Bernoulli **21**(4), 2015, 2217–2241
DOI: 10.3150/14-BEJ642

Optimal method in multiple regression with structural changes

FUQI CHEN* and SÉVÉRIEN NKURUNZIZA**

Mathematics and Statistics Department, University of Windsor, 401 Sunset Avenue, Windsor, Ontario N9B 3P4, Canada. E-mail: *chen111n@uwindsor.ca; **severien@uwindsor.ca

In this paper, we consider an estimation problem of the regression coefficients in multiple regression models with several unknown change-points. Under some realistic assumptions, we propose a class of estimators which includes as a special cases shrinkage estimators (SEs) as well as the unrestricted estimator (UE) and the restricted estimator (RE). We also derive a more general condition for the SEs to dominate the UE. To this end, we generalize some identities for the evaluation of the bias and risk functions of shrinkage-type estimators. As illustrative example, our method is applied to the "gross domestic product" data set of 10 countries whose USA, Canada, UK, France and Germany. The simulation results corroborate our theoretical findings.

Keywords: ADB; ADR; change-points; multiple regression; pre-test estimators; restricted estimator; shrinkage estimators; unrestricted estimator

1. Introduction

In this paper, we study the multivariate regression models with multiple change-points occurring at unknown times. The target parameters are the regression coefficients while the unknown change points are treated as nuisance parameters. More specifically, we are interested in scenario where imprecise prior information about the regression coefficients is available, that is, the target parameters may satisfy some restrictions.

The importance of change-points' model in literature is a primary source of our motivation. Indeed, the regression model with change-points has been applied in many fields. For example, this model was used in Broemeling and Tsurumi [4] for the US demand for money, as well as in Lombard [11] for the effect of sudden changes in wind direction of the flight of a projectile. It was also analyse the DNA sequences (see, e.g., Braun and Muller [3] and Fu and Curnow [5,6]). To give some recent references, we quote Bai and Perron [1], Zeileis *et al.* [20], Perron and Qu [16] among others.

More specifically, the method in Perron and Qu [16] is based on a global least squares procedure. Generally, when the restriction holds, the restricted estimator (RE) dominates the unrestricted estimator (UE). However, it is well known that the RE may performs poorly when the restrictions is seriously violated.

Over the years, shrinkage estimation has become a useful tool in deriving the method which combines in optimal way both imprecise prior knowledge from a hypothesized restriction and the sample information. For more details about such a technique, we refer to James and Stein [8],

1350-7265

Baranchick [2], Judge and Bock [9], and the references therein. Also, to give some recent contributions about shrinkage methods, we quote Saleh [18], Nkurunziza and Ahmed [15], Nkurunziza [13] and Tan [19], among others.

To the best of our knowledge, in context of multiple regression model with unknown changes-points, shrinkage method has received, so far, less attention. Thus, we hope to fill this gap by developing a class of shrinkage-type estimators which includes as special cases the UE, RE, James–Stein type and positive shrinkage estimators as well as pre-test estimators. We also prove that the proposed shrinkage estimators (SEs) dominate in mean square error sense the UE. The technique in this paper extends, in two ways the method given in literature.

First, the asymptotic dependance structure between the shrinking factor (i.e., the difference between the UE and the RE) and the RE is more general than that given in the quoted papers. In particular, the asymptotic variance of RE and the asymptotic variance of (UE − Re) are not positive definite matrices as in the problem studied in Judge and Mittelhammer [10]. This is justified by the fact that, since the hypothesized restriction is linear, these quantities are asymptotically equivalent to the nonsurjective linear (equivalent here to noninjective linear) transformations of the UE for which the asymptotic variance is positive definite matrix. In this case, it is impossible for the asymptotic variance of RE or that of (UE − Re) to be positive definite matrix. To make the justification more precise, let $\mathbf{A}$ be a nonrandom $n \times m$-matrix with the rank $n_0 < n$, let B be a nonrandom n-column vector, and let F be n-column random vector whose variance is a positive definite matrix $\boldsymbol{\Psi}$. Further, let $G = AF + B$, that is a nonsurjective linear transformation of the random vector F. Then, $\mathrm{Var}(G) = \mathbf{A}\boldsymbol{\Psi}\mathbf{A}'$ which cannot be a positive definite matrix since $\mathrm{rank}(\mathbf{A}\boldsymbol{\Psi}\mathbf{A}') = n_0 < n$.

Second, we derive a more general condition for the SEs to dominate the UE. To this end, we generalize Theorem 1 and Theorem 2 of Judge and Bock [9] which are useful in computing the bias and the risk functions of shrinkage-type estimators. As far as the underlying asymptotic results are concerned, another difference, with the work in Judge and Mittelhammer [10], consists in the fact that we derived the joint asymptotic normality under weaker conditions than that in the quoted paper. Indeed, in Judge and Mittelhammer [10], the covariance–variance of the error terms is a scalar matrix (see the first paragraph of Section 2 in Judge and Mittelhammer [10]) and thus, the errors term are both homoscedastic and uncorrelated. In addition, in the quoted paper, the regressors are assumed nonrandom. In this paper, the errors term do not need to be homoscedastic and/or uncorrelated, and they may also be nonstationary stochastic processes. Further, the regressors may be random and in addition, they may be correlated with the error terms. In summary, the proposed method is applicable to the statistical model with familiar regularity conditions as assumed in Judge and Mittelhammer [10], see the last sentence of Section 2.4, as well as in unfamiliar regularity conditions for which the dependance structure of the errors and regressors terms is as weak as that of mixingale array. The model considered here takes also an account for the possibility of the change-points phenomenon and, because of this, the derivation of the joint asymptotic normality between the UE and RE is mathematically challenging. Moreover, the established results extend that given for example in Perron and Qu [16].

In concluding this introduction, note that due to the conditions discussed above which are weaker than that in the literature, the construction of shrinkage-type estimators cannot be obtained by applying the results given in the quoted papers. Further, the derivation of the asymptotic distributional risk (ADR) of shrinkage estimators (SEs) is challenging and the instrumental identities in Judge and Bock [9], Theorems 1 and 2, are not useful. This motivated us to generalize these identities. This constitutes one of the aspects of the main results which are significant in reflecting the difference with the quoted works. The second aspect, of the main results which is significant in reflecting the difference with the quoted works, can be viewed from the fact that the established ADR has some extra terms and the risk dominance condition of SEs looks quite complicated.

The rest of this paper is organized as follows. Section 2 describes the statistical model and outlines the proposed estimation strategies. Section 3 gives the joint asymptotic normality of the unrestricted and restricted estimators. In Section 4, we introduce a class of shrinkage-type of estimators for the coefficients and derive its asymptotic distribution risks. Section 5 presents some simulation studies and an illustrative analysis of a real data set. Section 6 gives some concluding remarks and, for the convenience of the reader, technical proofs are given in the Appendix.

2. Statistical model and assumptions

In this section, we present the statistical model as well as the main regularity conditions. As mentioned above, in this paper, we focus on the model with change-points. Nevertheless, the proposed method is useful in linear model without change-points. In this last case, the derivation of the joint asymptotic normality between the RE and UE is not as mathematically involved as in case of the model with change-points.

