Journal of Statistical Research *ISSN 0256 - 422 X*
2010, Vol. 44, No. 2, pp. 243-266
Bangladesh

ESTIMATING AND PRETESTING IN SOME MULTIDIMENSIONAL INTEREST RATES MODELS

S. Ejaz Ahmed
Department of Mathematics and Statistics, University of Windsor
401 Sunset Avenue, Windsor, Ontario, Canada N9B 3P4
Email: seahmed@uwindsor.ca

Sévérien Nkurunziza
Department of Mathematics and Statistics, University of Windsor
401 Sunset Avenue, Windsor, Ontario, Canada N9B 3P4
Email: severien@uwindsor.ca

summary

We consider the general estimation problem of the drift parameter matrix in a multi-factors Vasicek model. We also develop estimation theory for the drift parameters under natural restrictions. In particular, we propose shrinkage and pretest estimators when the natural restrictions may or may not hold. Based on the asymptotic properties of both unrestricted and restricted maximum likelihood estimators (MLE), we examine the relative performance of the shrinkage and pretest estimators. Finally, we appraise the properties of the listed estimators based on a simulation study, insofar as applied implementation of the procedure is concerned.

Keywords and phrases: Gaussian process; Diffusion process; Interest rate model; MLE; Shrinkage and pretest estimators; Vasicek model

1 Introduction

In this paper, we consider inference problem concerning the drift parameter matrix of some multidimensional processes of the interest rates structure. In particular, we are interested in case where the drift parameter matrix is suspected to lie in a certain hyperplane. More specifically, we study the model for which the process under consideration follows a multidimensional Ornstein-Uhlenbeck process.

First, let us notice that the Ornstein-Uhlenbeck process has been extensively used in modelling diverse phenomenon in different scientific fields such as biology, ecology and finance. To give some examples, we quote Engen and Sther (2005), Engen *et al.* (2002), Nkurunziza (2010), Schbel and Zhu (1999), respectively. Particularly, in finance, Ornstein-Uhlenbeck process is mostly known as Vasicek process and is used in modelling the term structure of the interest rates. For more details about the interpretation of this model, we refer to Vasicek (1977), Abu-Mostafa (2001) among others.

Let $\{W(t), t \geqslant 0\}$ be a Wiener process. The instantaneous interest rate $r(t)$ is governed by the stochastic differential equation (SDE)

$$dr(t) = \theta\left(\alpha - r(t)\right)dt + \sigma dW(t), \quad \alpha, \theta, \sigma > 0, \tag{1.1}$$

where α denotes a steady-state interest rate (or the long-term mean), θ is the speed of converging to the steady-state, and σ is a volatility or "randomness level". The component $\theta\left(\alpha - r(t)\right)$ of the above relation represents the drift term. Thus, α and θ are so-called the drift parameters. The component $\sigma dW(t)$ is the diffusion term of the process whose σ is so-called the diffusion parameter. In the following sub-section we consider a more general form of the above model.

1.1 Multidimensional Model

It is noticed that the model in (1.1) is univariate while the term structure of the interest rates is embedded in a large macroeconomic system. To this end, Langetieg (1980) developed a multivariate version of the model (1.1), namely "multi-factors Vasicek model":

$$dr_k(t) = \sum_{j=1}^{p} \theta_{kj}\left(\alpha_j - r_j(t)\right)dt + \sigma_k dW_k(t), \quad k = 1, 2, \ldots, p, \tag{1.2}$$

where $\{W_k, t \geqslant 0\}$, $k = 1, 2, \ldots, p$ are Wiener processes possibly correlated. The above model takes into account the arbitrary number of economic relationships. In some economic perspectives, such multiple factors model has the advantages of allowing the analysis of the simultaneous impact of correlated factors on the behavior of interest rates. Also, the model (1.2) includes the case where $\theta_{kj} = 0$ for all $1 \leqslant j \neq k \leqslant p$ and with $\theta_{jj} > 0$ for all $j = 1, 2, \ldots, p$. In this particular scenario, we have

$$dr_k(t) = \theta_{kk}\left(\alpha_k - r_k(t)\right)dt + \sigma_k dW_k(t), \quad k = 1, 2, \ldots, p, \tag{1.3}$$

Thereafter, we refer this model as the "non-interdependence model." The interest rate Vasicek model has been extensively used in literature. For example, Georges (2003) used the model to analyze the maturity structure of the public debt using Canadian and Danish data. Abu-Mostafa (2001) calibrates the correlated multi-factors Vasicek model of interest rates, and then applied it to Japanese Yen swaps market and U.S. Dollar yield market. Liptser and Shiryayev (1978), Basawa and Rao (1980) and Kutoyants (2004) considered the inference problem concerning the drift parameters of the model (1.1) and derived its maximum likelihood estimator (MLE). Our goal is here to consider some inference problem for the drift parameters of multivariate Vasicek model. Our parameter of interest is $\boldsymbol{\theta}$ (that is a $p \times p$-matrix). For the sake of brevity, we consider that α is either known or available from the previous studies. For this reason, we let $X_k(t) = r_k(t) - \alpha_k$. Then, from the multi-factors Vasicek model (1.2), we get

$$dX_k(t) = -\sum_{j=1}^{p} \theta_{kj} X_j(t)dt + \sigma_k dW_k(t), \quad k = 1, 2, \ldots, p, \tag{1.4}$$

where $\theta_k > 0$, $\sigma_k > 0$, $k = 1, 2, \ldots, p$. We assume that $\{W_k(t), t \geqslant 0, k = 1, 2 \ldots, p\}$ is a p-dimensional Wiener process. The independence assumption between the p different Wiener processes is made to simplify the analysis. Similar results can be established for the case where these Wiener processes are pairwise jointly Gaussian with a non-zero correlation coefficient.

For the mathematical convenience, we assume a continuous observation of the trajectory in order to derive theoretical results. However, in practice, the observations are collected at discrete times $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$ and thus, the continuous time modelling is derived through some approximations. In order to guarantee the accuracy of such approximations, we assume that the observation times are dense. Our statistical procedure is applied by replacing each stochastic integral by its corresponding discrete Riemann-Itô sum. Theoretically, it is well known that the resulting new estimator is asymptotically equivalent to the original estimator obtained under continuous time sampling (see e.g. Le Breton, 1976), subject the fact that the observation times are very closely spaced. Nevertheless, there is a loss of information due to discretizationas as discussed in Dacunha-Castelle and Florens-Zmirou (1986) and Florens-Zmirou (1989).

Motivated by diverse applications of the above model, we consider the estimation of the drift parameter $p \times p$-matrix $\boldsymbol{\theta} = (\theta_{ij})$, $1 \leqslant i \leqslant p$, $1 \leqslant j \leqslant p$. However, in many experiment the parameter space is usually restricted and thus, there may be some interrelationships among the components of the parameter matrix. In practice, the experimenters may encounter one of the two following scenarios,

$$\boldsymbol{L}_1 \boldsymbol{\theta} \boldsymbol{L}_2 = \boldsymbol{d}, \quad \text{or} \quad \boldsymbol{L}_1^* \boldsymbol{\theta} = \boldsymbol{D}^* \tag{1.5}$$

where $\boldsymbol{L}_1$ is a $q \times p$-known full rank matrix with $q < p$, and $\boldsymbol{L}_2$ and $\boldsymbol{d}$ are known $p \times 1$ and $q \times 1$-column vectors, respectively. Further, $\boldsymbol{L}_1^*$ and $\boldsymbol{D}^*$ are $q \times p$-known matrix full rank with $q < p$. For a suitable choice of $\boldsymbol{L}_1$, $\boldsymbol{L}_2$, $\boldsymbol{d}$ and $\boldsymbol{d}^*$, the above two restrictions include the equality of the parameters, see Section 3. The estimation and testing of homogeneity of parameters is of great interest. For example, such a problem arises naturally in model selection procedures. An illustrative case concerns the situation where several countries have decided to unify their economic policy, and the measurements, of a certain economic index, have been taken in these countries. In this case, it is reasonable to suspect the homogeneity of the several drift parameters. The main contribution of this paper is to suggest some improved estimators of $\boldsymbol{\theta}$ with high estimation accuracy when $\boldsymbol{\theta}$ is suspected to satisfy the restriction as given in (1.5). As a particular case, an asymptotic test for the equality of parameters is also suggested. Hence, we have extended the single factor inference problem to the multidimensional. Indeed, the problem of estimation in multidimensional diffusion processes is in general difficult and has received, as a consequence, less attention than the simpler one-dimensional case. Further, the problem studied here is more complex since, in addition, the parameter matrix may satisfy or not some constraints. In Nkurunziza and Ahmed (2010), the authors study similar problem, but they do not provide Pretest (Pretest/Shrinkage) Estimators.

The rest of this paper is organized as follows. Section 2 gives preliminaries results. Namely, in this section, we present the MLE (unrestricted and restricted MLE) and the shrinkage and pretest estimators as well as their respective asymptotic results. In Section 3, we establish the supremacy of shrinkage and pretest estimator over the MLE. Section 4 deals with a particular scenario where we consider the analysis of a non-interdependence model. In Section 5, we study the performance of the proposed method through simulation studies. Section 6 offers some recommendations and conclusions. Finally, some technical results are outlined in the Appendix.

2 Preliminary Results

Let $\{X_1(t), 0 \leqslant t \leqslant T\}$, $\{X_2(t), 0 \leqslant t \leqslant T\}$,..., $\{X_p(t), 0 \leqslant t \leqslant T\}$ be diffusion processes whose diffusion equations are given by

$$dX_k(t) = -\sum_{j=1}^{p} \theta_{kj} X_j(t)dt + \sigma_k dW_k(t) \qquad X_k(0) \text{ fixed }, \tag{2.1}$$

where $\sigma_k > 0$, $k = 1, 2, \ldots, p$ and $\{W_k(t), 0 \leqslant t \leqslant T\}$, $k = 1, 2, \ldots, p$ are p-independent Wiener processes. Let $\boldsymbol{\Sigma}$ be the diagonal matrix whose diagonal entrees are $\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$, i.e.,
$\boldsymbol{\Sigma} = \text{diag}\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2\right)$. Also, let $\boldsymbol{X}(t)$ and $\boldsymbol{W}(t)$ be column vectors given by

$$\boldsymbol{X}(t) = (X_1(t), X_2(t), \ldots, X_p(t))', \quad \boldsymbol{W}(t) = (W_1(t), W_2(t), \ldots, W_p(t))',$$

for $0 \leqslant t \leqslant T$. Let $\boldsymbol{\theta} = (\theta_{ij})_{1 \leqslant i,j \leqslant p}$, from relation (2.1), we have

$$d\boldsymbol{X}(t) = -\boldsymbol{\theta}\boldsymbol{X}(t)dt + \boldsymbol{\Sigma}^{\frac{1}{2}} d\boldsymbol{W}(t), \quad 0 \leqslant t \leqslant T. \tag{2.2}$$

To this end, we consider the following classical statistical hypotheses testing problem:

$$H_0 : \boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2 = \boldsymbol{d} \text{ versus } H_1 : \boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2 \neq \boldsymbol{d}. \tag{2.3}$$

In Section 4, we consider the the non-interdependence case that was introduced in (1.3), Section 1.

