsilon)$ a $r \times r$ -random matrix Wishart variate with degrees of freedom n, and parameter Σ , and non-centrality parameter matrix Υ . Also, let $\chi_n^2(\delta)$ denote a random chi-square variate with n degrees of freedom, non-centrality parameter δ .

Corollary 2.5. Let $\Xi = L_1' \left(L_1 \Sigma^{-1} L_1' \right)^{-1} L_1$. If the hypotheses of Proposition 2.4 hold, then

$$\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)^{-\frac{1}{2}}\boldsymbol{\xi}_{T}^{\prime}\boldsymbol{\Xi}\boldsymbol{\xi}_{T}\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)^{-\frac{1}{2}}\xrightarrow[T\rightarrow\infty]{\mathcal{L}}W_{r}\left(q,\boldsymbol{I}_{r},\boldsymbol{\delta}^{*^{\prime}}\boldsymbol{\Xi}\boldsymbol{\delta}^{*}\right),$$

and trace $\left[\boldsymbol{\xi}_{T}^{\prime}\boldsymbol{\Xi}\boldsymbol{\xi}_{T}\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)^{-1}\right] \xrightarrow[T \to \infty]{\mathcal{L}} \chi_{qr}^{2}\left(\operatorname{trace}\left(\boldsymbol{\delta}^{*^{\prime}}\boldsymbol{\Xi}\boldsymbol{\delta}^{*}\right)\right).$

The proof is outlined in the Appendix A. Based on Corollary 2.5, we propose the following test statistic for taking care of uncertain prior information given by the constraint in (1.6)

$$\psi(T) = \operatorname{trace}\left[\boldsymbol{\xi}_T' \boldsymbol{\Xi} \boldsymbol{\xi}_T \left(\boldsymbol{L}_2' \boldsymbol{\Sigma}^{-1} \boldsymbol{L}_2\right)^{-1}\right]. \tag{2.8}$$

3. Shrinkage estimation strategy (SES)

In this subsection, we present a class of estimators that combine both UMLE and RMLE. Such estimators are so-called "shrinkage estimators". For more details, we refer to Judge and Bock [9], Ahmed and Saleh [2], Nkurunziza and Ahmed [14] among others. In particular, we consider the following class of estimators $\left\{\widetilde{\boldsymbol{\theta}} + \{1-c\psi(T)^{-1}\}(\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}}) : c \in [0,2(qr-2)), qr>2\right\}$, where $\psi(T)$ is the test statistic given in (2.8). As often the case in classical shrinkage strategy, we consider the shrinkage estimator (SE) that corresponds to c=qr-2. Namely, SE $\widehat{\boldsymbol{\theta}}^S$ of $\boldsymbol{\theta}$ is defined as $\widehat{\boldsymbol{\theta}}^S=\widetilde{\boldsymbol{\theta}}+\{1-(qr-2)/\psi(T)\}\left(\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}}\right)$. Since qr>2, and since $\psi>0$ with probability one, we have $\psi<(qr-2)$ if and only if $1-(qr-2)\psi^{-1}<0$. This causes a possible over-shrinking, and thus, following Ahmed [2], the shrinkage estimator should not be used as an estimator in its own right, but as a tool for developing the positive-rule shrinkage estimators (PSE), $\widehat{\boldsymbol{\theta}}^{S+}$ that is defined as

$$\widehat{\boldsymbol{\theta}}^{S+} = \widetilde{\boldsymbol{\theta}} + \max\{\left(1 - (qr - 2)\psi^{-1}\right), 0\} \left(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}\right). \tag{3.1}$$

In the following subsection, we present the asymptotic distributional risk (ADR) and asymptotic distributional bias (ADB). For more details about the concepts of ADR and ADB, we refer to Ahmed and Saleh [2]. For the sake of simplicity, we assume that the model in (1.2) satisfies the similar regularities conditions as given in [11] (pp. 212–213).

3.1. Asymptotic distributional risk and asymptotic distributional Bias

As often observed in shrinkage strategy, the effective domain of risk dominance of PSE or SE over MLE is a small neighborhood of the constraint (pivot), that is here $L_1\theta L_2 - d = 0$. Also, as the observation period T increases, this domain becomes narrower. This justifies the choice of the class of local alternatives given in (2.7), and for the optimality criterion, we consider the *quadratic loss function* of the form

$$L\left(\widehat{\boldsymbol{\theta}}^{\star}, \boldsymbol{\theta}; \mathbf{W}\right) = T \times \operatorname{trace}\left[\boldsymbol{L}_{2}^{\prime}\left(\widehat{\boldsymbol{\theta}}^{\star} - \boldsymbol{\theta}\right)^{\prime} \mathbf{W}\left(\widehat{\boldsymbol{\theta}}^{\star} - \boldsymbol{\theta}\right) \boldsymbol{L}_{2}\right], \tag{3.2}$$

where **W** is a nonnegative definite matrix. Using the distribution of $\sqrt{T} \left(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta} \right) \boldsymbol{L}_2$ and taking the expected value of both sides of (3.2), we get the expected loss, so-called the *quadratic risk* and this will be