2.1. The linear model without change-points

We consider the multiple linear regression model with T observations for which the response is a T-column vector $Y = (y_1, \ldots, y_T)'$, the regressors is a $T \times q_0$-matrix $\bar{Z}$, the regression coefficients is a q_0-column vector δ, and the errors term is a T-column vector u. In particular, we have let

$$Y = \bar{Z}\delta + u. \tag{2.1}$$

Further, we consider the scenario where a prior knowledge about δ exists with some uncertainty. More specifically, we consider the case where δ is suspected to satisfy the following restriction

$$R\delta = r, \tag{2.2}$$

where R is a known $k \times q_0$-matrix with rank $k \leq q_0$, and r is a known k-column vector. Under some regularities conditions on the error terms and the regressors, the shrinkage estimator for the parameter δ is available in literature. To give some references, we quote Saleh [18], Hossain *et al.* [7] among others. The shrinkage estimators given in the quoted papers are members of the class of shrinkage estimators which is established in this paper. Further, the established condition for the risk dominance of shrinkage estimators is more general than that given for example, in Saleh [18], Hossain *et al.* [7].

The proposed methodology is applicable to the model in (2.1) and (2.2) provided that the conditions on the error and regressors terms are such that, as T tends to infinity,

1. the matrices $T^{-1}\bar{Z}^{0\prime}\bar{Z}^0$ and $T^{-1}(\bar{Z}^{0\prime}uu'\bar{Z}^0)$ converge in probability to nonrandom $q_0 \times q_0$-positive and definite matrices;
2. $T^{-1/2}\bar{Z}^{0\prime}u$ converges in distribution to a Gaussian random vector whose variance–covariance is the limit in probability of $T^{-1}\bar{Z}^{0\prime}\bar{Z}^0$.

These two points are generally satisfied in classical regression models where the error terms are homoscedastic and independent, with linearly independent regressors. In the sequel, we consider a very general model with change-points and heteroscedastic as well as possibly correlated errors term. The assumptions of the model are discussed in the next subsection.

2.2. The model with change-points

Briefly, we consider the multiple linear regression model with T observations and m unknown breaks points $T_1, \dots, T_m$ with $1 < T_1 < \cdots < T_m < T$. Here, it is important to stress that the number of change-points m is known. For convenience, let $T_0 = 1$ and $T_{m+1} = T$. Namely, let

$$Y = \bar{Z}\delta + u, \tag{2.3}$$

where $Y = (y_1, \dots, y_T)'$ is a vector of T dependent variables, $\bar{Z}$ is a $T \times (m+1)q$-matrix of regressors given by $\bar{Z} = \operatorname{diag}(Z_1, \dots, Z_{m+1})$ with $Z_1 = (z_1, \dots, z_{T_1})'$, and for $j = 2, 3, \dots, m+1$, $\mathbf{Z}_j = (\mathbf{z}_{T_{j-1}+1}, \dots, \mathbf{z}_{T_j})'$, $\mathbf{z}_{T_{i-1}+1}$ is a q-column vector for $i = 1, 2, \dots, m+1$. Here, $u = (u_1, \dots, u_T)'$ is the set of disturbances and δ is the $(m+1)q$ vector of coefficients. Also, let R be a known $k \times (m+1)q$-matrix with rank k, $k \le (m+1)q$ and let r be a known k-column vector. We consider the case where δ may satisfy or not the following restrictions

$$R\delta = r. \tag{2.4}$$

Let $\{T_1^0, \dots, T_m^0\}$ be the true values of the break times $\{T_1, \dots, T_m\}$, and $\bar{Z}^0 = \operatorname{diag}(Z_1^0, \dots, Z_{m+1}^0)$, where $Z_i^0 = (z_{T_{i-1}^0+1}, \dots, z_{T_i^0})'$. Set $\delta = (\delta_1', \delta_2', \dots, \delta_{m+1}')'$ where for $i = 1, 2, \dots, m+1$ δ_i is a q-column vector.

To estimate the unknown parameters $(\delta_1', \dots, \delta_{m+1}', T_1, \dots, T_{m+1})'$ based only on the sample information given in $\{Y, Z\}$, one can use the least squares principle as described, for example in Perron and Qu [16]. Also, in case the restriction in (2.4) holds, it is common to use the restricted least squares methods in order to estimate the target parameter. This gives the restricted estimator (RE) of $(\delta, T_1, \dots, T_m)$. In particular, concerning the change-points, let $\{\tilde{T}_1, \dots, \tilde{T}_m\}$ denote the RE of the true change points from restricted OLS and let $\{\hat{T}_1, \dots, \hat{T}_m\}$ be the unrestricted estimators (UE). Also, let $\hat{\delta}$ and $\tilde{\delta}$ be, respectively, the UE and RE for the regression coefficients δ. Then, following the framework in Perron and Qu [16], let $\mathrm{SSR}_T^R(T_1, \dots, T_m)$ and $\mathrm{SSR}_T^U(T_1, \dots, T_m)$ be the sum of square residuals from the RE and UE OLS regression evaluated at the partition $\{T_1, \dots, T_m\}$, respectively. We have

$$\begin{aligned}
(\tilde{T}_1, \dots, \tilde{T}_m) &= \arg\min_{T_1,\dots,T_m} \mathrm{SSR}_T^R(T_1, \dots, T_m), \\
(\hat{T}_1, \dots, \hat{T}_m) &= \arg\min_{T_1,\dots,T_m} \mathrm{SSR}_T^U(T_1, \dots, T_m).
\end{aligned} \tag{2.5}$$

The optimality of the proposed method is based on the asymptotic properties of the UE and RE. In particular, in Section 3, we establish as a preliminary step the joint asymptotic normality of the UE and RE. To this end, we present below the regularities conditions. To simplify the notation, let the $\mathcal{L}_2$-norm of random matrix X be defined by $\|X\|_2 = (\sum_a \sum_b \mathrm{E}|X_{a,b}|^2)^{1/2}$, and let $\{\mathcal{F}_i, i = 1, 2, \dots\}$ be a filtration. Also, let $\mathrm{o}_p(a)$ denote a random quantity such that $\mathrm{o}_p(a)/a$ converges in probability to 0, let $\mathrm{O}_p(a)$ denote a random quantity such that $\mathrm{O}_p(a)/a$ is bounded in probability. Similarly, let $\mathrm{o}(a)$ denote a nonrandom quantity such that $\mathrm{o}(a)/a$ converges to 0, let $\mathrm{O}(a)$ denote a nonrandom quantity such that $\mathrm{O}(a)/a$ is bounded. We also use the notations $\xrightarrow[T\to\infty]{\mathrm{d}}$ and $\xrightarrow[T\to\infty]{\mathrm{P}}$ to stand for convergence in distribution and convergence in probability respectively.

Assumptions (Regularity conditions).