2.1 First Order Asymptotic and MLE

In this subsection, we recall some preliminary results about MLE which are used in deriving Shrinkage/Pretest Estimators. For more details, we refer to Nkurunziza and Ahmed (2010).

Let

$$\boldsymbol{U}_T = \int_0^T d\boldsymbol{X}(t)\boldsymbol{X}'(t), \quad \boldsymbol{D}_T = \int_0^T \boldsymbol{X}(t)\boldsymbol{X}'(t)dt. \tag{2.4}$$

Conditionally to $\boldsymbol{X}_0$, let $\widehat{\boldsymbol{\theta}}_T$ be the MLE of $\boldsymbol{\theta}$ satisfying the model (2.2).

Proposition 2.1. ***(Unrestricted MLE)*** *If the model in (2.2) holds then, the unrestricted MLE is*

$$\widehat{\boldsymbol{\theta}}_T = -\boldsymbol{U}_T\boldsymbol{D}_T^{-1} = \left(\widehat{\theta}_{ij}\right)_{1\leqslant i,j\leqslant p}. \tag{2.5}$$

Proof. See Nkurunziza and Ahmed (2010). □

Further, Proposition 2.2 gives the restricted MLE (RMLE). Let $\widetilde{\boldsymbol{\theta}}_T$ denote RMLE of $\boldsymbol{\theta}$ under H_0 and let $\boldsymbol{J} = \boldsymbol{\Sigma}\boldsymbol{L}_1'\left(\boldsymbol{L}_1\boldsymbol{\Sigma}\boldsymbol{L}_1'\right)^{-1}$.

Proposition 2.2. ***(Restricted MLE)*** *Assume that relation (2.2) holds. Then, under H_0 the restricted MLE is*

$$\widetilde{\boldsymbol{\theta}}_T = \widehat{\boldsymbol{\theta}}_T - \boldsymbol{J}\left(\boldsymbol{L}_1\widehat{\boldsymbol{\theta}}_T\boldsymbol{L}_2 - \boldsymbol{d}\right)\left(\boldsymbol{L}_2'\boldsymbol{D}_T^{-1}\boldsymbol{L}_2\right)^{-1}\boldsymbol{L}_2'\boldsymbol{D}_T^{-1}. \tag{2.6}$$

Proof. The proof follows directly by maximizing (A.1) under the first constraint in (1.5). Indeed, let $\boldsymbol{\lambda}$ be q-column vector of Lagrangian multipliers and let $\mathfrak{L}_{\boldsymbol{\lambda}}(\boldsymbol{\theta})$ be the Lagrangian, we have

$$\mathfrak{L}_{\boldsymbol{\lambda}}(\boldsymbol{\theta}) = \text{ trace}\left(\boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_T\right) + \frac{1}{2}\text{ trace}\left(\boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\boldsymbol{D}_T\right) + \boldsymbol{\lambda}'\left(\boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2 - \boldsymbol{d}\right),$$

and then, $\widetilde{\boldsymbol{\theta}}_T$ is the solution to the equation $\frac{\partial\mathfrak{L}_{\boldsymbol{\lambda}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \boldsymbol{0}$, $\quad \frac{\partial\mathfrak{L}_{\boldsymbol{\lambda}}(\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}} = \boldsymbol{0}$. Then, using the fact that

$$\partial\,\text{trace}\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\right)/\partial\boldsymbol{X} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B} + \boldsymbol{A}'\boldsymbol{X}\boldsymbol{B}', \quad \text{and} \quad \partial\,\text{trace}\left(\boldsymbol{A}\boldsymbol{X}\right)/\partial\boldsymbol{X} = \boldsymbol{A}',$$

along with some algebraic computations, we get the stated result. □

We established (see Proposition 2.3 given bellow) that $\widehat{\boldsymbol{\theta}}_T$ and $\widetilde{\boldsymbol{\theta}}_T$ are strongly consistent for $\boldsymbol{\theta}$. Moreover, as stated in Proposition 2.4, these estimators are asymptotically normal as T tends to infinity.

2.2 Large Sample Results and Test Statistic

In this sub-section, we establish asymptotic normality of the estimators and then use these results to build a test statistic for the testing problem on hand. As intermediate result, we present a proposition which shows that the UMLE $\widehat{\boldsymbol{\theta}}_T$ is strongly consistent. Also, under the null hypothesis in (2.3), the following proposition shows that the RMLE $\widetilde{\boldsymbol{\theta}}_T$ strong consistency.

Proposition 2.3. ***(Strong consistency)*** *Assume that the model (2.2) holds. Then,*

$$\Pr\left\{\lim_{T\to\infty}\widehat{\boldsymbol{\theta}}_T=\boldsymbol{\theta}\right\}=1 \quad \textit{and} \quad \Pr\left\{\lim_{T\to\infty}\widetilde{\boldsymbol{\theta}}_T=\boldsymbol{\theta}-\boldsymbol{J}\left(\boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2-\boldsymbol{d}\right)\left(\boldsymbol{L}_2'\boldsymbol{L}_2\right)^{-1}\boldsymbol{L}_2\right\}=1.$$

In particular, if $\boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2=\boldsymbol{d}$*, then* $\Pr\left\{\lim_{T\to\infty}\widetilde{\boldsymbol{\theta}}_T=\boldsymbol{\theta}\right\}=1.$

Note that the second statement of the proposition follows directly from the first statement. Also, note that the first statement corresponds to that given in Lipter and Shiryayev (1978, Theorem 17.4) and Kutoyants (2004) for the univariate case. Thus, the ideas of proof are the same as given in the quoted references. To give another reference where similar proof is given, we quote Nkurunziza and Ahmed (2007, Proposition 5, p. 17).

For the convenience of the reader, we present bellow Lemma 2.1 and Lemma 2.2 which are useful in deriving the asymptotic results.

Lemma 2.1. *Let* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be* $m\times p$ *and* $n\times q$ *matrices, respectively. Further, let* $\boldsymbol{X}\sim\mathcal{N}_{p\times n}\left(\boldsymbol{\mu},\Omega\otimes\Phi\right)$ *where* $A\otimes B$ *stands for the Kronecker product of the matrices* A *and* B*. Then,*

$$\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\sim\mathcal{N}_{m\times q}\left(\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{B},\,\left(\boldsymbol{B}'\Omega\boldsymbol{B}\right)\otimes\left(\boldsymbol{A}\Phi\boldsymbol{A}'\right)\right).$$

The proof of this Lemma 2.1 follows from the following well known algebraic properties on matrices vectorization. First, note that for a $m\times p$-matrix $\boldsymbol{A}$, one can write $\boldsymbol{A}=(A_1,A_2,\ldots,A_p)$, $A_j\in\mathbb{R}^p$, $j=1,2,\ldots,p$, where $\mathbb{R}^p$ denotes the p dimensional real space. Further, let $\mathrm{Vec}(\boldsymbol{A})$ denote the np-column vector obtained by stacking together the columns of $\boldsymbol{A}$ one underneath the other, i.e. $\mathrm{Vec}(\boldsymbol{A})=\left(A_1',A_2',\ldots,A_p'\right)'$.

Lemma 2.2. *Let* $\boldsymbol{A}$*,* $\boldsymbol{B}$ *and* $\boldsymbol{C}$ *be matrices such that* $\boldsymbol{ABC}$ *is well defined. Also, let* $\boldsymbol{I}$ *denote an identity matrix.*

$$Vec(\boldsymbol{ABC})=\left(\boldsymbol{I}\otimes\boldsymbol{AB}\right)Vec(\boldsymbol{C}) \quad \textit{and} \quad Vec(\boldsymbol{AB})=\left(\boldsymbol{I}\otimes\boldsymbol{A}\right)Vec(\boldsymbol{B})=\left(\boldsymbol{B}'\otimes\boldsymbol{I}\right)Vec(\boldsymbol{A}).$$

By using Lemma 2.1 and Lemma 2.2, and ergodic theory, we establish the following proposition.

Proposition 2.4. ***(Asymptotic normality)*** *Let* $W_p\left(n,\,\boldsymbol{\Sigma}\right)$ *denote a* $p\times p$*-random matrix whose distribution is Wishart with parameter* $\boldsymbol{\Sigma}$ *and degrees of freedom* n*. Also, assume that the model* (2.2) *holds and suppose that* $\boldsymbol{X}_0$ *has the same moment as the r.v. that follows the invariant distribution. Further, let* $\boldsymbol{V}_0=\mathrm{E}\left(\boldsymbol{X}_0\boldsymbol{X}_0'\right)$*. Then,*

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}\right)'\xrightarrow[T\to\infty]{\mathcal{L}}\mathcal{N}_{p\times p}\left(\boldsymbol{0},\,\boldsymbol{\Sigma}\otimes\boldsymbol{V}_0^{-1}\right) \ \textit{and}\ T\,\boldsymbol{V}_0^{-\frac{1}{2}}\left(\widehat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}\right)'\boldsymbol{\Sigma}^{-1}\left(\widehat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}\right)\boldsymbol{V}_0^{-\frac{1}{2}}\xrightarrow[T\to\infty]{\mathcal{L}}W_p\left(p,\,\boldsymbol{I}_p\right).$$

Proof. We have

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

and then

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T-\theta\right)'=-\left(\frac{1}{T}\boldsymbol{D}_T\right)^{-1}\left(\frac{1}{\sqrt{T}}\int_0^T \boldsymbol{X}(t)d\boldsymbol{W}'(t)\right)\boldsymbol{\Sigma}^{\frac{1}{2}}.$$

Form ergodicity theorem, we have

$$\frac{1}{T}\boldsymbol{D}_T \xrightarrow[T\to\infty]{a.s} \boldsymbol{V}_0. \tag{2.7}$$

Further, from Proposition 1.34 or Theorem 2.8 of Kutoyants (2004, p. 61 and p. 121), we have

$$\left(\frac{1}{\sqrt{T}}\int_0^T \boldsymbol{X}(t)d\boldsymbol{W}'(t)\right) \xrightarrow[T\to\infty]{\mathcal{L}} \boldsymbol{Z}\sim\mathcal{N}_{p\times p}\left(\mathbf{0},\,\boldsymbol{I}_p\otimes\boldsymbol{V}_0\right). \tag{2.8}$$