denoted by $R_T^o\left(\widehat{\boldsymbol{\theta}}^\star,\boldsymbol{\theta};\mathbf{W}\right)$. In particular, if $\widehat{\boldsymbol{\Sigma}}_1(T)\otimes\widehat{\boldsymbol{\Sigma}}_2(T)=T$ $\mathbf{E}\left[\operatorname{vec}\left(\left(\widehat{\boldsymbol{\theta}}^\star-\boldsymbol{\theta}\right)\boldsymbol{L}_2\right)\left(\operatorname{vec}\left(\left(\widehat{\boldsymbol{\theta}}^\star-\boldsymbol{\theta}\right)\boldsymbol{L}_2\right)\right)^T\right]$, with $\widehat{\boldsymbol{\Sigma}}_1(T)$ nonsingular matrix, we get $R_T^o\left(\widehat{\boldsymbol{\theta}}^\star,\boldsymbol{\theta};\mathbf{W}\right)=\operatorname{trace}\left(\mathbf{W}\widehat{\boldsymbol{\Sigma}}_2(T)\right)\operatorname{trace}\left(\widehat{\boldsymbol{\Sigma}}_1(T)\right)$. Thus, whenever $\lim_{T\to\infty}\widehat{\boldsymbol{\Sigma}}_1(T)=\boldsymbol{\Sigma}_{01}$ and $\lim_{T\to\infty}\widehat{\boldsymbol{\Sigma}}_2(T)=\boldsymbol{\Sigma}_{02}$ exists, $R_T^o\left(\widehat{\boldsymbol{\theta}}^\star,\boldsymbol{\theta};\mathbf{W}\right)\xrightarrow[T\to\infty]{}R^o\left(\widehat{\boldsymbol{\theta}}^\star,\boldsymbol{\theta};\mathbf{W}\right)=\operatorname{trace}\left(\mathbf{W}\boldsymbol{\Sigma}_{02}\right)$ trace $(\boldsymbol{\Sigma}_{01})$. Let $\widetilde{G}_T(\mathbf{u})$ denote the cumulative distribution function (cdf) of $\sqrt{T}\left(\widehat{\boldsymbol{\theta}}^\star-\boldsymbol{\theta}\right)\boldsymbol{L}_2$ and let \widetilde{G} be the cdf of a matrix variate and square-integrable $\boldsymbol{\rho}^*$. Also, suppose that $\sqrt{T}\left(\widehat{\boldsymbol{\theta}}^\star-\boldsymbol{\theta}\right)\boldsymbol{L}_2\xrightarrow[T\to\infty]{}\boldsymbol{\rho}^*$, i.e. $\widetilde{G}_T\xrightarrow[T\to\infty]{}\widetilde{G}$ (at all points of continuity of \widetilde{G}), and let $\boldsymbol{\Sigma}_{\widetilde{G}}^{(1)}\otimes\boldsymbol{\Sigma}_{\widetilde{G}}^{(2)}=\mathrm{E}\left(\boldsymbol{\rho}\boldsymbol{\rho}^{\star'}\right)$ with $\boldsymbol{\Sigma}_{\widetilde{G}}^{(1)}$ and $\boldsymbol{\Sigma}_{\widetilde{G}}^{(2)}$ respectively $k\times k$ and $q\times q$ -matrices. From Fatou's Lemma, we have $\mathrm{trace}\left(\mathbf{W}\boldsymbol{\Sigma}_{\widetilde{G}}^{(2)}\right)$ trace $\left(\boldsymbol{\Sigma}_{\widetilde{G}}^{(1)}\right)\leqslant R^o\left(\widehat{\boldsymbol{\theta}}^\star,\boldsymbol{\theta};\mathbf{W}\right)$, and the left-hand side of the above inequality is termed the asymptotic distributional risk (ADR), i.e. the ADR of $\widehat{\boldsymbol{\theta}}^\star$ is defined as

$$ADR\left(\widehat{\boldsymbol{\theta}}^{\star}, \boldsymbol{\theta}; \mathbf{W}\right) = trace\left(E\left(\boldsymbol{\rho}^{*} \boldsymbol{W} \boldsymbol{\rho}^{*'}\right)\right). \tag{3.3}$$

In general, ADR $(\widehat{\boldsymbol{\theta}}^{\star}, \boldsymbol{\theta}; \mathbf{W}) \leq \mathbb{R}^{o}(\widehat{\boldsymbol{\theta}}^{\star}, \boldsymbol{\theta}; \mathbf{W})$ and additional assumptions are required for the equality to hold. The study of such assumptions is beyond the scope of this paper. Here, as in [2], Nkurunziza and Ahmed [14] among others, the performance of the estimators of $\boldsymbol{\theta}$ are based on theirs ADR. More specifically, the estimator $\widehat{\boldsymbol{\theta}}_{1}^{\star}$ is asymptotically more efficient than $\widehat{\boldsymbol{\theta}}_{2}^{\star}$ if ADR $(\widehat{\boldsymbol{\theta}}_{1}^{\star}, \boldsymbol{\theta}; \mathbf{W}) \leq \text{ADR}(\widehat{\boldsymbol{\theta}}_{2}^{\star}, \boldsymbol{\theta}; \mathbf{W})$.

Even though the shrinkage estimators are, in general, biased, the bias is accompanied by reduction in risk, and hence, it does not have a serious impact on risk assessment. Recall that the asymptotic bias is defined as $B_T^0\left(\widehat{\boldsymbol{\theta}}^\star, \boldsymbol{\theta}\right) = E\left[\sqrt{T}\left(\widehat{\boldsymbol{\theta}}^\star - \boldsymbol{\theta}\right)\boldsymbol{L}_2\right]$, and then, the asymptotic distributional bias (ADB) is defined as $E\left[\boldsymbol{\rho}^*\right]$. Below, Theorem 3.1–3.3 give ADR and ADB of the shrinkage estimators. The derivation of ADR and ADB of the shrinkage estimators are based on Proposition 2.4 and Corollary 2.5, along with the results on the normal distribution parametric model. To simplify the notation, let $H_{\nu}(x; \Delta) = P\{\chi_{\nu}^2(\Delta) \leq x\}$, $x \geqslant 0$ with $\Delta = \operatorname{trace}\left(\delta^{*'}\Xi\delta^*\right)$.