($\mathcal{A}_1$) *Let $L_p = (T_{p+1}^0 - T_p^0)$, $p = 1, \dots, m$, then $(1/L_p)\sum_{t=T_p^0+1}^{T_p^0+[L_p v]} z_t z_t' \xrightarrow{p} Q_p(v)$ a nonrandom positive definite matrix uniformly in $v \in [0, 1]$. Besides, there exists an $L_0 > 0$ such that for all $L_p > L_0$, the minimum eigenvalues of $(1/L_p)\sum_{t=T_p^0+1}^{T_p^0+L_p} z_t z_t'$ and of $(1/L_p)\sum_{t=T_p^0-L_p}^{T_p^0} z_t z_t'$ are bounded away from 0.*

($\mathcal{A}_2$) *The matrix $\sum_{t=i_1}^{i_2} z_t z_t'$ is invertible for $0 \le i_2 - i_1 \le \varepsilon_0 T$ for some $\varepsilon_0 > 0$.*

($\mathcal{A}_3$) *$T_p^0 = [T\lambda_p^0]$, where $p = 1, \dots, m+1$ and $0 < \lambda_1^0 < \cdots < \lambda_m^0 < \lambda_{m+1}^0 = 1$.*

($\mathcal{A}_4$) *The minimization problem defined by (2.5) is taken over all possible partitions such that $T_i - T_{i-1} > \tau T$ $(i = 1, \dots, m+1)$ for some $\tau > 0$.*

($\mathcal{A}_5$) *For each segment, (T_{p-1}^0, T_p^0), $p = 1, \dots, m+1$, set $X_{pi} = T^{-1/2} z_{T_{p-1}^0+i} u_{T_{p-1}^0+i}$ and set $\mathcal{F}_{p,i} = \mathcal{F}_{T_{p-1}^0+i}$. We assume that $\{X_{pi}, \mathcal{F}_{p,i}\}$ forms a L^2-mixingale array of size $-1/2$. That is, there exist nonnegative constants $\{c_{pi} : i \ge 1\}$ and $\psi(j)$, $j \ge 0$ such that $\psi(j) \downarrow 0$ as $j \to \infty$ and for $i \ge 1$, $j \ge 0$, with*

$$\begin{aligned} \big\|\mathrm{E}(X_{pi}|\mathcal{F}_{p,i-j})\big\|_2 &\le c_{pi}\psi(j), \\ \big\|X_{pi} - \mathrm{E}(X_{pi}|\mathcal{F}_{p,i+j})\big\|_2 &\le c_{pi}\psi(j+1), \qquad \psi(j) = \mathrm{O}\big(j^{-1/2-\varepsilon}\big) \end{aligned}$$

for some $\varepsilon > 0$. Also, let $L_p = T_{p+1}^0 - T_p^0$, and define l_p, b_p and $r_p = [L_p/b_p]$ such that $b_p \ge l_p + 1$, $l_p \ge 1$, $b_p \le L_p$. We assume that as $b_p \xrightarrow[L_p\to\infty]{} \infty$, $l_p \xrightarrow[L_p\to\infty]{} \infty$, $b_p/L_p \to 0$, and $l_p/b_p \to 0$.

($\mathcal{A}_6$) *For $p = 1, \dots, m+1$, for $s = 1, \dots, q$, $\{X_{pi,s}^2/c_{pi}^2, i = 1, 2, \dots\}$ is uniformly integrable;*

$$\max_{1\le i\le L_p} c_{pi} = \mathrm{o}\big(b_p^{-1/2}\big); \qquad \sum_{i=1}^{r_p}\Big(\max_{(i-1)b_p+1\le t\le ib_p} c_{pt}\Big)^2 = \mathrm{O}\big(b_p^{-1}\big)$$

and

$$\sum_{i=1}^{r_p}\left(\sum_{t=(i-1)b_p+l_p+1}^{ib_p} X_{pt}\right)\left(\sum_{t=(i-1)b_p+l_p+1}^{ib_p} X_{pt}\right)' \xrightarrow[L_p\to\infty]{p} \Sigma_p.$$

Moreover, let $V_{j,i}=\sum_{t=(i-1)b_j+l_j+1}^{ib_j} X_{j,t}$, $j=1,2,\ldots,m+1$. *Let* $r_{(1)}=\min_{1\le j\le m}(r_{p_j})$, *let* $r_{(m)}=\max_{1\le j\le m}(r_{p_j})$, *and let* $L_{\min}=\min(L_1,\ldots,L_{m+1})$. *We have*

1. $\sum_{i=r_{(1)}+1}^{r_{(m)}}(\max_{(i-1)b_j+1\le t\le ib_j} c_{jt})^2=\mathrm{o}(b_j^{-1})$, $j=1,2,\ldots,m+1$.
2. $\sum_{i=1}^{r_{(1)}}(V'_{1,i},V'_{2,i},\ldots,V'_{m+1,i})'(V'_{1,i},V'_{2,i},\ldots,V'_{m+1,i})\xrightarrow[L_{\min}\to\infty]{p}\Omega$, *where* Ω *is nonrandom positive definite matrix.*

For the interpretation of Assumptions $(\mathcal{A}_1)$–$(\mathcal{A}_4)$, we refer to Perron and Qu [16]. In summary, Assumptions $(\mathcal{A}_1)$ and $(\mathcal{A}_2)$ are usually imposed in multiple linear regressions with structural changes. Further, Assumption $(\mathcal{A}_3)$ guarantees to have asymptotically distinct change points and Assumption $(\mathcal{A}_4)$ puts a lower bound on the distance between breaks. As mentioned in Perron and Qu [16], this assumption is stronger than the similar condition literature. As justified in the quoted paper, this is the cost needed to allow the heterogeneity and serial correlation in the errors. Assumptions $(\mathcal{A}_5)$–$(\mathcal{A}_6)$ are needed to establish the asymptotic normality of the UE. Note that Assumption $(\mathcal{A}_5)$ considers the case of mixingale random variables, which allow both the regressors and the errors in each break to be a form of different distributions and asymptotically weak dependencies.

3. The joint asymptotic distribution of the UE and RE

In this section, we derive the asymptotic joint normality for the restricted and unrestricted OLS. Under Assumptions $(\mathcal{A}_1)$–$(\mathcal{A}_4)$, $T^{-1}\bar{Z}^{0\prime}\bar{Z}^0$ converges in probability to a nonrandom $q(m+1)\times q(m+1)$-positive and definite matrix. Hereafter, we denote this matrix by Γ. Also, under Assumption $(\mathcal{A}_6)$, $T^{-1}(\bar{Z}^{0\prime}uu'\bar{Z}^0)$ converges in probability to Ω, which is a nonrandom $q(m+1)\times q(m+1)$-positive and definite matrix. Further, under Assumptions $(\mathcal{A}_5)$–$(\mathcal{A}_6)$, we establish the following lemma which is crucial in establishing the joint asymptotic of the UE and RE.

Lemma 3.1. *Under Assumptions* $(\mathcal{A}_1)$–$(\mathcal{A}_6)$, $T^{-1/2}\bar{Z}^{0\prime}u\xrightarrow[T\to\infty]{d}\mathcal{N}_{(m+1)q}(0,\Omega)$.

The proof is given in the Appendix B. Also, note that if the restriction in (2.4) does not hold, the asymptotic distribution of $\tilde{\delta}$ may degenerate. Thus, in order to derive the joint asymptotic normality, we consider the following sequence of local alternative,

$$H_{1T}: R\delta=r+\frac{\mu}{\sqrt{T}},\qquad T=1,2,\ldots, \tag{3.1}$$

with $\|\mu\| < \infty$. To simplify the notation, let $\hat{\delta}$ and $\tilde{\delta}$ denote, respectively, the UE and RE of δ. Let $J_0 = \Gamma^{-1}R'(R\Gamma^{-1}R')^{-1}$, and let I_m denote $m \times m$ identity matrix. Further, let

$$\mu_1 = -J_0\mu, \qquad \Sigma_{11} = \Gamma^{-1}\Omega\Gamma^{-1}, \qquad \Sigma_{12} = \Gamma^{-1}\Omega\Gamma^{-1}\big(I_{(m+1)q} - R'J_0'\big),$$
$$\Sigma_{21} = \Sigma_{12}', \qquad \Sigma_{22} = (I_{(m+1)q} - J_0R)\Gamma^{-1}\Omega\Gamma^{-1}\big(I_{(m+1)q} - R'J_0'\big),$$
$$\Lambda_{11} = J_0R\Sigma_{11}R'J_0', \qquad \Lambda_{12} = J_0R\Sigma_{12}, \qquad \Lambda_{21} = \Lambda_{12}', \qquad \Lambda_{22} = \Sigma_{22}.$$