Hence, combining (2.7), (2.8) and Slutsky's theorem, we get

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T-\theta\right)' \xrightarrow[T\to\infty]{\mathcal{L}} -\boldsymbol{V}_0^{-1}\boldsymbol{Z}\boldsymbol{\Sigma}^{\frac{1}{2}},$$

and using Lemma 2.1, we have

$$-\boldsymbol{V}_0^{-1}\boldsymbol{Z}\boldsymbol{\Sigma}^{\frac{1}{2}}\sim\mathcal{N}_{p\times p}\left(\mathbf{0},\,\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}\otimes\boldsymbol{V}_0^{-1}\boldsymbol{V}_0\boldsymbol{V}_0^{-1}\right),$$

and that proves the first statement of the proposition. The second statement follows directly by applying the properties of Wishart distribution (see for example De Gunst, 1987), and that completes the proof. □

For power computation and related purposes we consider the following set of local alternatives,

$$K_T:\boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2=\boldsymbol{d}+\frac{\boldsymbol{\delta}}{\sqrt{T}},\quad T>0 \tag{2.9}$$

where $\boldsymbol{\delta}$ is a nonzero p-column vector with $\|\boldsymbol{\delta}\|<\infty$. Also, let $\boldsymbol{V}_0$ be the variance-covariance of invariant distribution and let $\boldsymbol{\Sigma}^*=\boldsymbol{J}\boldsymbol{L}_1\boldsymbol{\Sigma}$. Further, let

$$\boldsymbol{\varrho}_T=\sqrt{T}\,\left(\widehat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}\right)\boldsymbol{L}_2,\quad \boldsymbol{\xi}_T=\sqrt{T}\,\left(\widehat{\boldsymbol{\theta}}_T-\widetilde{\boldsymbol{\theta}}_T\right)\boldsymbol{L}_2,\quad \boldsymbol{\zeta}_T=\sqrt{T}\,\left(\widetilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}\right)\boldsymbol{L}_2.$$

Proposition 2.5. *Assume that the model (2.2) holds. Under the local alternative as given in (2.9),*

(i)

$$\begin{pmatrix}\boldsymbol{\varrho}_T\\ \boldsymbol{\xi}_T\end{pmatrix} \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}_{2p}\left(\begin{pmatrix}\mathbf{0}\\ \boldsymbol{J}\boldsymbol{\delta}\end{pmatrix},\,\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\begin{pmatrix}\boldsymbol{\Sigma} & \boldsymbol{\Sigma}^*\\ \boldsymbol{\Sigma}^* & \boldsymbol{\Sigma}^*\end{pmatrix}\right).$$

(ii)

$$\begin{pmatrix} \boldsymbol{\zeta}_T \\ \boldsymbol{\xi}_T \end{pmatrix} \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}_{2p}\left(\begin{pmatrix} -\boldsymbol{J}\boldsymbol{\delta} \\ \boldsymbol{J}\boldsymbol{\delta} \end{pmatrix}, \left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\begin{pmatrix} \boldsymbol{\Sigma}-\boldsymbol{\Sigma}^* & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}^* \end{pmatrix}\right).$$

Proof. (i) By some usual computations, we have $(\boldsymbol{\varrho}_T', \boldsymbol{\xi}_T')' = (\boldsymbol{I}_p, \boldsymbol{L}_1'\boldsymbol{J}')'\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right)\boldsymbol{L}_2 + (\boldsymbol{0}, \boldsymbol{\delta}'\boldsymbol{J}')$. Statement in *(i)* follows directly by using Proposition 2.4 and by the application of the Slutsky theorem.

(ii) It can be verified that $(\boldsymbol{\zeta}_T', \boldsymbol{\xi}_T')' = (\mathbf{I}_p - \boldsymbol{L}_1'\boldsymbol{J}', \boldsymbol{L}_1'\boldsymbol{J}')'\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right)\boldsymbol{L}_2 + (-\boldsymbol{I}_p, \boldsymbol{I}_p)'\boldsymbol{J}\boldsymbol{\delta}$. Then, the result in *(ii)* follows from Proposition 2.4 and the Slutsky's theorem, that completes the proof.

□

From Proposition 2.5, we establish the following corollary.

Corollary 2.1. Let $\boldsymbol{\Xi} = \boldsymbol{L}_1'\left(\boldsymbol{L}_1\boldsymbol{\Sigma}\boldsymbol{L}_1'\right)^{-1}\boldsymbol{L}_1\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}$ and let $\boldsymbol{\delta}^* = \boldsymbol{J}\boldsymbol{\delta}$. If Proposition 2.5 holds, we have

$$\boldsymbol{\xi}_T'\boldsymbol{\Xi}\boldsymbol{\xi}_T \xrightarrow[T\to\infty]{\mathcal{L}} \chi_q^2\left(\boldsymbol{\delta}^{*\prime}\boldsymbol{\Xi}\boldsymbol{\delta}^*\right), \quad \text{and under } H_0 \quad \boldsymbol{\xi}_T'\boldsymbol{\Xi}\boldsymbol{\xi}_T \xrightarrow[T\to\infty]{\mathcal{L}} \chi_q^2.$$

The proof of Corollary 2.1 is given in the Appendix. It should be noted that the convergence in Proposition 2.5 and Corollary 2.1 holds uniformly in $\boldsymbol{\delta}^*$ on every compact subset of $\mathbb{R}^p$. This allows us to conclude that, the statements in Proposition 2.5 and Corollary 2.1 still hold under the null hypothesis H_0. Based on these results, we consider the following test statistic, when $\boldsymbol{\Sigma}$ is known,

$$\varphi(T) = T\boldsymbol{L}_2'\left(\widehat{\boldsymbol{\theta}}_T - \widetilde{\boldsymbol{\theta}}_T\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\left(\widehat{\boldsymbol{\theta}}_T - \widetilde{\boldsymbol{\theta}}_T\right)\boldsymbol{L}_2. \tag{2.10}$$

By Corollary 2.1, under H_0, the test statistic follows a central chi-square distribution with q degrees of freedom

$$\varphi(T) \xrightarrow[T\to\infty]{\mathcal{L}} \chi_q^2 \quad (\text{ under } H_0).$$

Also, we denote $\chi_{q;\alpha}^2$ a nonnegative real number such that $\Pr\left\{\chi_q^2 > \chi_{q;\alpha}^2\right\} = \alpha$. Further, concerning the testing problem (2.3), for the $\boldsymbol{\Sigma}$ known case, we suggest the following test

$$\Psi = I\left(\varphi(T) > \chi_{q;\alpha}^2\right) \tag{2.11}$$

where $I(A)$ denotes the indicator function of an event A. Note that, when $\boldsymbol{\Sigma}$ is unknown, the test statistic (2.10) can be modified by replacing $\boldsymbol{\Sigma}$ with its strongly consistent estimator. The new test is asymptotically α-level, and it is asymptotically as powerful as the test given in (2.11). Let Π_Ψ denote the power function of the test Ψ.

Corollary 2.2. Under the model (2.2), the test Ψ in (2.11) is asymptotically α-level test for the testing problem (2.3). Further, under the conditions of Corollary 2.1, we have

$$\lim_{T\to\infty} \Pi_\Psi \left(\boldsymbol{\theta} + \frac{\boldsymbol{\delta}}{\sqrt{T}}\right) = Pr\left\{\chi_q^2\left(\boldsymbol{\delta}^{*'}\boldsymbol{\Xi}\boldsymbol{\delta}^*\right) > \chi_{q;\alpha}^2\left(\boldsymbol{\delta}^{*'}\boldsymbol{\Xi}\boldsymbol{\delta}^*\right)\right\}.$$

The proof is straightforward from Corollary 2.1.

Remark 1. We obtained the MLE (2.6) by assuming that the whole sample paths of (1.4) are observable. The continuous time process has the advantage to be mathematically more convenient. However, in practice the data are collected in discrete times and thus here we keep in mind the fact that a continuous time process is derived through some approximations of a process that is observed in discrete times. In order to evaluate the suggested estimator, the stochastic integrals are replaced with their corresponding discrete Riemann-Itô sums.

Remark 2. Noting that for the diffusion process (1.4), $\boldsymbol{\Sigma}$ is known (equals to the quadratic variation). For the corresponding incomplete sample paths, the covariance matrix $\boldsymbol{\Sigma}$ becomes unknown. Also, for the special case $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_p$, the estimator $\widetilde{\boldsymbol{\theta}}_T$ in (2.6) and test statistic in (2.11) does not depend on σ. Nevertheless, $\boldsymbol{\Sigma}$ can be replaced by its corresponding strongly consistent estimator $\widehat{\boldsymbol{\Sigma}}$. Then, by Slutsky theorem we can get the similar result when $\boldsymbol{\Sigma}$ is unknown. The strongly consistent estimator for the diffusion parameter is based on quadratic variation.

2.3 Shrinkage and Pretest Estimation

We will use the results of the preceding subsections to build some alternative estimators of $\boldsymbol{\theta}$. Our interest here is the estimation of $\boldsymbol{\theta}$ when it is suspected $\boldsymbol{L}_1\,\boldsymbol{\theta}\,\boldsymbol{L}_2 = \boldsymbol{d}$. It is advantageous to utilize this information in the estimation process to construct improved estimation for the drift parameters. It is reasonable then to move the unrestricted maximum likelihood estimator (UMLE) of $\boldsymbol{\theta}$ close to the RMLE. We define a linear shrinkage estimator (LSE) as

$$\widehat{\boldsymbol{\theta}}^S = \beta\widetilde{\boldsymbol{\theta}}_T + (1-\beta)\widehat{\boldsymbol{\theta}}_T,$$

where $\beta \in [0,1]$ denotes the shrinkage intensity. Noting that for $\beta = 1$ the shrinkage estimate equals the $\widetilde{\boldsymbol{\theta}}_T$ whereas for $\beta = 0$ the UMLE is recovered. The key advantage of this construction is that it outperforms the UMLE in important part of the parameter space. However, the key question in this type of estimator is how to select an optimal value for the shrinkage factor β. In some situations, it may suffice to fix the parameter β at some given value. The second choice, is to choose the parameter β in a data-driven fashion by explicitly minimizing a suitable risk function. A common but also computationally intensive approach is to estimate β by using cross-validation; we refer to Friedman (1989) among others. On the other hand, from a Bayesian perspective one can employ the empirical Bayes technique to infer β. In this case β is treated as a hyper-parameter and that may be estimated from the data by optimizing the marginal likelihood.