Theorem 3.1. If the hypotheses of Proposition 2.4 hold, the ADB functions of the estimators are given by

$$ADB\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}\right) = \mathbf{0}, \quad ADB\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}\right) = \boldsymbol{\delta}^*, \quad ADB\left(\widehat{\boldsymbol{\theta}}^S, \boldsymbol{\theta}\right) = -\boldsymbol{\delta}^*(qr - 2)E\left\{\chi_{qr+2}^{-2}(\Delta)\right\}$$

$$ADB\left(\widehat{\boldsymbol{\theta}}^{S+}, \boldsymbol{\theta}\right) = -\boldsymbol{\delta}^*\left[H_{qr+2}(qr + 2; \Delta) + (qr - 2)E\left\{\chi_{qr+2}^{-2}(\Delta)I\left(\chi_{qr+2}^2(\Delta) > (qr - 2)\right)\right\}\right].$$
(3.4)

The proof is outlined in the Appendix A.

Remark 3.2. Note that the component $\delta *$ is common to the ADB of $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\theta}}^S$ and $\widehat{\boldsymbol{\theta}}^{S+}$, and thus, these expressions differ only by scalar factors Δ . On one hand, the bias of the $\widetilde{\boldsymbol{\theta}}$ is an unbounded function of Δ . On the other hand, the ADB of both $\widehat{\boldsymbol{\theta}}^S$ and $\widehat{\boldsymbol{\theta}}^{S+}$ are bounded in Δ . Also, since $\mathrm{E}\{\chi_{qr+2}^{-2}(\Delta)\}$ is a decreasing log-convex function of Δ , the ADB of $\widehat{\boldsymbol{\theta}}^S$ starts from the origin at $\Delta=0$, increases to a maximum, and then decreases to 0. Further, the bias curve of $\widehat{\boldsymbol{\theta}}^{S+}$ remains below the curve of the SE $\widehat{\boldsymbol{\theta}}^S$ for all values of Δ .

Theorem 3.3. If the hypotheses of Proposition 2.4 hold, the ADR functions of the estimators are given by

$$\begin{split} \operatorname{ADR}\left(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta};\mathbf{W}\right) &= \operatorname{trace}\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)\operatorname{trace}(\mathbf{W}\boldsymbol{\Sigma}), \\ \operatorname{ADR}\left(\widetilde{\boldsymbol{\theta}},\boldsymbol{\theta};\mathbf{W}\right) &= \operatorname{trace}\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)\operatorname{trace}(\mathbf{W}\boldsymbol{\Sigma}) - \operatorname{trace}\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)\operatorname{trace}(\mathbf{W}\boldsymbol{\Sigma}^{*}) + \operatorname{trace}\left(\boldsymbol{\delta}^{*'}\boldsymbol{W}\boldsymbol{\delta}^{*}\right), \\ \operatorname{ADR}\left(\widehat{\boldsymbol{\theta}}^{S},\boldsymbol{\theta};\mathbf{W}\right) &= \operatorname{ADR}(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta};\mathbf{W}) + \operatorname{trace}\left(\boldsymbol{\delta}^{*'}\boldsymbol{W}\boldsymbol{\delta}^{*}\right)\left((qr)^{2} - 4\right)\operatorname{E}\left(\chi_{qr+4}^{-4}(\Delta)\right) \\ &- (qr-2)\operatorname{trace}\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)\operatorname{trace}\left(\mathbf{W}\boldsymbol{\Sigma}^{*}\right)\left\{2\operatorname{E}\left(\chi_{qr+2}^{-2}(\Delta)\right) - (qr-2)\operatorname{E}\left(\chi_{qr+2}^{-4}(\Delta)\right)\right\}, \\ \operatorname{ADR}\left(\widehat{\boldsymbol{\theta}}^{S+},\boldsymbol{\theta};\mathbf{W}\right) &= \operatorname{ADR}(\widehat{\boldsymbol{\theta}}^{S},\boldsymbol{\theta};\mathbf{W}) - \operatorname{trace}\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)\operatorname{trace}\left(\mathbf{W}\boldsymbol{\Sigma}\right)H_{qr+2}(qr-2;\Delta) \\ &+ 2(qr-2)\operatorname{trace}\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)\operatorname{trace}\left(\mathbf{W}\boldsymbol{\Sigma}^{*}\right)\operatorname{E}\left\{\chi_{qr+2}^{-2}(\Delta)\mathbf{1}_{\left\{\chi_{qr+2}^{2}(\Delta)\leq(qr-2)\right\}}\right\} \\ &- (qr-2)^{2}\operatorname{trace}\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)\operatorname{trace}\left(\mathbf{W}\boldsymbol{\Sigma}^{*}\right)\operatorname{E}\left\{\chi_{qr+2}^{-4}(\Delta)\mathbf{1}_{\left\{\chi_{qr+2}^{2}(\Delta)\leq(qr-2)\right\}}\right\} \\ &- (qr-2)\operatorname{trace}\left(\boldsymbol{\delta}^{*'}\boldsymbol{W}\boldsymbol{\delta}^{*}\right)\left[2\operatorname{E}\left\{\chi_{qr+2}^{-2}(\Delta)\right)\mathbf{1}_{\left\{\chi_{qr+2}^{2}(\Delta)\leq(qr-2)\right\}}\right\} \\ &- 2\operatorname{E}\left\{\chi_{qr+4}^{-2}(\Delta)\mathbf{1}_{\left\{\chi_{qr+4}^{2}(\Delta)\leq(qr-2)\right\}}\right\} + (qr-2)\operatorname{E}\left\{\chi_{qr+4}^{-4}(\Delta)\mathbf{1}_{\left\{\chi_{qr+4}^{2}(\Delta)\leq(qr-2)\right\}}\right\}\right\} \\ &+ \operatorname{trace}\left(\boldsymbol{\delta}^{*'}\boldsymbol{W}\boldsymbol{\delta}^{*}\right)\left\{2H_{qr+2}(qr-2;\Delta) - H_{qr+4}(qr-2;\Delta)\right\}. \end{split}$$

The proof is outlined in the Appendix.