Lemma 3.2. *Under Assumptions* $(\mathcal{A}_1)$–$(\mathcal{A}_6)$, *and the sequence of local alternative in* (3.1),

$$\begin{pmatrix} \sqrt{T}\big(\hat{\delta} - \delta^0\big) \\ \sqrt{T}\big(\tilde{\delta} - \delta^0\big) \end{pmatrix} \xrightarrow[T\to\infty]{d} \begin{pmatrix} \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \sim \mathcal{N}_{2(m+1)q}\left(\begin{pmatrix} 0 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right);$$
$$\begin{pmatrix} \sqrt{T}(\hat{\delta} - \tilde{\delta}) \\ \sqrt{T}\big(\tilde{\delta} - \delta^0\big) \end{pmatrix} \xrightarrow[T\to\infty]{d} \begin{pmatrix} \varepsilon_5 \\ \varepsilon_4 \end{pmatrix} \sim \mathcal{N}_{2(m+1)q}\left(\begin{pmatrix} -\mu_1 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}\right).$$

From the above result, it should be noted that $(\varepsilon_5, \varepsilon_4)'$, the limit in distribution $(\sqrt{T}(\hat{\delta} - \tilde{\delta}), \sqrt{T}(\tilde{\delta} - \delta^0))$ are not uncorrelated as for example in Saleh [18], Theorem 3, page 375, Hossain *et al.* [7], among others. Further, note that Λ_{11} and Λ_{22} are not positive definite matrices as the case in Judge and Mittelhammer [10]. Because of that, the construction of shrinkage-type estimators cannot be obtained by applying the results given in the literature.

4. Shrinkage estimator and related asymptotic properties

It is well known that under the restriction in (2.4), the RE dominates in mean square error sense the UE. However, if the restriction in (2.4) is seriously violated, the RE performs poorly. In some scenarios, the prior restriction in (2.4) is subjected to some uncertainty that may be induced by the change in the phenomenon underlying the regression model in (2.3). Under such an uncertainty, it is of interest to propose a statistical method which combine in optimal way the sample information and an uncertain information given in (2.4).

In this section, we introduce a class of shrinkage estimators which encloses the UE, RE as well as Stein-type estimator, and positive part Stein-type estimator. To simplify some notations, let $A = R'(R\Gamma^{-1}\Omega\Gamma^{-1}R')^{-1}R$, and $\hat{A} = R'(R\hat{\Gamma}^{-1}\hat{\Omega}\hat{\Gamma}^{-1}R')^{-1}R$, where $\hat{\Omega}$ and $\hat{\Gamma}$ denote consistent estimators of Ω and Γ, respectively. Also, as in Nkurunziza [14], let h be continuous (except on a number of finite points), real-valued and integrable function (with respect to the Gaussian measure). We consider the following class of estimators

$$\hat{\beta}(h) = \tilde{\delta} + h\big(T(\tilde{\delta} - \hat{\delta})'\hat{A}(\tilde{\delta} - \hat{\delta})\big)(\hat{\delta} - \tilde{\delta}). \tag{4.1}$$

It should be noted that for the case where $h \equiv 0$, $\hat{\beta}(0)$ is the RE $\tilde{\delta}$. Also, if $h \equiv 1$, we have the UE, that is, $\hat{\beta}(1) = \hat{\delta}$. Further, by choosing a suitable h one can get the pretest estimators as given for example in Saleh [18], Hossain *et al.* [7], among others. Finally, the James–Stein estimator $\hat{\delta}^s$ and Positive-Rule Stein estimator $\hat{\delta}^{s+}$ are members of the class in (4.1). Indeed, let

k denote the rank of the matrix $\mathbf{R}$ as defined in (2.4). By taking $h(x) = 1 - (k-2)/x$, $x > 0$, and $h(x) = \max\{0, 1 - (k-2)/x\}$, $x > 0$ we get $\hat{\delta}^s$ and $\hat{\delta}^{s+}$, respectively. More precisely, we have $\hat{\delta}^s = \tilde{\delta} + (1 - \frac{k-2}{\psi})(\hat{\delta} - \tilde{\delta})$, $\hat{\delta}^{s+} = \tilde{\delta} + (1 - \frac{k-2}{\psi})^+(\hat{\delta} - \tilde{\delta})$ where $\psi = T(\tilde{\delta} - \hat{\delta})'\hat{A}(\tilde{\delta} - \hat{\delta})$, with $x^+ = \max(0, x)$.

In order to evaluate the performance of the proposed estimators, we consider the quadratic loss function $L(\theta, d) = (d - \theta)'W(d - \theta)$, where W is a symmetric nonnegative definite matrix, and use the asymptotic distributional risk (ADR) as defined, for example, in Saleh [18]. For the convenience of the reader, we recall that the ADR of an estimator $\hat{\boldsymbol{\theta}}$ is defined as $\mathrm{ADR}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \mathbf{W}) = \mathrm{E}[\boldsymbol{\rho}_0'\mathbf{W}\boldsymbol{\rho}_0]$, with $\boldsymbol{\rho}_0$ the limit in distribution of $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ as T tends to infinity, and $\mathbf{W}$ is a certain weight nonnegative definite matrix.

In the sequel, we set $\Delta = \mu_1' A \mu_1$ and assume that the weight matrix W satisfies $W = A^{1/2} W^* A^{1/2}$, with W^* a symmetric nonnegative definite matrix. We establish below a lemma which gives the ADR of estimators which are members of the class in (4.1). Briefly, the derivation of this lemma is based on the identity, established in Appendix C, which generalizes Theorem 2 in Judge and Bock [9]. In particular, this lemma is useful in deriving ADR of $\hat{\delta}$, $\tilde{\delta}$, $\hat{\delta}^s$ and $\hat{\delta}^{s+}$.

Lemma 4.1. *Suppose that Assumptions* $(\mathcal{A}_1)$–$(\mathcal{A}_6)$ *and the sequence of local alternative in* (3.1) *hold. Then*

$$\begin{aligned}
&\mathrm{ADR}\big(\hat{\beta}(h), \delta^0, W\big) \\
&\quad = \mathrm{ADR}\big(\tilde{\delta}, \delta^0, W\big) - 2\mathrm{E}\big[h\big(\chi^2_{k+2}(\Delta)\big)\big]\mu_1' W \mu_1 \\
&\qquad - 2\mathrm{E}\big[h\big(\chi^2_{k+2}(\Delta)\big)\big]\mu_1' A \Lambda_{12} W \mu_1 + 2\mathrm{E}\big[h\big(\chi^2_{k+2}(\Delta)\big)\big]\operatorname{trace}(\Lambda_{12} W \Lambda_{11} A) \\
&\qquad + 2\mathrm{E}\big[h\big(\chi^2_{k+4}(\Delta)\big)\big]\mu_1' A \Lambda_{12} W \mu_1 \\
&\qquad + \mathrm{E}\big[h^2\big(\chi^2_{k+2}(\Delta)\big)\big]\operatorname{trace}(W\Lambda_{11}) + \mathrm{E}\big[h^2\big(\chi^2_{k+4}(\Delta)\big)\big]\mu_1' W \mu_1.
\end{aligned} \tag{4.2}$$

Proof. The proof of this lemma follows directly by combining Lemma 3.2, Theorem C.2 and Lemma C.3. □

From Lemma 4.1, by taking $h(x) = 1$, $h(x) = 0$, $h(x) = 1 - \frac{k-2}{x}$ and $h(x) = \max\{0, (1 - \frac{k-2}{x})\}$, we establish the following corollary which gives the ADR of the estimators $\hat{\delta}$, $\tilde{\delta}$, $\hat{\delta}^s$ and $\hat{\delta}^{s+}$, respectively.