Here we treat β as the degree of trust in the prior information $\boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2 = \boldsymbol{d}$. The value of $\beta \in [0,1]$ may be assigned by the experimenter according to her/his prior belief in the prior value $\boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2 = \boldsymbol{d}$. Ahmed and Krzanowski (2004), Bickel and Doksum (2001), Chiou and Miao (2007) and others pointed out that such an estimator yields smaller mean squared error (MSE) when a priori information is correct or nearly correct, however at the expense of poorer performance in the rest of the parameter space induced by the prior information. In the present context, we will demonstrate that $\widehat{\boldsymbol{\theta}}^S$ will have a smaller MSE than $\widehat{\boldsymbol{\theta}}_T$ near the restriction, that is, $\boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2 = \boldsymbol{d}$. However, $\widetilde{\boldsymbol{\theta}}_T$ becomes considerably biased and inefficient when the restriction may not be judiciously justified. Accordingly, when the prior information is rather suspicious, it may be reasonable to construct a shrinkage preliminary test estimator (SPE) denoted by $\widehat{\boldsymbol{\theta}}^{SP}$ which incorporates a preliminary test on $\boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2 = \boldsymbol{d}$. Thus, the estimator $\widehat{\boldsymbol{\theta}}_T$ and $\widetilde{\boldsymbol{\theta}}_T$ is selected depending upon the outcome of the preliminary test. Thus, the shrinkage preliminary test estimator (SPE) is defined as

$$\widehat{\boldsymbol{\theta}}^{SP} = \widehat{\boldsymbol{\theta}}_T I\left(\varphi(T) \geq c_{T,\alpha}\right) + \widehat{\boldsymbol{\theta}}^S I\left(\varphi(T) < c_{T,\alpha}\right), \tag{2.12}$$

where $\varphi(T)$ is the test statistic for the null hypothesis $H_o : \boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2 = \boldsymbol{d}$, which is defined in (2.10) (see subsection 2.2). The critical value $c_{T,\alpha}$ converges to $\chi^2_{q,\alpha}$ as $T \to \infty$. Thus, the critical value $c_{T,\alpha}$ of $\varphi(T)$ may be approximated by $\chi^2_{q,\alpha}$, the upper $100\alpha\%$ critical value of the χ^2-distribution with q degree of freedom. If we substitute $\beta = 1$ in (2.12) we get

$$\widehat{\boldsymbol{\theta}}^{P} = \widehat{\boldsymbol{\theta}}_T I\left(\varphi(T) \geq c_{T,\alpha}\right) + \widetilde{\boldsymbol{\theta}}_T I\left(\varphi(T) < c_{T,\alpha}\right). \tag{2.13}$$

The estimator $\widehat{\boldsymbol{\theta}}^P$ is known as the usual preliminary estimator (PE), due to Bancroft (1944). The SPE may be viewed as an improved PE which represents both UE and PE for $\beta = 0$ and $\beta = 1$, respectively. For a discussion about preliminary testing in various context, we refer to Leeb (2003), Danilov and Magnus (2004), Khan and Ahmed (2006), Ahmed (2005), Kibria and Saleh (2006), Reif (2006), Ahmed et al. (2006), Zoua et al. (2007) and Pardoa and Menndez (2008). For a comprehensive discussion on preliminary test and shrinkage estimations, we refer to Saleh (2006) and references therein.

3 Main Result

In this section, we compare the performance of the proposed shrinkage and pre-test estimator with respect to the UMLE and RMLE. In particular, we study the behavior of the risk functions of these estimators. Here, the most difficulty consists in the fact that the closed form of the finite sample distributions of the maximum likelihood estimators $\widehat{\boldsymbol{\theta}}_T$ and $\widetilde{\boldsymbol{\theta}}_T$ are not known unless we adopt a sequential plans (see e.g. Liptser and Shiryayev, 1978, p. 219). Furthermore, the finite sample distribution theory of pretest estimators is not simple to obtain. This difficulty has been largely overcome by asymptotic methods (see e.g. Ahmed *et al.*, 2006 among others).

These asymptotic methods relate primarily to *convergence in distribution* which may not generally guarantee *convergence in quadratic risk*. This technicality has been taken care of by the introduction of *asymptotic distributional risk* (ADR) (Sen (1986)), which, in turn, is based on the concept of a *shrinking neighborhood* of the pivot for which the ADR serves a useful and interpretable role in *asymptotic risk analysis*.

It is well known that, even for normal distribution, the effective domain of risk dominance of SPE or PE over MLE is a small neighborhood of the chosen pivot (viz., $\boldsymbol{L}_1\boldsymbol{\theta}\boldsymbol{L}_2 = \boldsymbol{d}$); and as we make the observation period T large, this domain becomes narrower. This justifies the choice of the local alternatives given in (2.9).

Now we introduce the following optimality criterion. For an estimator $\widehat{\boldsymbol{\theta}}^\star$ of $\boldsymbol{\theta}$, we consider a *quadratic loss function* of the form

$$L\left(\widehat{\boldsymbol{\theta}}^\star, \boldsymbol{\theta}; \mathbf{W}\right) = \boldsymbol{L}_2' \left[\sqrt{T}\left(\widehat{\boldsymbol{\theta}}^\star - \boldsymbol{\theta}\right)\right]' \mathbf{W} \left[\sqrt{T}\left(\widehat{\boldsymbol{\theta}}^\star - \boldsymbol{\theta}\right)\right] \boldsymbol{L}_2, \tag{3.1}$$

where $\mathbf{W}$ is a positive semi-definite (p.s.d) matrix. Using the distribution of $\sqrt{T}\left(\widehat{\boldsymbol{\theta}}^\star - \boldsymbol{\theta}\right) \boldsymbol{L}_2$ and taking the expected value both sides of (3.1), we get the expected loss that would be called the *quadratic risk* $\mathrm{R}_T^o\left(\widehat{\boldsymbol{\theta}}^\star, \boldsymbol{\theta}; \mathbf{W}\right) = \mathrm{trace}\left(\mathbf{W}\widehat{\boldsymbol{\Sigma}}_T\right)$, where $\widehat{\boldsymbol{\Sigma}}_T$ is the dispersion matrix of $\sqrt{T}\left(\widehat{\boldsymbol{\theta}}^\star - \boldsymbol{\theta}\right) \boldsymbol{L}_2$. Whenever $\lim_{T\to\infty} \widehat{\boldsymbol{\Sigma}}_T = \boldsymbol{\Sigma}$ exists, $\mathrm{R}_T^o\left(\widehat{\boldsymbol{\theta}}^\star, \boldsymbol{\theta}; \mathbf{W}\right) \xrightarrow[T\to\infty]{} \mathrm{R}^o\left(\widehat{\boldsymbol{\theta}}^\star, \boldsymbol{\theta}; \mathbf{W}\right) =$ trace$(\mathbf{W}\boldsymbol{\Sigma})$, which is termed the *asymptotic risk*. To set up notation, let $\tilde{G}_T(\mathbf{u})$ ($\mathbf{u} \in \mathbb{R}^p$) denote the distribution of $\sqrt{T}\left(\widehat{\boldsymbol{\theta}}^\star - \boldsymbol{\theta}\right) \boldsymbol{L}_2$. Also, suppose that $\tilde{G}_T \to \tilde{G}$ (at all points of continuity of $\tilde{G}$), as $T \to \infty$, and let $\boldsymbol{\Sigma}_{\tilde{G}}$ be the dispersion matrix of $\tilde{G}$. Then the *asymptotic distributional risk* (ADR) of $\widehat{\boldsymbol{\theta}}^\star$ is defined as

$$\mathrm{R}^o\left(\widehat{\boldsymbol{\theta}}^\star, \boldsymbol{\theta}; \mathbf{W}\right) = \mathrm{trace}\left(\mathbf{W}\boldsymbol{\Sigma}_{\tilde{G}}\right) + \left[B\left(\widehat{\boldsymbol{\theta}}^\star, \boldsymbol{\theta}\right)\right]' \boldsymbol{W} \left[B\left(\widehat{\boldsymbol{\theta}}^\star, \boldsymbol{\theta}\right)\right], \tag{3.2}$$

where $B\left(\widehat{\boldsymbol{\theta}}^\star, \boldsymbol{\theta}\right) = \int_{\mathbb{R}^p} \mathbf{x} d\tilde{G}(\mathbf{x})$, that is called *asymptotic distributional bias* (ADB) of $\widehat{\boldsymbol{\theta}}^\star$.

Two crucial results to the study of ADR and ADB of the suggested estimators are given in Proposition 2.5 and Corollary 2.1. Indeed, from these results, we apply the results on the (normal distribution) parametric model, and thereby give the main results of this subsection. The results are presented without derivation since the proofs are similar to that given in Ahmed (2001). Let $\Delta = \boldsymbol{\delta}^{*'}\boldsymbol{\Xi}\boldsymbol{\delta}^*$ and let $H_\nu(x\,; \Delta) = \mathrm{P}\{\chi_\nu^2(\Delta) \leq x\}$, $x \geqslant 0$.

Theorem 1. *If Proposition 2.5 holds, then, the* ADB *functions of the estimators are given as follows:*

$$\begin{aligned} \mathrm{ADB}\left(\widehat{\boldsymbol{\theta}}_T\right) &= \mathbf{0}, \quad \mathrm{ADB}\left(\widetilde{\boldsymbol{\theta}}_T\right) = -\boldsymbol{\delta}^*, \quad \mathrm{ADB}\left(\widehat{\boldsymbol{\theta}}^S\right) = -\beta\,\boldsymbol{\delta}^* \\ \mathrm{ADB}\left(\widehat{\boldsymbol{\theta}}^P\right) &= -\boldsymbol{\delta}^*\left[H_{q+2}\left(\chi_{q,\alpha}^2; \Delta\right)\right], \quad \mathrm{ADB}\left(\widehat{\boldsymbol{\theta}}^{SP}\right) = -\beta\boldsymbol{\delta}^*\left[H_{q+2}\left(\chi_{q,\alpha}^2; \Delta\right)\right]. \end{aligned} \tag{3.3}$$

Since for the ADB of $\widetilde{\boldsymbol{\theta}}_T$, $\widehat{\boldsymbol{\theta}}^S$, $\widehat{\boldsymbol{\theta}}^P$ and $\widehat{\boldsymbol{\theta}}^{SP}$, the component $\boldsymbol{\delta}^*$ is common and they differ only by scalar factors, it suffices to compare the scalar factors Δ only. It is clear that bias of the $\widetilde{\boldsymbol{\theta}}_T$ and $\widehat{\boldsymbol{\theta}}^S$ are unbounded function of $||\boldsymbol{\delta}^*||$. On the other hand, the ADB of both

$\widehat{\boldsymbol{\theta}}^P$ and $\widehat{\boldsymbol{\theta}}^{SP}$ are bounded in Δ. The ADB of $\widehat{\boldsymbol{\theta}}^P$ starts from the origin at $\Delta = 0$, increases to a maximum, and then decreases towards 0. The risk characteristic $\widehat{\boldsymbol{\theta}}^{SP}$ is similar to $\widehat{\boldsymbol{\theta}}^P$. Interestingly, the bias curve of $\widehat{\boldsymbol{\theta}}^{SP}$ remains below the curve of $\widehat{\boldsymbol{\theta}}^P$ for all values of Δ.