Remark 3.4. First, it is noticed that for the special case where r=1, Theorems 3.1 and 3.3 give the same result as in [14]. Second, for a suitable choice of the matrix \mathbf{W} , the risk dominance of the estimators are similar to that established under normal distribution. More specifically, the following corollary shows that both $\widehat{\boldsymbol{\theta}}^{S+}$ and $\widehat{\boldsymbol{\theta}}^{S}$ outperform $\widehat{\boldsymbol{\theta}}$. At first glance, this may seem to be in contradiction with the fact that the model (1.2) satisfies the condition of locally asymptotically normal (LAN) family. Recall that for a LAN family, MLE is optimal in class of regular (smooth) estimators (see [4], p. 249, among others). However, as in classical multivariate normal samples, the proposed shrinkage estimators (or Stein-type estimators) are not regular (see [4], pp. 249–250). So, here as classical shrinkage estimators, $\widehat{\boldsymbol{\theta}}^S$ and $\widehat{\boldsymbol{\theta}}^{S+}$ showcase the well known Stein's phenomenon.

Corollary 3.5. Let
$$\operatorname{ch_{max}}(\boldsymbol{A})$$
 denote the largest eigenvalue of the matrix \boldsymbol{A} and let $\boldsymbol{W} \in \{\boldsymbol{W}: 0 < (qr+2) \operatorname{ch_{max}}(\boldsymbol{W}\boldsymbol{\Sigma}^*) \leqslant 2 \operatorname{trace}(\boldsymbol{W}\boldsymbol{\Sigma}^*) \operatorname{trace}(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2)\}$. Then, $\operatorname{ADR}\left(\widehat{\boldsymbol{\theta}}^{S+}, \boldsymbol{\theta}; \mathbf{W}\right) \leqslant \operatorname{ADR}\left(\widehat{\boldsymbol{\theta}}^{S}, \boldsymbol{\theta}; \mathbf{W}\right) < \operatorname{ADR}\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \mathbf{W}\right)$ for all $\Delta \in [0, \infty)$.

The proof follows from Theorem 3.3 along with Courant's Theorem and some algebraic computations. For the convenience of the reader, the proof is outlined in the Appendix.

4. Simulation study

In this section, we carry out a Monte Carlo simulation study to examine risk (namely MSE) performance of all estimators under consideration. To this end, we consider $V(X^{(i)}(t)) = -X^{(i)}(t)$, i = 1, 2, ..., 3. Different values of p are explored, but in order to save the space, we present the results for p = 3 and p = 4 only. Also, for each i = 1, 2, ..., m, we choose $\Sigma_i = I_p$. Further, we choose different matrices $\theta^{(i)}$, i = 1, 2, ..., m, with m = 2 and m = 3. For the constraint in (1.6), we choose d = 0, and d = 1 where d = 1 where d = 1 is given as

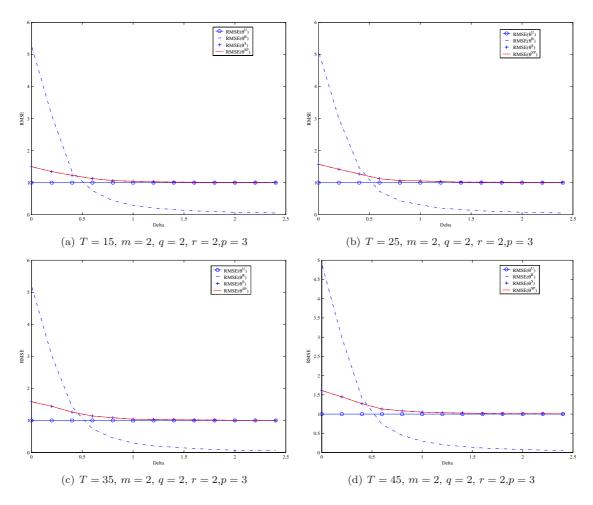


FIGURE 1. Relative efficiency vs. Δ .

in (1.5). Also, we choose some matrices L_2 , L_1^* according to the value of p. In particular, for p=3, we choose,

$$\boldsymbol{\theta}^{(1)} = \left(\begin{array}{cccc} 0.9501 & 0.4860 & 0.4565 \\ 0.2311 & 0.8913 & 0.0185 \\ 0.6068 & 0.7621 & 0.8214 \end{array} \right), \quad \boldsymbol{\theta}^{(2)} = \left(\begin{array}{cccc} 0.4447 & 0.9218 & 0.4057 \\ 0.6154 & 0.7382 & 0.9355 \\ 0.7919 & 0.1763 & 0.9169 \end{array} \right),$$

$$m{L}_2^* = \left(egin{array}{ccc} 1 & & 0 \\ 1 & & 0 \\ 1 & & 1 \end{array}
ight), \quad ext{ and } \quad m{L}_1^* = \left(egin{array}{ccc} -1 & 1 & & 0 \\ 0 & -1 & & 1 \end{array}
ight).$$