Corollary 4.1. *Suppose that the conditions of Lemma* 4.1 *hold*, *then*

$$\begin{aligned}
&\mathrm{ADR}\big(\hat{\delta}, \delta^0, W\big) \\
&\quad = \operatorname{trace}\big(W\Gamma^{-1}\Omega\Gamma^{-1}\big), \\
&\mathrm{ADR}\big(\tilde{\delta}, \delta^0, W\big) \\
&\quad = \operatorname{trace}\big[W(I_{q(m+1)} - J_0 R)\Gamma^{-1}\Omega\Gamma^{-1}\big(I_{q(m+1)} - R' J_0'\big)\big] + \mu_1' W \mu_1,
\end{aligned}$$

$$
\begin{aligned}
&\operatorname{ADR}(\hat{\delta}^s, \delta^0, W) \\
&\quad = \operatorname{ADR}(\hat{\delta}, \delta^0, W) - 2(k-2)\mathrm{E}\big[\chi_{k+2}^{-2}(\Delta)\big]\operatorname{trace}\big(W(\Lambda_{11}+\Lambda_{12})\big) \\
&\qquad + (k^2-4)\mathrm{E}\big[\chi_{k+4}^{-4}(\Delta)\big]\mu_1' W \mu_1 + (k-2)^2\mathrm{E}\big[\chi_{k+2}^{-4}(\Delta)\big]\operatorname{trace}(W\Lambda_{11}) \\
&\qquad + 4(k-2)\mathrm{E}\big[\chi_{k+4}^{-4}(\Delta)\big]\mu_1' A\Lambda_{12} W \mu_1, \\
&\operatorname{ADR}(\hat{\delta}^{s+}, \delta^0, W) \\
&\quad = \operatorname{ADR}(\hat{\delta}^s, \delta^0, W) \\
&\qquad + 2\mathrm{E}\big(I\big(\chi_{k+2}^2(\Delta) < k-2\big) - (k-2)\chi_{k+2}^{-2}(\Delta) I\big(\chi_{k+2}^2(\Delta) < k-2\big)\big)\mu_1' W \mu_1 \\
&\qquad + 2\mathrm{E}\big(I\big(\chi_{k+2}^2(\Delta) < k-2\big) - (k-2)\chi_{k+2}^{-2}(\Delta) I\big(\chi_{k+2}^2(\Delta) < k-2\big)\big)\mu_1' A\Lambda_{12} W \mu_1 \\
&\qquad - 2\mathrm{E}\big(I\big(\chi_{k+2}^2(\Delta) < k-2\big) - (k-2)\chi_{k+2}^{-2}(\Delta) I\big(\chi_{k+2}^2(\Delta) < k-2\big)\big)\operatorname{trace}(W\Lambda_{12}) \\
&\qquad - 2\mathrm{E}\big(I\big(\chi_{k+4}^2(\Delta) < k-2\big) - (k-2)\chi_{k+4}^{-2}(\Delta) I\big(\chi_{k+4}^2(\Delta) < k-2\big)\big)\mu_1' A\Lambda_{12} W \mu_1 \\
&\qquad - \mathrm{E}\big(I\big(\chi_{k+2}^2(\Delta) < k-2\big) - 2(k-2)\chi_{k+2}^{-2}(\Delta) I\big(\chi_{k+2}^2(\Delta) < k-2\big) \\
&\qquad\qquad + (k-2)^2\chi_{k+2}^{-4}(\Delta) I\big(\chi_{k+2}^2(\Delta) < k-2\big)\big)\operatorname{trace}(W\Lambda_{11}) \\
&\qquad - \mathrm{E}\big(I\big(\chi_{k+4}^2(\Delta) < k-2\big) - 2(k-2)\chi_{k+4}^{-2}(\Delta) I\big(\chi_{k+4}^2(\Delta) < k-2\big) \\
&\qquad\qquad + (k-2)^2\chi_{k+4}^{-4}(\Delta) I\big(\chi_{k+4}^2(\Delta) < k-2\big)\big)\mu_1' W \mu_1.
\end{aligned}
\tag{4.3}
$$

It should be noted that the expressions in Corollary 4.1 are more general than that, for example, in Saleh [18], page 377, and Hossain *et al.* [7] for which $\Lambda_{12} = \mathbf{0}$.

From Corollary 4.1, we establish the following corollary which shows that shrinkage estimators dominate the UE. It is noticed that, due to the asymptotic dependance structure between the shrinking factor and the restricted estimator, the above dominance condition looks quite complicated. To simplify the notation, let $\mathrm{Ch}_{\max}(\Pi)$ denote the largest eigenvalue of Π, and let $\mathrm{Ch}_{\min}(\Pi)$ denote the smallest eigenvalue of Π. Further, let $\Pi_0 = A^{1/2}(\Lambda_{11} + 4\Lambda_{12}/(k+2))W\Lambda_{11}A^{1/2}$, $\Pi^* = (\Pi_0 + \Pi_0')/2$.

Corollary 4.2. *Suppose that Assumptions* $(\mathcal{A}_1)$–$(\mathcal{A}_6)$ *hold, and let* W *be nonnegative definite matrix such that* $\operatorname{trace}(W\Lambda_{12}) \le 0$, $-\mathrm{Ch}_{\min}(W\Lambda_{11}) \le \mathrm{Ch}_{\min}(W\Lambda_{12})$ *and* $\operatorname{trace}(W(\Lambda_{11} + \Lambda_{12})) \ge \max(-\operatorname{trace}(W\Lambda_{12}), (k+2)\,\mathrm{Ch}_{\max}(\Pi^*)/4)$. *Then*,

$$
\operatorname{ADR}(\hat{\delta}^{s+}, \delta^0, W) \le \operatorname{ADR}(\hat{\delta}^s, \delta^0, W) \le \operatorname{ADR}(\hat{\delta}, \delta^0, W), \qquad \text{for all } \Delta \ge 0. \tag{4.4}
$$

Remark 4.1. It should be noted that the conditions for the shrinkage estimators to dominate the unrestricted estimator are more general than given for example in Hossain *et al.* [7], Corollary 4.2, Saleh [18], pages 358, 360, 382, the relations (7.4.8), (7.4.31) and (7.8.35).

Indeed, in the quoted work, we have $\Lambda_{12} = \mathbf{0}$. In this special case, the above condition can be rewritten as $\{W : \frac{\text{trace}(W\Lambda_{11})}{\text{Ch}_{\max}(W\Lambda_{11})} \geq \frac{k+2}{4}\}$ and this set contains $\{W : \frac{\text{trace}(W\Lambda_{11})}{\text{Ch}_{\max}(W\Lambda_{11})} \geq \frac{k+2}{2}\}$ which given in the above quoted works.

5. Illustrative data set and numerical evaluation

5.1. Simulation study

In this section, we present some Monte Carlo simulation results to evaluate the performances of the proposed estimators. This is done by comparing the relative mean square efficiencies (RMSE) of the estimators with respect to the UE, $\hat{\delta}$. Recall that $\text{RMSE}(\delta^*) = \text{risk}(\hat{\delta})/\,\text{risk}(\delta^*)$, where δ^* is the proposed estimator. Note that, a relative efficiency greater than one indicates the degree of superiority of the proposed estimator over $\hat{\delta}$. To save the space of this paper, we report only two cases.