Theorem 2. If Proposition 2.5 holds, then, the ADR *functions of the estimators are given as follows:*

$$\begin{aligned}
\mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right) &= \left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\mathrm{trace}\left(\mathbf{W}\boldsymbol{\Sigma}\right),\\
\mathrm{ADR}\left(\widetilde{\boldsymbol{\theta}}_T;\mathbf{W}\right) &= \mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right) - \left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\mathrm{trace}\left(\mathbf{W}\boldsymbol{\Sigma}^*\right) + \boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^*,\\
\mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}^S;\mathbf{W}\right) &= \mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right) - \beta(2-\beta)\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\mathrm{trace}(\mathbf{W}\boldsymbol{\Sigma}^*) + \beta^2\boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^*\\
\mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}^P;\mathbf{W}\right) &= \mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right) - \left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\mathrm{trace}\left(\boldsymbol{W}\boldsymbol{\Sigma}^*\right)H_{q+2}(\chi^2_{q,\alpha};\Delta)\\
&\quad +\boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^*\left\{2H_{q+2}\left(\chi^2_{q,\alpha};\Delta\right) - H_{q+4}\left(\chi^2_{q,\alpha};\Delta\right)\right\}\\
\mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}^{SP};\mathbf{W}\right) &= \mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right) - \beta(2-\beta)\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\mathrm{trace}\left(\boldsymbol{W}\boldsymbol{\Sigma}^*\right)H_{q+2}\left(\chi^2_{q,\alpha};\Delta\right)\\
&\quad +\boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^*\left\{2\beta H_{q+2}\left(\chi^2_{q,\alpha};\Delta\right) - \beta(2-\beta)H_{q+4}\left(\chi^2_{q,\alpha};\Delta\right)\right\}
\end{aligned} \tag{3.4}$$

The proof is similar to that given in Ahmed (2001). Clearly, the ADR of $\widehat{\boldsymbol{\theta}}_T$ is constant (independent of $\boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^*$). If $\beta > 0$, the ADR $\left(\widehat{\boldsymbol{\theta}}^S;\mathbf{W}\right)$ is a straight line in terms of $\boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^*$ which intersects the ADR $\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right)$ whenever $\boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^* = (2-\beta)\mathrm{trace}\left(\boldsymbol{W}\boldsymbol{\Sigma}^*\right)/\beta$. At and near the null hypothesis the ADR of $\widehat{\boldsymbol{\theta}}^S$ is less than the ADR of $\widehat{\boldsymbol{\theta}}_T$.

Comparing ADR $\left(\widehat{\boldsymbol{\theta}}^{SP};\mathbf{W}\right)$ with ADR $\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right)$,

$$\mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}^{SP};\mathbf{W}\right) \leqslant \mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right)\ ,\ \text{when}$$

$$\Delta \leqslant (2-\beta)\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\ \mathrm{trace}\left(\mathbf{W}\boldsymbol{\Sigma}^*\right)H_{q+2}(\chi^2_{q,\alpha};\Delta)\left\{2H_{q+2}(\chi^2_{q,\alpha};\Delta) - (2-\beta)H_{q+4}(\chi^2_{q,\alpha};\Delta)\right\}^{-1}. \tag{3.5}$$

Further, as α, the level of the statistical significance, approaches 1, ADR $\left(\widehat{\boldsymbol{\theta}}^{SP};\mathbf{W}\right)$ converges to ADR $\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right)$. Also, when Δ increases and tends to infinity, the ADR $\left(\widehat{\boldsymbol{\theta}}^{SP};\mathbf{W}\right)$ approaches the ADR $\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right)$. Broadly speaking, for larger values of Δ, the value of the ADR $\left(\widehat{\boldsymbol{\theta}}^{SP};\mathbf{W}\right)$ increases, reaches its maximum after crossing the ADR $\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right)$ and then monotonically decreases and approaches the ADR $\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right)$. On the other hand,

$$\mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}^{P};\mathbf{W}\right) \leqslant \mathrm{ADR}\left(\widehat{\boldsymbol{\theta}}_T;\mathbf{W}\right)\quad \text{when}$$

$$\Delta \leqslant \left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\ \mathrm{trace}\left(\mathbf{W}\boldsymbol{\Sigma}^*\right)H_{q+2}(\chi^2_{q,\alpha};\Delta)\left\{2H_{q+2}(\chi^2_{q,\alpha};\Delta) - H_{q+4}(\chi^2_{q,\alpha};\Delta)\right\}^{-1}. \tag{3.6}$$

By comparing the right hand side of equation (3.5) to the right hand side of (3.6), we noticed that the range of the parameter space in (3.6) is smaller than that in (3.5).

Noting that under the null hypothesis, $\widehat{\boldsymbol{\theta}}^{P}$ dominates $\widehat{\boldsymbol{\theta}}^{SP}$. However, the picture is somewhat different when the hypothesis error grows. As the hypothesis error increases then $\widehat{\boldsymbol{\theta}}^{SP}$ performs better than $\widehat{\boldsymbol{\theta}}^{P}$. Denote $\boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^{*} = \Delta^{*}$ and for a given β, let Δ^{*}_{β} be a point in the parameter at which the risk of $\widehat{\boldsymbol{\theta}}^{SP}$ and $\widehat{\boldsymbol{\theta}}^{P}$ intersect. Then for $\Delta^{*} \in (0, \Delta^{*}_{\beta}]$, $\widehat{\boldsymbol{\theta}}^{P}$ performs better than $\widehat{\boldsymbol{\theta}}^{SP}$, while for $\Delta^{*} \in [\Delta^{*}_{\beta}, \infty)$, $\widehat{\boldsymbol{\theta}}^{SP}$ dominates $\widehat{\boldsymbol{\theta}}^{P}$. Further for large values of β (close to 1) the $(0, \Delta^{*}_{\beta}]$ may not be significant. Nonetheless, $\widehat{\boldsymbol{\theta}}^{P}$ and $\widehat{\boldsymbol{\theta}}^{SP}$ share a common asymptotic property that as the hypothesis error grows and tends to ∞, their ADRs converge to a common limit, that is, to the ADR of $\widehat{\boldsymbol{\theta}}_{T}$.

Thus, in the light of above findings we suggest to use $\widehat{\boldsymbol{\theta}}^{SP}$, since it has a good control on ADR when the restriction may not be judiciously imposed.

In the following section, we consider the non-interdependence case.

4 The Non-interdependence Model

In this section, we briefly give a special case for which the drift terms of the model in (2.1) do not contain interdependence parameters and that corresponds to the case where the matrix parameter $\boldsymbol{\theta}$ is a diagonal matrix. Note that this case is also presented in Nkurunziza and Ahmed (2010). We consider the estimation problem of a p-column vector $(\theta_{11}, \theta_{22}, \ldots, \theta_{pp})'$ when it is suspected that

$$\theta_{11} = \theta_{22} = \cdots = \theta_{pp}. \tag{4.1}$$

For this case, the notation $\boldsymbol{\theta}$ denotes the p-column parameter vector $(\theta_{11}, \theta_{22}, \ldots, \theta_{pp})'$, and then, for this particular case, we set for the simplicity sake, $(\theta_{11}, \theta_{22}, \ldots, \theta_{pp})' = (\theta_{1}, \theta_{2}, \ldots, \theta_{p})' = \boldsymbol{\theta}$. Accordingly, the constraint (4.1) has the same form as first constraint in (1.5) with $\boldsymbol{d}$ a $p-1$-column zero vector and

$$\boldsymbol{L}_1 = \begin{pmatrix} 1 & -1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & -1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -1 \end{pmatrix}_{(p-1)\times p} .$$

To make a connection with the problem studied in Section 2, note that here $q = p-1$. Also, let

$$\boldsymbol{V}(t) = \mathrm{diag}\left(X_1(t), X_2(t), \ldots, X_p(t)\right), \quad 0 \leqslant t \leqslant T.$$

From relation (2.1), we get

$$d\boldsymbol{X}(t) = -\boldsymbol{V}(t)\boldsymbol{\theta}dt + \boldsymbol{\Sigma}^{\frac{1}{2}}d\boldsymbol{W}(t), \quad 0 \leqslant t \leqslant T. \tag{4.2}$$

Finally, based on constraint in (4.1),

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_p \text{ versus } H_1 : \theta_j \neq \theta_k \text{ for some } 1 \leqslant j < k \leqslant p. \tag{4.3}$$

Further, let

$$\widetilde{\boldsymbol{U}}_T = \operatorname{diag}(\boldsymbol{U}_T) = \left(\int_0^T X_1(t)dX_1(t), \int_0^T X_2(t)dX_2(t), \ldots, \int_0^T X_p(t)dX_p(t)\right)',$$

and let

$$\widetilde{\boldsymbol{D}}_T = \operatorname{diag}(\boldsymbol{D}_T) = \int_0^T \boldsymbol{V}^2(t)dt, \quad \widehat{\boldsymbol{\theta}}_T^* = -\widetilde{\boldsymbol{D}}_T^{-1}\widetilde{\boldsymbol{U}}_T = \left(\widehat{\theta}_1^*, \widehat{\theta}_2^*, \ldots, \widehat{\theta}_p^*\right)'. \tag{4.4}$$

Proposition 4.1. *Let $\boldsymbol{e}_p$ be a p-column vector whose all entrees are equal to 1. We have*

$$\widehat{\boldsymbol{\theta}}_T = \left(\widehat{\theta}_1^{*+}, \widehat{\theta}_2^{*+}, \ldots, \widehat{\theta}_p^{*+}\right)' \text{ and } \widetilde{\boldsymbol{\theta}}_T = \left(\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{D}}_T\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{D}}_T\widehat{\boldsymbol{\theta}}_T. \tag{4.5}$$

Under this special model and for the $\boldsymbol{\Sigma}$ known case, we get the following test statistic,

$$\varphi(T) = T\ \left(\widehat{\boldsymbol{\theta}}_T - \widetilde{\boldsymbol{\theta}}_T\right)'\boldsymbol{\Sigma}^{-1}\ \left(\widehat{\boldsymbol{\theta}}_T - \widetilde{\boldsymbol{\theta}}_T\right) \tag{4.6}$$

and accordingly, for a given $0 < \alpha < 1$, we suggest the asymptotically α-level test

$$\Psi = I\left(\varphi(T) > \chi^2_{p-1;\alpha}\right), \quad \text{where} \quad \Pr\left\{\chi^2_{p-1} > \chi^2_{p-1;\alpha}\right\} = \alpha. \tag{4.7}$$

As for Section 2, when $\boldsymbol{\Sigma}$ is unknown, the test statistic in (4.6) is modified by replacing $\boldsymbol{\Sigma}$ by a its strongly consistent estimator and again, the obtained new test is asymptotically α-level test, as powerful as the test Ψ .