Further, for p = 4, we choose

$$\boldsymbol{\theta}^{(1)} = \begin{pmatrix} 0.5226 & 0.2714 & 0.1365 & 0.2987 \\ 0.8801 & 0.2523 & 0.0118 & 0.6614 \\ 0.1730 & 0.8757 & 0.8939 & 0.2844 \\ 0.9797 & 0.7373 & 0.1991 & 0.4692 \end{pmatrix}, \quad \boldsymbol{\theta}^{(2)} = \begin{pmatrix} 0.0648 & 0.5155 & 0.5798 & 0.2091 \\ 0.9883 & 0.3340 & 0.7604 & 0.3798 \\ 0.5828 & 0.4329 & 0.5298 & 0.7833 \\ 0.4235 & 0.2259 & 0.6405 & 0.6808 \end{pmatrix},$$

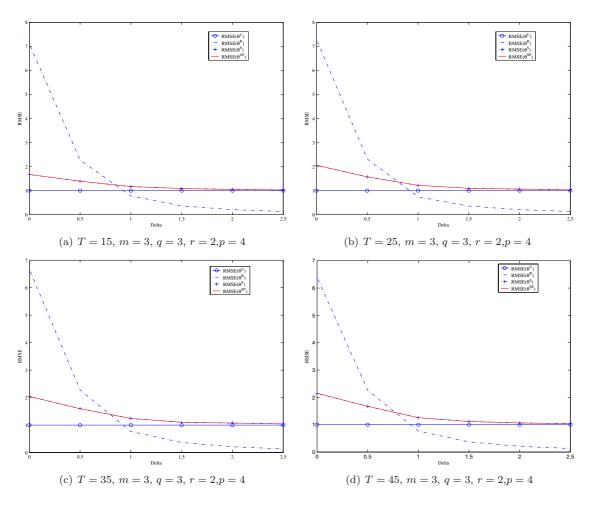


FIGURE 2. Relative efficiency $vs. \Delta$.

$$L_2^* = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad \text{and} \quad L_1^* = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

Also, the short and medium time periods of observation have been considered. Namely, we consider T=15, T=25, T=35, and T=45, and we perform 1000 replications. The comparison between estimators is based on the quantity called the relative mean square efficiency (RMSE) of the estimators with respect to $\hat{\theta}$. In passing, recall that RMSE is defined as RMSE (proposed estimator) = risk $(\hat{\theta})$ /risk (proposed estimator). Accordingly,

$$\text{RMSE}\left(\widehat{\boldsymbol{\theta}}\right) \ = \ \frac{\text{risk}\left(\widehat{\boldsymbol{\theta}}\right)}{\text{risk}\left(\widehat{\boldsymbol{\theta}}\right)}, \ \text{RMSE}\left(\widetilde{\boldsymbol{\theta}}\right) = \frac{\text{risk}\left(\widehat{\boldsymbol{\theta}}\right)}{\text{risk}\left(\widetilde{\boldsymbol{\theta}}\right)}, \ \text{RMSE}\left(\widehat{\boldsymbol{\theta}}^S\right) = \frac{\text{risk}\left(\widehat{\boldsymbol{\theta}}\right)}{\text{risk}\left(\widehat{\boldsymbol{\theta}}^S\right)}, \ \text{RMSE}\left(\widehat{\boldsymbol{\theta}}^{S+}\right) = \frac{\text{risk}\left(\widehat{\boldsymbol{\theta}}\right)}{\text{risk}\left(\widehat{\boldsymbol{\theta}}^{S+}\right)}.$$

Thus, a relative efficiency greater than one indicates the degree of superiority of the estimator over $\hat{\theta}$. Figures 1 and 2 highlight the performance of shrinkage strategy. Both Figures 1 and 2 corroborate the theoretical findings

as presented in Section 2. Indeed,

- (i) the asymptotic behavior of the SES is stable and, as theoretically expected, around the pivot, the SES dominate the MLE;
- (ii) around the pivot, $\hat{\theta}$ dominates shrinkage estimators. However, shrinkage estimators dominate $\tilde{\theta}$ as the constraint is seriously violated.

5. Conclusion

In this paper, we studied the inference problem concerning m-drift parameter matrices of k-multivariate diffusion processes, when these parameter matrices may satisfy certain constraints. We developed shrinkage estimation methods for the m-parameter matrices. As demonstrated in Section 2, the established shrinkage estimators improve the performance of the maximum likelihood estimator. This theoretical finding has been confirmed by the simulation study results. Further, our simulation finding show that the suggested SES is robust in the sense that it preserves a good performance whether or not the constraint holds. The proposed strategy is useful for example in econometric and/or financial modeling, particularly for model assessment and variable selection.

Acknowledgements. The author would like to acknowledge the financial support received from Natural Sciences and Engineering Research Council of Canada. Also, the author would like to thank the Editors-in-Chief and anonymous referees for useful comments.

APPENDIX A. SOME TECHNICAL DETAILS

Proof of Proposition 2.4. Using relation (2.6), we get $\left(\varrho_{T}^{'}, \zeta_{T}^{'}\right)^{'} = \left(\boldsymbol{I}_{mp}, \boldsymbol{I}_{mp} - \boldsymbol{L}_{1}^{'}\boldsymbol{J}^{'}\right)^{'} \varrho_{T} - \boldsymbol{J}\boldsymbol{\delta}$. Then, using Proposition 2.3 and Slutsky's theorem, we get $\left(\varrho_{T}^{'}, \zeta_{T}^{'}\right)^{'} \xrightarrow{\mathcal{L}} \left(\boldsymbol{I}_{p}, \boldsymbol{I}_{mp} - \boldsymbol{L}_{1}^{'}\boldsymbol{J}^{'}\right)^{'} \varrho - \boldsymbol{J}\boldsymbol{\delta}$ where $\varrho \sim \mathcal{N}_{mp \times r}\left(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{L}_{2}^{'}\boldsymbol{\Sigma}_{0}^{-1}\boldsymbol{L}_{2}\right)$. Hence,