Case 1: the number of unknown parameters is small, with $m = 3, q = 2$; $\delta^0 = (\delta_1^{0'}, \delta_2^{0'}, \delta_3^{0'}, \delta_4^{0'})'$ with $\delta_1^0 = \delta_3^0 = (1, 2)'$ and $\delta_2^0 = \delta_4^0 = \mathbf{0}$ (i.e., the zero vector), and the sample sizes are set to be $T = 40$ with the change points given by $(10, 20, 30, 40)$. Also, we set $T = 100$ with the change-points $(25, 50, 75, 100)$. Further, the restriction is such that $\mathbf{R} = [E_1, E_2, E_3, E_4, -E_1, -E_2, E_5, E_6]$ where, for $j = 1, 2, \ldots, 6$, E_j is a 6-column vector with all components equal to zero except the jth component which equal to 1.

Case 2: the number of unknown parameters is relative large by setting $m = 4$, $q = 5$, $\delta^0 = (\delta_1^{0'}, \delta_2^{0'}, \delta_3^{0'}, \delta_4^{0'}, \delta_5^{0'})'$ with $\delta_1^0 = \delta_3^0 = \delta_5^0 = (1, 2, 3, 4, 5)'$, $\delta_2^0 = \delta_4^0 = \mathbf{0}$ and the sample sizes are $T = 100$ and $T = 500$ with the change-points $(20, 40, 60, 80, 100)$ and $(100, 200, 300, 400, 500)$, respectively. Further, the restriction R is set to be a 8×25 matrix with

$$R_{1,1} = R_{2,2} = R_{3,3} = R_{4,4} = R_{5,5} = R_{6,6} = R_{7,19} = R_{8,20} = 1,$$
$$R_{1,11} = R_{2,12} = R_{3,13} = R_{4,14} = R_{5,15} = -1,$$

and the rest elements of R are set to be 0.

In each case, we let $z_{T_i} \sim \mathcal{N}_q(1, \Sigma)$, where Σ is a $q \times q$ symmetric matrix such that $\Sigma_{a,b} = |0.5|^{|a-b|}$. Also, we let $u_i \sim \mathcal{N}(0, \sigma^2)$, $1 \leq \sigma^2 \leq 2$, and compute the related RMSE based on the 1000 replications.

The results of the simulation studies are given in Figures 1 and 2. In summary, the results corroborate the theoretical finding (given in Corollary 4.2) for which the proposed shrinkage estimators dominate the unrestricted estimator. We also construct, and present in Appendix C, Figures 3–6 which give some histograms of the UE and RE of the change points. The results given in Figures 3–6 suggest that both the unrestricted and the restricted methods work well in estimating the change points.

5.2. Data analysis

In this subsection, we illustrate the application of the proposed estimation strategy to the real data set. As a real data set, we consider a historical (log) gross domestic product (GDP) data set

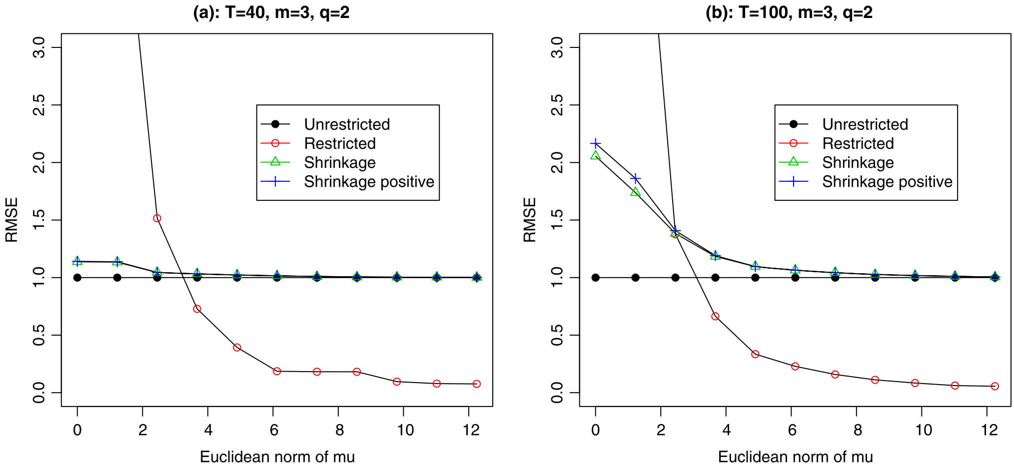

Figure 1. RMSE of the restricted and shrinkage estimators (case 1).

from 1870 to 1986 for 10 different countries. This data set is used for example in Perron and Yabu [17], and these authors pointed out that most GDP series presented in the given data set are characterized by at least one major shift and therefore change-point model is applicable. For

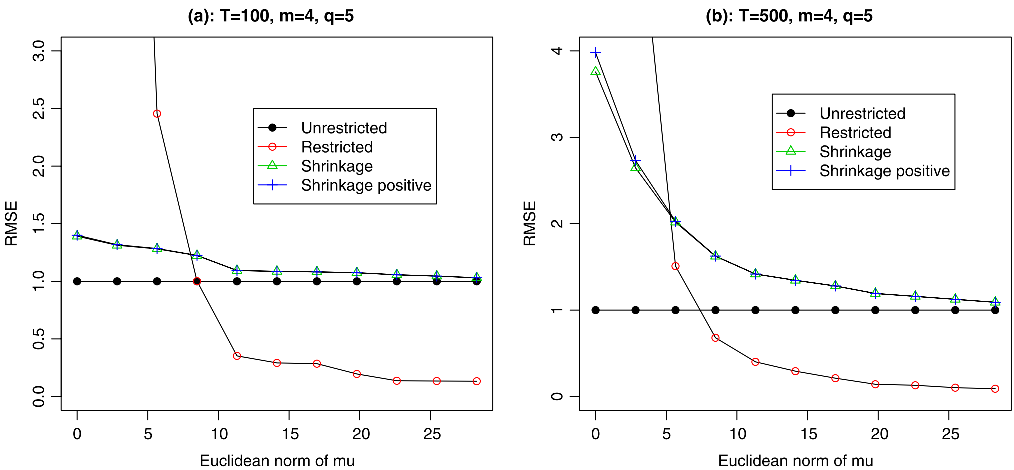

Figure 2. RMSE of the restricted and shrinkage estimators (case 2).

each GDP series, we consider the following model:

$$Y_t = \begin{cases} \delta_1'\left(1, t, t^{1.5}, t^2\right)', & \text{if } t = 1, \ldots, T_1, \\ \delta_2'\left(1, t, t^{1.5}, t^2\right)', & \text{if } t = T_1 + 1, \ldots, 117, \end{cases}$$

with $1 \leq T_1 \leq 117$, for $i = 1, 2$, δ_i is a 4-column vector. The uncertain restriction is given by $R\delta = r$ with

$$R = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and $r = \{0, 0, 0, 0\}'$. In practice, the hypothesized restriction means that the log(GDP) is suspected to have a linear trend. For the given data, we first use the proposed method to calculate the unrestricted and the restricted estimators of the change-point $\hat{T}_1$ and $\tilde{T}_1$ as well as the estimators $\hat{\delta}$, $\tilde{\delta}$, $\hat{\delta}^s$ and $\hat{\delta}^{s+}$. For the change-point T_1 which is a nuisance parameter here, we do not compute the shrinkage estimators. The obtained unrestricted and restricted estimate of the change-point $\hat{T}_1$ and $\tilde{T}_1$ are given in Table 1. In order to save the space of this paper, we do not report here the point estimates of $\hat{\delta}, \tilde{\delta}, \hat{\delta}^s, \hat{\delta}^{s+}$, but these values are available upon request. Further, we calculate the MSE of each type of estimators, by applying the bootstrap method to the residuals. Recall that, in this paper, the change-points are treated as the nuisance parameters. Thus, the construction of the shrinkage estimators for the change-points is beyond the scope of this paper.