The shrinkage and pretest estimators will have the same form as that presented in Section 2. We obtain these estimators by replacing the unrestricted, restricted estimators and test statistic of Section 2 by the unrestricted, restricted estimators and test statistic derived in this section. Let $\bar{\theta}$ be the common value of $\theta_1 = \theta_2 = \cdots = \theta_p$ under H_0. For the asymptotic power and derivation of ADR of all estimators, we modify local alternative as

$$K_T : \boldsymbol{\theta} = \bar{\theta}\boldsymbol{e}_p + \frac{\boldsymbol{\delta}}{\sqrt{T}} \tag{4.8}$$

where $\boldsymbol{\delta}$ is a p-column vector with different direction than $\boldsymbol{e}_p$. Also, we assume that $\|\boldsymbol{\delta}\| < \infty$. With the following minor adjustment in the asymptotic results of Section 2, we get parallel expressions for ADB and ADR results, and the performance of the estimators remain intact. We denote

$$\underline{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\boldsymbol{V}_0^{-1}, \quad \Upsilon = \left(\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p', \quad \boldsymbol{\delta}^* = \boldsymbol{\delta} - \Upsilon\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\delta}, \quad \underline{\boldsymbol{\Sigma}}^* = \underline{\boldsymbol{\Sigma}} - \Upsilon \quad \text{and} \quad \underline{\boldsymbol{\Sigma}}^- = \underline{\boldsymbol{\Sigma}}^*.$$

In this case, under local alternative and if T tends to infinity, the test statistic in $\varphi(T)$ converges in distribution to a noncentral χ^2 random variable with $p-1$ degrees of freedom ($q = p-1$) and non-centrality parameter $\Delta = \boldsymbol{\delta}^{*'}\boldsymbol{\Xi}\boldsymbol{\delta}^{*}$, where $\boldsymbol{\Xi} = \underline{\boldsymbol{\Sigma}}^{-1} - \underline{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}$.

To save the space, we do not display the expression for ADB and ADR of the estimators under the special cases. Indeed, it suffices to replace q by $p-1$. The performance of shrinkage and pretest estimator is now much superior since $q = p-1$. In other words, we have less parameters to estimate with the same sample information as compared to general estimation problem in Section 2.

5 Simulation Study

In this section, we utilize the theoretical methodology of the proceeding section to determine the practical performance of the shrinkage and pretest estimators relative to $\widehat{\boldsymbol{\theta}}_T$ in a simulating setting. This setting is intended to replace the circumstances as an applied researcher is likely to face in the estimation of $\boldsymbol{\theta}$.

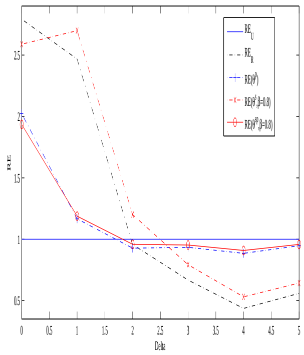
A Monte Carlo simulation experiment is designed to investigate the risk (namely MSE) behavior of the all estimators under consideration. In an effort to conserve space, we report detailed results only for $p = 3$ and $p = 5$ with the shrinkage parameter $\beta = 0.2$, $\beta = 0.5$ and $\beta = 0.8$. Also, we consider the null hypothesis $H_0 : \boldsymbol{\theta} = \overline{\theta}\mathbf{e}_p'$. The length of the time period of observation $T = 15$, $T = 30$ are considered. Also, the chosen value for $\overline{\theta}$ is 0.2 and 2500 replications have been performed.

The relative efficiency of the estimators with respect to $\widehat{\boldsymbol{\theta}}_T$ is defined by $RE = \text{MSE}(\widehat{\boldsymbol{\theta}}_T)/\text{MSE}(\hat{\boldsymbol{\theta}}^l)$, $l = R, S, P, SP$.

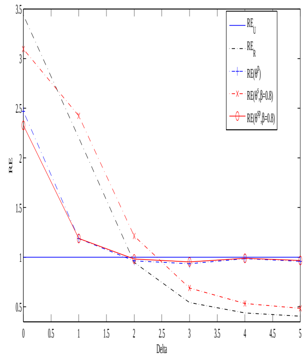
Thus, a relative efficiency larger than one indicates the degree of superiority of the estimator $\hat{\boldsymbol{\theta}}^l$ ($l = R, S, P, SP$) over $\widehat{\boldsymbol{\theta}}_T$. The results are graphically reported in **FIG.** 1(a)-1(f) for $p = 3$ and in **FIG.** 2(a)-2(f) for $p = 5$. Graphically, **FIG.** 1 and **FIG.** 2 indicate that a substantial improvement of 50% or more over $\widehat{\boldsymbol{\theta}}_T$ seems quite realistic depending on the values of Δ, p and β. Our simulation study findings are in agreement with our theoretical results developed in this paper.

6 Conclusion

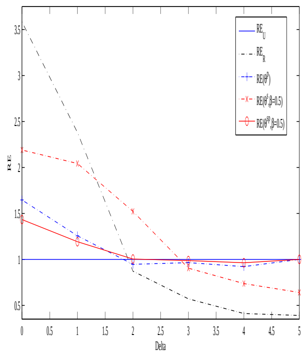
We develop the inferential statistics for drift parameter matrix of a multivariate Ornstein-Uhlenbeck process. The analytical and numerical results presented in this paper should have great applicability to the researchers in this area. For this paper, MLE, shrinkage and pretest estimators are considered for estimation drift parameter matrix of a multivariate Ornstein-Uhlenbeck process. The shrinkage and pretest estimators are presented as alternative to the UMLE and RMLE. We show that $\widehat{\boldsymbol{\theta}}^{SP}$ displays superior performance to the other three estimators for a significant range of the parameter space. On the basis of our simulation experiment results, this range appears to be substantial, in terms of yielding a large reductions in MSE. The results of our simulation experiments also indicate several



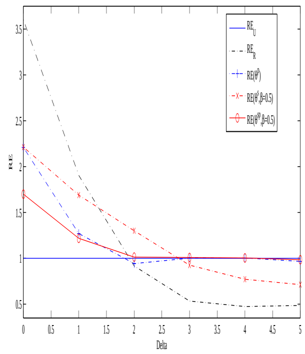
(a) $\beta = 0.8$, $T = 15$ and $p = 3$



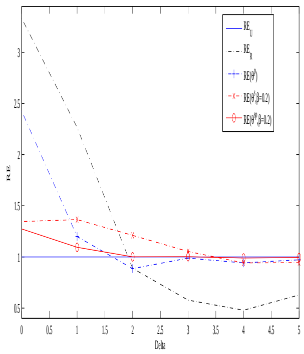
(b) $\beta = 0.8$, $T = 30$ and $p = 3$



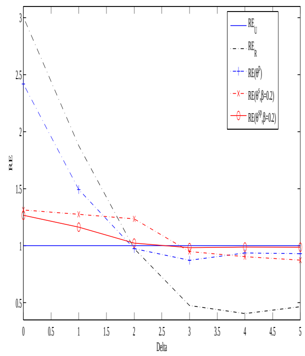
(c) $\beta = 0.5$, $T = 15$ and $p = 3$



(d) $\beta = 0.5$, $T = 30$ and $p = 3$

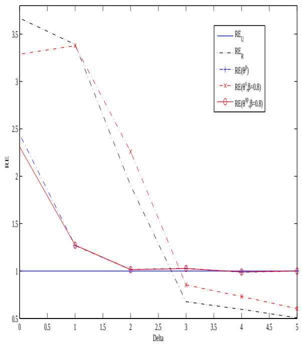


(e) $\beta = 0.2$, $T = 15$ and $p = 3$

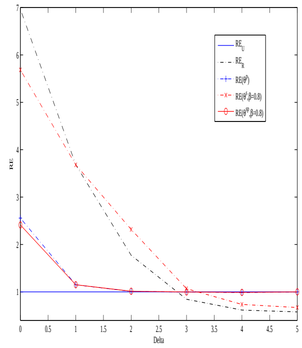


(f) $\beta = 0.2$, $T = 30$ and $p = 3$

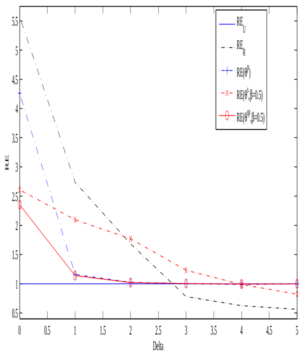
Figure 1: Relative efficiency vs Δ



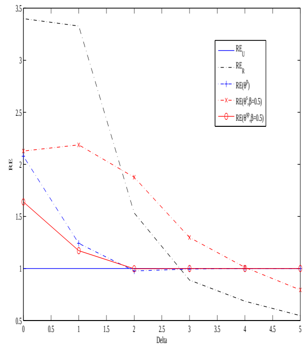
(a) $\beta = 0.8$, $T = 15$ and $p = 5$



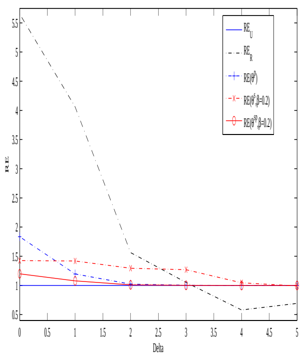
(b) $\beta = 0.8$, $T = 30$ and $p = 5$



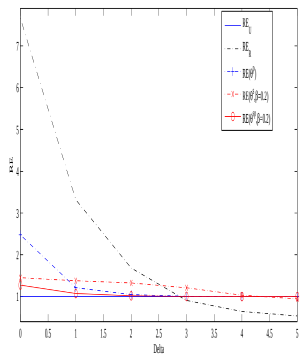
(c) $\beta = 0.5$, $T = 15$ and $p = 5$



(d) $\beta = 0.5$, $T = 30$ and $p = 5$



(e) $\beta = 0.2$, $T = 15$ and $p = 5$



(f) $\beta = 0.2$, $T = 30$ and $p = 5$

Figure 2: Relative efficiency vs Δ

factors for which an applied researcher would wish to take into account in practical situations. The simulation results also provide largely good news about the behavior of suggested estimators. In summary, our simulation study provides strong evidence that corroborates with the developed asymptotic theory related to MLE , RMLE and shrinkage and pretest estimators.