$$\left(\boldsymbol{\varrho}_{T}^{'},\boldsymbol{\zeta}_{T}^{'}\right)^{'}\xrightarrow[T\to\infty]{\mathcal{L}}\mathcal{N}_{mp\times r}\left(-\boldsymbol{J}\boldsymbol{\delta},\left(\boldsymbol{I}_{mp},\,\boldsymbol{I}_{mp}-\boldsymbol{L}_{1}^{'}\boldsymbol{J}^{'}\right)^{'}\boldsymbol{\Sigma}\left(\boldsymbol{I}_{mp},\,\boldsymbol{I}_{mp}-\boldsymbol{L}_{1}\boldsymbol{J}\right)\otimes\boldsymbol{L}_{2}^{'}\boldsymbol{\Sigma}_{0}^{-1}\boldsymbol{L}_{2}\right).\tag{A.1}$$

Further,

$$\begin{pmatrix} \boldsymbol{\varrho}_T \\ \boldsymbol{\xi}_T \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_p & \mathbf{0} \\ \boldsymbol{I}_p & -\boldsymbol{I}_p \end{pmatrix} \begin{pmatrix} \boldsymbol{\varrho}_T \\ \boldsymbol{\zeta}_T \end{pmatrix}. \tag{A.2}$$

Therefore, by combining (A.1) and (A.2), we get the first statement of the proposition. Further, we have $\zeta_T = (I_{mp}, -I_{mp}) (\varrho'_T, \xi'_T)'$, and then, using the first statement of the proposition and Slutsky's theorem we get the second statement and that completes the proof.

Proof of Corollary 2.5. From Proposition 2.4, under local alternative, $\boldsymbol{\xi}_T \xrightarrow[T \to \infty]{\mathcal{L}} \boldsymbol{Z} \sim \mathcal{N}_{mp \times r} (\boldsymbol{\delta}^*, \boldsymbol{\Sigma}^* \otimes (\boldsymbol{L}_2' \boldsymbol{\Sigma}^{-1} \boldsymbol{L}_2))$, and then

$$(L_{2}'\Sigma^{-1}L_{2})^{-\frac{1}{2}} \xi_{T}'\Xi \xi_{T} (L_{2}'\Sigma^{-1}L_{2})^{-\frac{1}{2}} \xrightarrow[T \to \infty]{\mathcal{L}} (L_{2}'\Sigma^{-1}L_{2})^{-\frac{1}{2}} Z'\Xi Z (L_{2}'\Sigma^{-1}L_{2})^{-\frac{1}{2}}.$$
(A.3)

Further, using the well known properties of quadratic random matrices (see for example [5]), we get

$$\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)^{-\frac{1}{2}}\boldsymbol{Z}^{\prime}\boldsymbol{\Xi}\boldsymbol{Z}\left(\boldsymbol{L}_{2}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right)^{-\frac{1}{2}}\sim W_{r}\left(q,\boldsymbol{I}_{r},\boldsymbol{\delta}^{*^{\prime}}\boldsymbol{\Xi}\boldsymbol{\delta}^{*}\right). \tag{A.4}$$

Therefore, combining relations (A.3) and (A.4), we get $(L_2'\Sigma^{-1}L_2)^{-\frac{1}{2}}\xi_T'\Xi\xi_T(L_2'\Sigma^{-1}L_2)^{-\frac{1}{2}}\xrightarrow{\mathcal{L}}W_r(q, I_r, \delta^{*'}\Xi\delta^*)$, and then, trace $\left[\left(L_2'\Sigma^{-1}L_2\right)^{-\frac{1}{2}}\xi_T'\Xi\xi_T\left(L_2'\Sigma^{-1}L_2\right)^{-\frac{1}{2}}\right]\xrightarrow{\mathcal{L}}\chi_{rq}\left(\operatorname{trace}\left(\delta^{*'}\Xi\delta^*\right)\right)$, that completes the proof.

Proof of Theorem 3.1. Using Proposition 2.4, we have $ADB\left(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta}\right) = \mathrm{E}\left(\boldsymbol{\rho}\right) = \mathbf{0}$, and $ADB\left(\widetilde{\boldsymbol{\theta}},\boldsymbol{\theta}\right) = \mathrm{E}\left(\boldsymbol{\zeta}\right) = -\boldsymbol{\delta}^*$, that proves the two first statements. Also, we have $\mathrm{ADB}\left(\widehat{\boldsymbol{\theta}}^S,\boldsymbol{\theta}\right) = \mathrm{E}\left[\boldsymbol{\zeta} + \left(1 - (qr - 2)\varphi^{-1}\right)\boldsymbol{\xi}\right] = -\boldsymbol{\delta}^* + \mathrm{E}\left[\left(1 - (qr - 2)\varphi^{-1}!\right)\boldsymbol{\xi}\right]$. Further, let $\Delta = \mathrm{trace}\left(\boldsymbol{\delta}^{*'}\boldsymbol{\Xi}\boldsymbol{\delta}^*\right)$. Using the fact that $\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-\frac{1}{2}}$ is a symmetric and idempotent matrix, along with Theorem 2 in [9], we get $\mathrm{Vec}\left(\mathrm{ADB}\left(\widehat{\boldsymbol{\theta}}^S,\boldsymbol{\theta}\right)\right) = -\mathrm{Vec}\left(\boldsymbol{\delta}^*\right) + \mathrm{E}\left[\left(1 - (qr - 2)/\chi_{qr+2}^2\left(\Delta\right)\right)\right]\mathrm{Vec}\left(\boldsymbol{\delta}^*\right)$ and this proves the third statement of the theorem. The last statement of the theorem is proved by using same method.