As we can see from Table 1, the MSE of the restricted estimator is much smaller the MSE of the other estimators. This may indicate that the true value of the parameter vector lies in the neighborhood of the chosen restriction. Further, the MSE of the proposed shrinkage estimators is smaller than the MSE of the unrestricted estimator. The obtained result is in agreement with the above simulation study.

Table 1. Change-points and MSE

	Change-points		MSE			
Country	(UE)	(RE)	$\hat{\delta}$	$\tilde{\delta}$	$\hat{\delta}^s$	$\hat{\delta}^{s+}$
Australia	1907	1929	1.67004021	0.03936242	1.64839567	1.64839567
Canada	1931	1930	2.96623326	0.05474518	2.87279365	2.87279365
Denmark	1939	1939	3.99038175	0.04765026	3.93532691	3.93532691
France	1943	1943	12.1123258	0.1253509	11.9030741	11.9030741
Germany	1945	1954	11.4218637	0.1704905	11.3279191	11.3279191
Italy	1943	1943	10.2462836	0.1211837	10.2079175	10.2079175
Norway	1944	1948	7.09593981	0.03606614	6.92396377	6.92396377
Sweden	1924	1916	0.72605495	0.02192452	0.70854206	0.70854206
U.K.	1918	1919	0.61037392	0.01496282	0.58916536	0.57701458
U.S.	1940	1929	3.97869572	0.05967521	3.91443168	3.91443168

6. Conclusion

The goal of this research was to derive an improved estimation strategy for the regression coefficients in multiple linear model with unknown change-points under uncertain restrictions. In summary, we introduced a class of estimators which includes the UE $\hat{\delta}$, RE $\tilde{\delta}$, James–Stein Estimator $\hat{\delta}^s$ and Positive-Rule Stein Estimator $\hat{\delta}^{s+}$. The main difficulty consists in the fact that the random quantities $\tilde{\delta}-\delta$ and $\hat{\delta}-\tilde{\delta}$ are not asymptotically uncorrelated as this is the common case in literature. To tackle this difficulty, we generalized (in the Appendix C) Theorems 1–2 in Judge and Bock [9]. Under the conditions more general than that in literature, we established that $\hat{\delta}^s$ and $\hat{\delta}^{s+}$ dominate UE. The performance of SEs over the UE is confirmed by the simulation studies. They also show that SEs perform better than the RE when one moves far away from the hypothesized restriction. It should be noticed that, in this paper, the tools used for studying shrinkage estimators are based on noncentral chi-squares. One of the referees suggested to investigate if the obtained results can be improved by using more recent tools such as Stein's unbiased risk estimate. Research on this interesting idea is ongoing.

Another highlight of this paper consists in the fact that, in deriving the joint asymptotic normality of the UE and RE, we relax some conditions given in recent literature. In particular, we considered here the condition of L_2-mixingale with size $-1/2$, which allow both the regressors and the errors in each break to be a form of different distributions and asymptotically weak dependencies.

Appendix

In this section, we give some technical proofs underlying the results established in this paper. To set up additional notations, let $\|A\|$ denote the Euclidean norm for vector A. For a matrix B, let $\|B\|$ be the vector induced norm (i.e., $\|B\|=\sup_{x\neq 0}\|Bx\|/\|x\|$).

Appendix A: Technical results underlying the asymptotic properties

First, we establish the following proposition which plays a central role in deriving the joint asymptotic normality between the UE and RE. For the sake of simplicity, we set $D_{i,k^*}=X_{pi}-\mathrm{E}(X_{pi}|\mathcal{F}_{p,i+k^*})$ and set $D_{i,k^*,s}$ be the sth element in D_{i,k^*}.

Proposition A.1. *Suppose that Assumptions* ($\mathcal{A}_5$) *and* ($\mathcal{A}_6$) *hold. Then*,

$$\mathrm{E}\left(\sum_{r=1}^{L_p}(D_{i,k^*-1,s}-D_{i,k^*,s})^2\right)=\sum_{i=1}^{L_p}\mathrm{E}\big(\mathrm{E}^2(X_{pi,s}|\mathcal{F}_{p,i+k^*})\big)-\sum_{i=1}^{L_p}\mathrm{E}\big(\mathrm{E}^2(X_{pi,s}|\mathcal{F}_{p,i+k^*-1})\big),$$

$$\sum_{i=1}^{L_p}\sum_{j=l}^{i-1}\mathrm{E}\big[(D_{i,k^*-1,s}-D_{i,k^*,s})(D_{j,k^*-1,s}-D_{j,k^*,s})\big]=0$$

and

$$\sum_{i=1}^{L_p}\left[\mathrm{E}\left(\mathrm{E}^2(X_{pi,s}|\mathcal{F}_{p,i+k^*})\right)-\mathrm{E}\left(\mathrm{E}^2(X_{pi,s}|\mathcal{F}_{p,i+k^*-1})\right)\right]=\sum_{i=1}^{L_p}\left[\mathrm{E}\left(D^2_{i,k^*-1,s}\right)-\mathrm{E}\left(D^2_{i,k^*,s}\right)\right].$$

Proof. One can verify that

$$X_{pi}=\sum_{k^*=-\infty}^{\infty}\left[\mathrm{E}(X_{pi}|\mathcal{F}_{p,i+k^*})-\mathrm{E}(X_{pi}|\mathcal{F}_{p,i+k^*-1})\right]\qquad \text{a.s.}$$

Further, one can verify that

$$\begin{aligned}
&\mathrm{E}\left(\sum_{r=1}^{L_p}(D_{i,k^*-1,s}-D_{i,k^*,s})^2\right)\\
&\quad=\sum_{i=1}^{L_p}\mathrm{E}\left[\mathrm{E}^2(X_{pi,s}|\mathcal{F}_{p,i+k^*})\right]+\sum_{i=1}^{L_p}\mathrm{E}\left[\mathrm{E}^2(X_{pi,s}|\mathcal{F}_{p,i+k^*-1})\right]\\
&\qquad-2\sum_{i=1}^{L_p}\mathrm{E}\left[\mathrm{E}(X_{pi,s}|\mathcal{F}_{p,i+k^*-1})\mathrm{E}\left(\mathrm{E}(X_{pi,s}|\mathcal{F}_{p,i+k^*})|\mathcal{F}_{p,i+k^*-1}\right)\right],
\end{aligned}$$

and then, by using the properties of the conditional expected value, we prove the first statement. For the second statement, we have

$$\begin{aligned}
&\sum_{i=1}^{L_p}\sum_{j=1}^{i-1}\mathrm{E}\left[\left((D_{i,k^*-1,s}-D_{i,k^*,s})(D_{j,k^*-1,s}-D_{j,k^*,s})\right)\right]\\
&\quad=\sum_{i=1}^{L_p}\sum_{j=1}^{i-1}\mathrm{E}\left[(D_{j,k^*-1,s}-D_{j,k^*,s})\left(\mathrm{E}\left((D_{i,k^*-1,s}-D_{i,k^*,s})|\mathcal{F}_{p,j+k^*}\right)\right)\right]=0.
\end{aligned}$$