Acknowledgements

This research received financial support from Natural Sciences and Engineering Research Council of Canada. Also, the authors would like to thank the referee for helpful comments.

A Some technical details

Let μ_U be the probability measure induced by process $\{U(t), t \geq 0\}$, and let $\mu_{\boldsymbol{W}}$ be the measure induced by the Wiener process $\{\boldsymbol{W}(t), t \geq 0\}$. We have $\mu_U(B) = \mathrm{P}\{\omega : U_t(\omega) \in B\}$, where B is a Borel set. The following result plays a central role in establishing the MLE of $\boldsymbol{\theta}$.

Proposition A.1. *Conditionally to* $\boldsymbol{X}_0$*, the Radon-Nicodym derivative of* $\mu_{\boldsymbol{X}}$ *with respect to* $\mu_{\boldsymbol{W}}$ *is given by*

$$\frac{d\mu_{\boldsymbol{X}}}{d\mu_{\boldsymbol{W}}}(\boldsymbol{X}) = \exp\left\{-\mathrm{trace}\left(\boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_T\right) - \frac{1}{2}\mathrm{trace}\left(\boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\boldsymbol{D}_T\right)\right\}. \tag{A.1}$$

The proof is straightforward by applying Theorem 7.7 in Liptser and Shiryayev (1977). It should be noted that, the relation (A.1) has the same form as the equation 17.24 of Liptser and Shiryayev (1977) for the univariate case with the non-random initial value.

A.1 The interdependence model case

In univariate case, Lipter and Shiryayev (1978, Theorem 17.3, Lemma 17.3 and Theorem 17.4) and Kutoyants (2004) give some asymptotic results for the maximum likelihood estimator of the drift parameter of an Ornstein-Uhlenbech process. The ideas of proof are the same as given in the quoted references. However, the multidimensional case requires some additional attention and technicality. Because of that, and for the completeness, we outline the proofs of the main results. To this end, let

$$\boldsymbol{\varrho}_T = \sqrt{T}\ \left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right)\boldsymbol{L}_2, \quad \boldsymbol{\xi}_T = \sqrt{T}\ \left(\widehat{\boldsymbol{\theta}}_T - \widetilde{\boldsymbol{\theta}}_T\right)\boldsymbol{L}_2, \quad \boldsymbol{\zeta}_T = \sqrt{T}\ \left(\widetilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right)\boldsymbol{L}_2, \quad \Upsilon = \boldsymbol{\Sigma}\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2,$$

and let $\boldsymbol{\Sigma}^* = \boldsymbol{J}\boldsymbol{L}_1\boldsymbol{\Sigma}$.

Note that, the covariance-variance matrix of invariant distribution can be written as $\boldsymbol{C}\boldsymbol{\Sigma}$ where $\boldsymbol{C}$ is a positive definite matrix. Thus, through a linear transformation, one can take $\boldsymbol{C}$ as the identity matrix without loss of generalities. In particular, without loss of generalities, we assume that $\boldsymbol{X}_0$ has an invariant distribution with $\boldsymbol{V}_0 = \boldsymbol{\Sigma}$. However, under this assumption the estimator of $\boldsymbol{\theta}$ in (2.6) may no longer be MLE. Nevertheless, this estimator is a quasi-likelihood and preserves the same asymptotic optimality. Accordingly, in the sequel and in order to simplify the mathematical analysis, we work with $\boldsymbol{V}_0 = \boldsymbol{\Sigma}$.

Proposition A.2. *Assume that Proposition 2.4 holds. Then, under H_0 given in (2.3), we have*

$$\begin{pmatrix} \boldsymbol{\varrho}_T \\ \boldsymbol{\xi}_T \end{pmatrix} \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}_{2p}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}^* \\ \boldsymbol{\Sigma}^* & \boldsymbol{\Sigma}^* \end{pmatrix}\right) \quad \textit{and}$$

$$\begin{pmatrix} \boldsymbol{\zeta}_T \\ \boldsymbol{\xi}_T \end{pmatrix} \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}_{2p}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\begin{pmatrix} \boldsymbol{\Sigma}-\boldsymbol{\Sigma}^* & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^* \end{pmatrix}\right).$$

Proof. Under H_0, we have

$$\left(\boldsymbol{\varrho}_T', \boldsymbol{\xi}_T'\right)' = \left(I_p, \boldsymbol{L}_1'\boldsymbol{J}'\right)' \sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right)\boldsymbol{L}_2, \tag{A.2}$$

and then, the rest of proof is similar to that given for Proposition 2.5 and that completes the proof. □

A.1.1 Proof of Corollary 2.1

From Proposition 2.5, under local alternative, we have

$$\boldsymbol{\xi}_T \xrightarrow[T\to\infty]{\mathcal{L}} \boldsymbol{Z} \sim \mathcal{N}_p\left(\boldsymbol{\delta}^*, \boldsymbol{\Sigma}^*\right). \tag{A.3}$$

Therefore, under local alternative (2.9), $\boldsymbol{\xi}_T'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\boldsymbol{\xi}_T \xrightarrow[T\to\infty]{\mathcal{L}} \boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\boldsymbol{Z}$. Moreover, one can verify that

$$\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\boldsymbol{\Sigma}^*\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\boldsymbol{\Sigma}^*\right)^2 = \left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\boldsymbol{\Sigma}^*,$$

$$\begin{aligned} \operatorname{rank}\left[\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\boldsymbol{\Sigma}^*\right] &= q, \\ \boldsymbol{\delta}^{*\prime}\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\boldsymbol{\Sigma}^*\right)^2 &= \boldsymbol{\delta}^{*\prime}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\boldsymbol{\Sigma}^*, \end{aligned}$$

and that

$$\boldsymbol{\delta}^{*\prime}\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\right)\boldsymbol{\delta}^{*\prime} = \boldsymbol{\delta}^{*\prime}\boldsymbol{\Xi}\boldsymbol{\delta}^*.$$

Therefore, by using Theorem 4 in Styan (1970), we get $\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\boldsymbol{Z} \sim \chi_q^2\left(\boldsymbol{\delta}^{*\prime}\boldsymbol{\Xi}\boldsymbol{\delta}^*\right)$. Moreover, under H_0 in (2.3), by rewriting the same steps and by replacing $\boldsymbol{\delta}^*$ with $\mathbf{0}$, we prove that $\boldsymbol{\xi}_T'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{L}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_2\right)^{-1}\boldsymbol{\xi}_T \xrightarrow[T\to\infty]{\mathcal{L}} \chi_q^2$, that completes the proof. □

A.2 The non-interdependence model case

Proposition A.3. *If the model (2.2) holds and if $\boldsymbol{X}_0$ has the same moment as a r.v. that follows the invariant distribution, then*

(i) uniformly in $\boldsymbol{\theta}$ on every compact subset of $\mathbb{R}_+^{*p}$,

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}_p\left(\mathbf{0},\, \underline{\boldsymbol{\Sigma}}\right), \text{ and } T\ \left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right)' \underline{\boldsymbol{\Sigma}}^{-1}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right) \xrightarrow[T\to\infty]{\mathcal{L}} \chi_p^2;$$

(ii) if $\theta_1 = \theta_2 = \cdots = \theta_p$, we have, uniformly in $\overline{\theta}$ on every compact subset of $\mathbb{R}_+^*$,

$$\sqrt{T}\left(\widetilde{\boldsymbol{\theta}}_T - \bar{\theta}\boldsymbol{e}_p\right) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}_p\left(\mathbf{0},\, \boldsymbol{e}_p\boldsymbol{e}_p'\left(\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}_p\right)^{-1}\right).$$

Proof. The proof follows directly from Theorem 2.8 in Kutoyants (2004, p. 121) and that completes the proof. □

Let

$$\boldsymbol{\varrho}_T = \sqrt{T}\ \left(\widehat{\boldsymbol{\theta}}_T - \bar{\theta}\boldsymbol{e}_p\right),\, \boldsymbol{\xi}_T = \sqrt{T}\ \left(\widehat{\boldsymbol{\theta}}_T - \widetilde{\boldsymbol{\theta}}_T\right),\, \zeta_T = \sqrt{T}\ \left(\widetilde{\boldsymbol{\theta}}_T - \bar{\theta}\boldsymbol{e}_p\right),\, \Upsilon = \left(\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'$$

and let $\boldsymbol{\Sigma}^* = \underline{\boldsymbol{\Sigma}} - \Upsilon$.

Proposition A.4. *Assume that Proposition A.3 holds.*

(i) Under the local alternative hypothesis as given in (4.8), we have

$$\begin{pmatrix} \boldsymbol{\varrho}_T \\ \boldsymbol{\xi}_T \end{pmatrix} \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}_{2p}\left(\begin{pmatrix} \boldsymbol{\delta} \\ \boldsymbol{\delta}^* \end{pmatrix}, \begin{pmatrix} \underline{\boldsymbol{\Sigma}} & \boldsymbol{\Sigma}^* \\ \boldsymbol{\Sigma}^* & \boldsymbol{\Sigma}^* \end{pmatrix}\right) \quad \text{and}$$
$$\begin{pmatrix} \boldsymbol{\zeta}_T \\ \boldsymbol{\xi}_T \end{pmatrix} \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}_{2p}\left(\begin{pmatrix} \boldsymbol{\delta} - \boldsymbol{\delta}^* \\ \boldsymbol{\delta}^* \end{pmatrix}, \begin{pmatrix} \Upsilon & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^* \end{pmatrix}\right).$$

(ii) Under H_0 given in (4.3), we have

$$\begin{pmatrix} \boldsymbol{\varrho}_T \\ \boldsymbol{\xi}_T \end{pmatrix} \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}_{2p}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \underline{\boldsymbol{\Sigma}} & \boldsymbol{\Sigma}^* \\ \boldsymbol{\Sigma}^* & \boldsymbol{\Sigma}^* \end{pmatrix}\right) \quad \text{and}$$
$$\begin{pmatrix} \boldsymbol{\zeta}_T \\ \boldsymbol{\xi}_T \end{pmatrix} \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}_{2p}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Upsilon & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^* \end{pmatrix}\right).$$

Proof. (i) By some computations, we get

$$\begin{aligned}(\boldsymbol{\varrho}_T', \boldsymbol{\xi}_T')' = &\left(I_p,\, I_p - \left(\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T\boldsymbol{e}_p\right)^{-1}\boldsymbol{D}_T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\right)'\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right) \\ &+ \left(I_p,\, I_p - \left(\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T\boldsymbol{e}_p\right)^{-1}\boldsymbol{D}_T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\right)'\boldsymbol{\delta}.\end{aligned}$$

By using Proposition A.3 and the Slutsky theorem, we get the first statement in *(i)*. Further, in the similar way, we have

$$\begin{pmatrix} \boldsymbol{\zeta}_T \\ \boldsymbol{\xi}_T \end{pmatrix} = \begin{pmatrix} \left(\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T \\ I_p - \left(\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T \end{pmatrix} \sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right) + \begin{pmatrix} \left(\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T \\ I_p - \left(\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T \end{pmatrix}\boldsymbol{\delta}.$$

Again, from Proposition A.3 and the Slutsky theorem, we get the second statement in *(i)*.