Proposition A.1. Let $\Lambda = L_2' \Sigma^{-1} L_2$ and let c be a real number. If the hypotheses of Proposition 2.4 hold, then

$$\begin{split} E\left\{\mathrm{trace}\left[\left(1-c\psi^{-1}\right)^{2}\boldsymbol{\xi}'\boldsymbol{W}\boldsymbol{\xi}\right]\right\} &= E\left[\left(1-c\chi_{qr+2}^{-2}\left(\Delta\right)\right)^{2}\right]\mathrm{trace}\left(\boldsymbol{W}\boldsymbol{\Sigma}^{*}\right)\mathrm{trace}\left(\boldsymbol{\Lambda}\right) \\ &+ E\left[\left(1-c\chi_{qr+4}^{-2}\left(\Delta\right)\right)^{2}\right]\mathrm{trace}\left(\boldsymbol{\delta}^{*'}\boldsymbol{W}\boldsymbol{\delta}^{*}\right), \end{split}$$

and
$$E\left[\left(1-c\psi^{-1}\right)\eta'W\xi\right] = -E\left[\left(1-c\chi_{qr+2}^{-2}(\Delta)\right)\right]\delta^{*'}W\delta^{*}.$$

Proof. The proof of the first statement follows by using similar transformations as used in proof of the third and fourth statements of Theorem 3.1. Further, since $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are independent, we have $\mathrm{E}\big[\left(1-c\,\psi^{-1}\right)\boldsymbol{\eta}'\boldsymbol{W}\boldsymbol{\xi}\big]=(\mathrm{E}\left(\boldsymbol{\eta}\right))'\boldsymbol{W}$ $\mathrm{E}\left[\left(1-c\,\psi^{-1}\right)\boldsymbol{\xi}\right]=-\left(\boldsymbol{\delta}^*\right)'\boldsymbol{W}$ $\mathrm{E}\left[\left(1-c\,\psi^{-1}\right)\boldsymbol{\xi}\right]$, and then, following the same steps as used in proof of Theorem 3.1, we get $\mathrm{E}\left[\left(1-c\,\psi^{-1}\right)\boldsymbol{W}\boldsymbol{\xi}\right]=\mathrm{E}\left[\left(1-c\,\chi_{qr+2}^{-2}\left(\Delta\right)\right)\right]\boldsymbol{\delta}^*$, that completes the proof of the proposition.

Proof of Theorem 3.3. By applying Proposition 2.4, we establish the ADR of $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$. Further, we have ADR $\left(\widehat{\boldsymbol{\theta}}^S, \boldsymbol{\theta}; \mathbf{W}\right) = \mathbb{E}\left\{\operatorname{trace}\left[\left(\boldsymbol{\eta} + \left(1 - (qr - 2)\psi^{-1}\right)\boldsymbol{\xi}\right)'\boldsymbol{W}\left(\boldsymbol{\eta} + \left(1 - (qr - 2)\psi^{-1}\right)\boldsymbol{\xi}\right)\right]\right\}$, and that gives

$$ADR\left(\widehat{\boldsymbol{\theta}}^{S}, \boldsymbol{\theta}; \mathbf{W}\right) = ADR\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}; \mathbf{W}\right) + 2E\left\{trace\left[\boldsymbol{\eta}'\boldsymbol{W}\left(1 - (qr - 2)\psi^{-1}\right)\boldsymbol{\xi}\right]\right\} + E\left\{trace\left[\boldsymbol{\xi}'\boldsymbol{W}\left(1 - (qr - 2)\psi^{-1}\right)^{2}\boldsymbol{\xi}\right]\right\}. \quad (A.5)$$

Then, applying Proposition 5, we have

$$\begin{aligned} & \text{ADR}\left(\widehat{\boldsymbol{\theta}}^{S}, \boldsymbol{\theta}; \mathbf{W}\right) = \text{ADR}\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}; \mathbf{W}\right) - 2 \; \text{E}\left[\left(1 - \left(qr - 2\right) \chi_{qr+2}^{-2}\left(\Delta\right)\right)\right] \; \text{trace}\left(\boldsymbol{\delta}^{*'} \boldsymbol{W} \boldsymbol{\delta}^{*}\right) \\ & + \; \text{E}\left[\left(1 - \left(qr - 2\right) \chi_{qr+2}^{-2}\left(\Delta\right)\right)^{2}\right] \; \; \text{trace}\left(\boldsymbol{W} \boldsymbol{\Sigma}^{*}\right) \; \; \text{trace}\left(\boldsymbol{\Lambda}\right) + \; \text{E}\left[\left(1 - \left(qr - 2\right) \chi_{qr+4}^{-2}\left(\Delta\right)\right)^{2}\right] \; \; \text{trace}\left(\boldsymbol{\delta}^{*'} \boldsymbol{W} \boldsymbol{\delta}^{*}\right), \end{aligned}$$

and then, by some computations, we get

$$\begin{aligned} \text{ADR}\left(\widehat{\boldsymbol{\theta}}^{S},\boldsymbol{\theta};\mathbf{W}\right) &= \text{ADR}\left(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta};\mathbf{W}\right) + \text{trace}\left(\boldsymbol{\delta}^{*'}\boldsymbol{W}\boldsymbol{\delta}^{*}\right)\left(\left(qr\right)^{2} - 4\right) \text{E}\left(\chi_{qr+4}^{-4}(\Delta)\right) \\ &- \left(qr-2\right) \text{ trace}\left(\boldsymbol{L}_{2}'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_{2}\right) \text{trace}\left(\mathbf{W}\boldsymbol{\Sigma}^{*}\right) \left\{2 \text{ E}\left(\chi_{qr+2}^{-2}(\Delta)\right) - \left(qr-2\right) \text{ E}\left(\chi_{qr+2}^{-4}(\Delta)\right)\right\}, \end{aligned}$$

Further, by following the same steps, we establish the ADR of $\hat{\boldsymbol{\theta}}^{S+}$.