(ii) The proof of (ii) is similar to that given in (i). It suffices to take $\boldsymbol{\delta} = \mathbf{0}$ and the completes the proof.

□

Corollary A.1. Assume that the conditions of Proposition A.4 holds. Then, under the local alternative hypothesis as given in (4.8),

$$\boldsymbol{\xi}_T'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\xi}_T \xrightarrow[T\to\infty]{\mathcal{L}} \chi^2_{p-1}\left(\boldsymbol{\delta}*'\,\Xi\boldsymbol{\delta}*\right).$$

Furthermore, under H_0, we have

$$\boldsymbol{\xi}_T'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\xi}_T \xrightarrow[T\to\infty]{\mathcal{L}} \chi^2_{p-1}. \tag{A.4}$$

Proof. We have

$$\boldsymbol{\xi}_T'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\xi}_T = \boldsymbol{\varrho}_T'\left(\underline{\boldsymbol{\Sigma}}^{-1} - \underline{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\varrho}_T + \boldsymbol{\varrho}_T'\left(\Xi_T - \Xi\right)\boldsymbol{\varrho}_T, \tag{A.5}$$

where

$$\Xi_T = \left(I_p - \boldsymbol{D}_T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_p\left(\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T\boldsymbol{e}_p\right)^{-1}\right)\underline{\boldsymbol{\Sigma}}^{-1}\left(I_p - \left(\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_T\right), \tag{A.6}$$

and

$$\Xi = \underline{\boldsymbol{\Sigma}}^{-1} - \underline{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}. \tag{A.7}$$

Combining Proposition A.4 and the Slutsky theorem, we deduce that, under the local alternative hypothesis as given in (4.8),

$$\boldsymbol{\varrho}_T'\left[\Xi_T - \Xi\right]\boldsymbol{\varrho}_T \xrightarrow[T\to\infty]{\mathcal{L}} \mathbf{0} \quad \text{and} \quad \boldsymbol{\varrho}_T \xrightarrow[T\to\infty]{\mathcal{L}} \boldsymbol{Z} \sim \mathcal{N}_p\left(\boldsymbol{\delta}, \underline{\boldsymbol{\Sigma}}\right). \tag{A.8}$$

Moreover, $\underline{\boldsymbol{\Sigma}}\left(\underline{\boldsymbol{\Sigma}}^{-1} - \underline{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\right) = \underline{\boldsymbol{\Sigma}}\Xi$ is an idempotent matrix. Therefore,

$$\boldsymbol{\varrho}_T'\left(\underline{\boldsymbol{\Sigma}}^{-1} - \underline{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\right)\boldsymbol{\varrho}_T \xrightarrow[T\to\infty]{\mathcal{L}} \boldsymbol{Z}'\Xi\boldsymbol{Z} \sim \chi^2_r\left(\boldsymbol{\delta}'\Xi\boldsymbol{\delta}\right),$$

where

$$r = \text{rank}\,(\boldsymbol{\Xi}) = \text{trace}\left(\underline{\boldsymbol{\Sigma}}\left(\underline{\boldsymbol{\Sigma}}^{-1} - \underline{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}_p\right)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p'\underline{\boldsymbol{\Sigma}}^{-1}\right)\right) = p-1 \text{ and } \boldsymbol{\delta}'\boldsymbol{\Xi}\boldsymbol{\delta} = \boldsymbol{\delta}^{*'}\boldsymbol{\Xi}\boldsymbol{\delta}^{*} \quad (A.9)$$

Further, under H_0 in (4.3), since the relation in (A.8) holds with $\boldsymbol{\delta} = \mathbf{0}$, the statement (A.4) holds too and that completes the proof. □

References

[1] Abu-Mostafa, Y. S. (2001). Financial Model Calibration Using Consistency Hints. *IEEE transactions on neural networks,* **12***, 4, 791–808.*

[2] Ahmed, S. E., Hussein, A. A. and Sen, P. K. (2006). Risk comparison of some shrinkage M-estimators in linear models. *Journal of Nonparametric Statistics,* **18***, 401–415.*

[3] Ahmed, S. E. (2005). Assessing Process Capability Index for Nonnormal Processes. *Journal of Statistical Planning and Inference,* **129***, 195–206.*

[4] Ahmed, S. E. and Krzanowski, W. J. (2004). Biased estimation in a simple multivariate regression model. *Computational Statistics and Data Analysis,* **45***, 689–696.*

[5] Ahmed, S. E. (2001). *Shrinkage estimation of regression coefficients from censored data with multiple observations.* In S.E. Ahmed and N. Reid (Eds.), Empirical Bayes and and Likelihood inference (pp. 103–120). New York : Springer.

[6] Bancroft, T.A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Annals of Mathematical Statistics,* **15***, 190–204.*

[7] Basawa, V. I. and Rao, P. B. L. S. (1980). *Statistical inference for stochastic processes. London: Academic Press.*

[8] Bickel, V. P. and Doksum K. A. (2001). *Mathematical Statistics: Basic Ideas and Selected Topics, vol. 1,* 2^{nd} *ed. Prentice Hall.*

[9] Chiou, P. and Miao, W. (2007). Shrinkage estimation for the difference between a control and treatment mean. *Journal of Statistical Computation and Simulation,* **77***, 651–662.*

[10] Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics,* **19***, 263–284.*

[11] Danilov, D. and Magnus, J. R. (2004). On the harm that ignoring pretesting can cause. *J. Econometrics,* **122***, 27–46.*

[12] Engen, S. and Saether, B. E. (2005). Generalizations of the Moran Effect Explaining Spatial Synchrony in Population Fluctuations. *The American Naturalist,* **166***, 603–612.*

[13] Engen, S., Lande, R., Wall, T. and DeVries, J. P. (2002). Analyzing Spatial Structure of Communities Using the Two-Dimensional Poisson Lognormal Species Abundance Model. *The American Naturalist,* **160**,*1, 60–73.*

[14] Florens-Zmirou, D. (1989). Approximation discrete-time schemes for statistics of diffusion processes. *Statistics,* **20**, *547–557.*

[15] Friedman, J. H. (1989) . Regularized Discriminant Analysis. *J. Amer. Statist. Assoc.,* **84**, *405, 165–175.*

[16] Georges, P. (2003). *The Vasicek and CIR Models and the Expectation Hypothesis of the Interest Rate Term Structure.* The Bank of Canada and Department of Finance, Canada.

[17] Khan, B. U. and Ahmed, S. E. (2006). Comparisons of improved risk estimators of the multivariate mean vector. *Computational Statistics and Data Analysis,* **50**, *402–421.*

[18] Kibria, B. M. G. and Saleh, A. K. Md. E. (2006). Optimum critical value for pretest estimators. *Communications in Statistics-Simulation and Computation,* **35**, *2, 309–319.*

[19] Kutoyants, A. Y. (2004). *Statistical Inference for Ergodic Diffusion Processes.* New York: Springer. Langetieg, T. C. (1980). A Multivariate Model of the Term Structure. *The Journal of Finance,* **35**, *1, 71–97.*

[20] Le Breton, A. (1976). Stochastic Systems : Modeling, Identification and Optimization, I. *In Mathematical Programming Studies,* **5**, *124–144.*

[21] Leeb, H. (2003). The distribution of a linear predictor after model selection: Conditional finite sample distributions and asymptotic approximations. *Journal of Statistical Planning and Inference,* **134**, *64–89.*

[22] Liptser, R. S. and Shiryayev, A. N. (1978). *Statistics of Random Processes : Applications*, Vol. II. New York: Springer-Verlag.

[23] Liptser, R. S. and Shiryayev, A. N. (1977). *Statistics of Random Processes : Generale Theory*, Vol. I. New York: Springer-Verlag.

[24] Nkurunziza, S. and Ahmed, S. E. (2007). Shrinkage Drift Parameter Estimation for Multi-factor Ornstein-Uhlenbeck Processes. *University of Windsor. Technical report, WMSR* **07-06**.

[25] Nkurunziza, S. (2010). Testing Concerning the Homogeneity of Some Predator-Prey Populations. *Journal of Statistical Planning and Inference,* **140**, *323–333.*

[26] Nkurunziza, S. and Ahmed, S. E. (2010). Shrinkage drift parameter estimation for multi-factor Ornstein-Uhlenbeck processes. *Appl. Stochastic Models Bus. Ind.* **26**, *2, 103–124.*

[27] Pardoa, L. and Men'endez, M. L. (2008). On some pre-test and Stein-rule phi-divergence test estimators in the independence model of categorical data. *Journal of Statistical Planning and Inference,* **138**, *2163–2179.*

[28] Reif, J. (2006). Pitman Closeness in Classes of General Pre-Test Estimators and Regression Estimators. *Communications in Statistics-Theory and Methods,* **35**, *263–279.*

[29] Saleh, A. K. Md. E. (2006). *Preliminary test and Stein-Type Estimation with Applications.* John Wiley & Sons, New York.

[30] Schbel, R. and Zhu, J. (1999). Stochastic Volatility With an Ornstein-Uhlenbeck Process: An Extension. *European Finance Review,* **3**, *23–46.*

[31] Sen, P.K. (1986). On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators. *Sankhya, Series A,* **48**, *354–371.*

[32] Styan, G. P. H. (1970). Notes on the distribution of quadratic forms in singular normal variables. *Biometrika,* **57**,*3, 567–572.*

[33] Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics,* **5**, *177–188.*

[34] Zoua, G., Wan A. T. K., Wub, X., and Chen, T. (2007). Estimation of regression coefficients of interest when other regression coefficients are of no interest: The case of non-normal errors. *Statistics & Probability Letters,* **77**, *803–810.*