Proof of Corollary 3.5. Let $c_1 = \operatorname{trace} \left(\mathbf{L}_2' \mathbf{\Sigma}^{-1} \mathbf{L}_2 \right) \operatorname{trace} \left(\mathbf{W} \mathbf{\Sigma}^* \right)$ and let $c_2 = \operatorname{trace} \left(\boldsymbol{\delta}^{*'} \mathbf{W} \boldsymbol{\delta}^* \right)$. By using Theorem 3.3 and the identity $(qr-2) \operatorname{E} \left(\chi_{qr+2}^{-4}(\Delta) \right) = \operatorname{E} \left(\chi_{qr+2}^{-2}(\Delta) \right) - 2\Delta \operatorname{E} \left(\chi_{qr+4}^{-4}(\Delta) \right)$, we have $\operatorname{ADR}(\widehat{\boldsymbol{\theta}}^S, \boldsymbol{\theta}; \mathbf{W}) - \operatorname{ADR} \left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}, \mathbf{W} \right) = -(qr-2) \left\{ c_1 \operatorname{E} \left(\chi_{qr+2}^{-2}(\Delta) \right) + \left[2\Delta c_1 \operatorname{E} \left(\chi_{qr+2}^{-4}(\Delta) \right) - (qr+2)c_2 \operatorname{E} \left(\chi_{qr+2}^{-4}(\Delta) \right) \right] \right\}$. Then, since qr-2>0, it suffices to prove that for the given choice of \mathbf{W} , we have $2\Delta c_1 - (qr+2)c_2 \geq 0$, for all $\Delta \geq 0$. Obviously, the inequality holds if $c_2=0$. Further, if $c_2>0$, the inequality is equivalent to $\Delta c_1/c_2 \geq (qr+2)/2$, for all $\Delta \geq 0$. Let $ch_{\min}(\mathbf{A})$ denote the smallest eigenvalue of the matrix \mathbf{A} . By using Courant's Theorem along with some algebraic computations, we have $c_1/ch_{\max}(\mathbf{W} \mathbf{\Sigma}^*) \leq \Delta c_1/c_2 \leq c_1/ch_{\min}(\mathbf{W} \mathbf{\Sigma}^*)$, for all $\Delta \geq 0$, that proves the desired inequality. Further, by using Theorem 3.3 along with some algebraic computations, one can prove that $\operatorname{ADR}\left(\widehat{\boldsymbol{\theta}}^{S+}, \boldsymbol{\theta}; \mathbf{W}\right) \leq \operatorname{ADR}\left(\widehat{\boldsymbol{\theta}}^S, \boldsymbol{\theta}; \mathbf{W}\right)$ for all $\Delta \geq 0$, that completes the proof.

References

- [1] S.E. Ahmed, Shrinkage estimation of regression coefficients from censored data with multiple observations, in *Empirical Bayes* and *Likelihood inference*, edited by S.E. Ahmed and N. Reid. Springer, New York (2001) 103–120.
- [2] S.E. Ahmed and A.K.Md.E. Saleh, Improved nonparametric estimation of location vector in a multivariate regression model. J. Nonparametr. Stat. 11 (1999) 51–78.
- [3] A.R. Bergstrom, Continuous Time Econometric Modelling. Oxford University Press, Oxford (1990).
- [4] A. DasGupta, Asymptotic theory of statistics and probability. Springer Science & Business Media, New York (2008).
- [5] M.C.M. De Gunst, On the distribution of general quadratic functions in normal vectors. Stat. Neerl. 41 (1987) 245–251.
- [6] S. Engen, R. Lande, T. Wall and J.P. DeVries, Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. The American Naturalist 160 (2002) 60–73.
- [7] S. Iyengar, Diffusion models for neutral activity, in Statistics for the 21st Century: Methodologies for Applications of the Future, edited by C.R. Rao and G. Szekely. Marcel-Dekker (2000) 233–250.
- [8] A.J. Izenman, Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. Springer Science, Business Media, LLC (2008).
- [9] G.G. Judge and M.E. Bock, The statistical implication of pre-test and Stein-rule estimators in econometrics. Amsterdam, North Holland (1978).
- [10] I. Karatzas and S.E. Shreve, Brownian Motion and Stochastic Calculus. Springer-Verlag, New York (1991).
- [11] A.Y. Kutoyants, Statistical Inference for Ergodic Diffusion Processes, in *Springer Series in Statistics*. Springer-Verlag, London (2004).
- [12] R.S. Liptser and A.N. Shiryaev, Statistics of Random Processes: Generale Theory I. Springer-Verlag, New York (1977).
- [13] R.S. Liptser and A.N. Shiryaev, Statistics of Random Processes: Applications II. Springer-Verlag, New York (1978).
- [14] S. Nkurunziza and S.E. Ahmed, Shrinkage Drift Parameter Estimation for Multi-factor Ornstein-Uhlenbeck Processes. Appl. Stoch. Models Bus. Ind. 26 (2010) 103–124.
- [15] G. Papanicolaou, Diffusion in random media, in *Surveys in Applied Mathematics*, edited by J.B. Keller, D. McLaughlin and G. Papanicolaou. Plenum Press (1995) 205–255.