The third statement of the proposition follows from the similar algebraic computations. □

Lemma A.1. *Let $v^2_{L_p}=\sum_{i=1}^{L_p}c^2_{pi}$ and suppose that Assumptions $(\mathcal{A}_5)$ and $(\mathcal{A}_6)$ hold. Then*

$$\sum_{s=1}^{q}\mathrm{E}\left(\max_{j\le L_p}\left(\sum_{i=1}^{j}X_{pi,s}\right)^2\right)\le 16v^2_{L_p}\left[\sum_{k^*=0}^{\infty}\left(\sum_{i=0}^{k^*}\psi^{-2}(i)\right)^{-1/2}\right]^2.$$

The proof follows from Proposition A.1 and following the similar steps as in proof of Lemma 3.2 in Mcleish [12]. By using this lemma, one establishes the following corollary which plays a central role in establishing the joint asymptotic normality of UE and RE.

Corollary A.1. *Under Assumptions* $(\mathcal{A}_5)$ *and* $(\mathcal{A}_6)$, *then*

$$\sum_{s=1}^{q} \mathrm{E}\left[\left(\sum_{i=1}^{L_p} X_{pi,s}\right)^2\right] = \mathrm{O}\big(v_{L_p}^2\big).$$

Proof. From Lemma A.1,

$$\sum_{s=1}^{q} \mathrm{E}\left(\max_{j\le L_p}\left(\sum_{i=1}^{j} X_{pi,s}\right)^2\right) \le 16 v_{L_p}^2 \left[\sum_{k^*=0}^{\infty}\left(\sum_{i=0}^{k^*}\psi^{-2}(i)\right)^{-1/2}\right]^2, \tag{A.1}$$

and then, the proof follows directly from the fact that $\sum_{k^*=0}^{\infty}(\sum_{i=0}^{k^*}\psi^{-2}(i))^{-1/2} < \infty$. □

Corollary A.2. *Let* $v_i^2 = \sum_{(i-1)b_p+l_p+1}^{ib_p} c_{pt}^2$ *and suppose that Assumptions* $(\mathcal{A}_5)$ *and* $(\mathcal{A}_6)$ *hold. Then,* $\{\sum_{s=1}^{q}\max_{j\le ib_p}(\sum_{t=(i-1)b_p+l_p+1}^{j} X_{pt,s})^2/v_i^2, i=1,\dots,r_p, r_p\ge 1\}$ *is uniformly integrable. In particular,* $\{\sum_{s=1}^{q}(\sum_{t=(i-1)b_p+l_p+1}^{ib_p} X_{pt,s})^2/(v_i^2), i=1,\dots,r_p, r_p\ge 1\}$ *is uniformly integrable.*

Proof. Let $S_{j,s} = \sum_{i=1}^{j} X_{pi,s}$, $s=1,\dots,q$. By using the same arguments as used in proof of Lemma 3.5 in McLeish [12], one verifies that the set $\{\max_{j\le L_p}\sum_{s=1}^{q}\frac{S_{j,s}^2}{v_{L_p}^2}; L_p\ge 1\}$ is uniformly integrable. This completes the proof. □

Further, by using Lemma A.1, we establish the following proposition which is also useful in establishing the joint asymptotic normality of UE and RE. To simplify some notations, let $r_{\min} = \min_{1\le p\le m+1}(r_p)$, and let $L_{\min} = \min_{1\le p\le m+1}(L_p)$. Further, let $\mathcal{H}_i$ be the σ-field generated by $\{U_{ib_p}, U_{ib_p-1},\dots\}$, with U_i are random variables defined on $(\Omega,\mathcal{F},P)$ such that $\mathcal{H}_{i-1}\subseteq\mathcal{F}_{p,i-j}$, and let $V_{pi} = \sum_{t=(i-1)b_p+l_p+1}^{ib_p} X_{pt}$, let $W_{p,i} = \mathrm{E}(V_{pi}|\mathcal{H}_i) - \mathrm{E}(V_{pi}|\mathcal{H}_{i-1})$, $p=1,2,\dots,m+1$, $i=1,2,\dots,r_{\min}$.

Proposition A.2. *Suppose that Assumptions* $(\mathcal{A}_5)$ *and* $(\mathcal{A}_6)$ *hold. Then,*

$$\sum_{i=1}^{r_{\min}}\big[\big(V_{1,i}',\dots,V_{m+1,i}'\big)'\big(V_{1,i}',\dots,V_{m+1,i}'\big) - \big(W_{1,i}',\dots,W_{m+1,i}'\big)'\big(W_{1,i}',\dots,W_{m+1,i}'\big)\big] \xrightarrow[L_{\min}\to\infty]{p} 0.$$

The proof follows from Lemma A.1 along with some algebraic computations.

Proposition A.3. *Suppose that the conditions Proposition A.2 hold. Then,*

$$\sum_{i=1}^{r_{\min}} \left(W'_{1,i}, W'_{2,i}, \ldots, W'_{m+1,i}\right)' \left(W'_{1,i}, W'_{2,i}, \ldots, W'_{m+1,i}\right) \xrightarrow[L_{\min}\to\infty]{p} \Omega$$

and

$$\sum_{a=1}^{m+1}\sum_{i=1}^{r_a}\sum_{s=1}^{q} \mathrm{E}\left[(W_{a,i,s})^2 \mathbb{I}\left(\sum_{s=1}^{q} W_{a,i,s}^2 > \varepsilon\right)\right] \xrightarrow[L_{\min}\to\infty]{} 0, \qquad \textit{for all } \varepsilon > 0.$$

Proof. By using Assumption $(\mathcal{A}_6)$ along with Proposition A.2 and Slutsky's theorem, we establish the first statement. For the second statement, one verifies that, for each $a = 1, 2, \ldots, m+1$, $\{W_{a,i}, \mathcal{H}_i\}$ is a L_2-mixingale array of size $-1/2$. Then the rest of the proof follows from Corollary A.2. □

Appendix B: Asymptotic normality of the UE and RE

Proof of Lemma 3.1. Note that

$$T^{-1/2}\bar{Z}^{0\prime}u \equiv \left(\sum_{i=1}^{L_1} X'_{1,i}, \ldots, \sum_{i=1}^{L_{m+1}} X'_{m+1,i}\right)',$$

then

$$T^{-1/2}\bar{\mathbf{Z}}^{0\prime}u = \sum_{i=1}^{r_{\min}} \mathbf{W}_i + \boldsymbol{\Xi}^* + \left(\sum_{i=r_{\min}}^{r_1}\sum_{t=(i-1)b_1+1}^{ib_1} X'_{1,i}, \ldots, \sum_{i=r_{\min}}^{r_{m+1}}\sum_{t=(i-1)b_{m+1}+1}^{ib_{m+1}} X'_{m+1,i}\right)', \quad \text{(B.1)}$$

with $r_{\min} = \min_{1\le i\le m+1}(r_i)$ and $\boldsymbol{\Xi}^* = (\Xi_1^{*\prime}, \Xi_2^{*\prime}, \ldots, \Xi_{m+1}^{*\prime})'$, where

$$\Xi_j^* = \sum_{i=1}^{r_{\min}}\left(V_{ji} - W_{j,i} + \sum_{t=(i-1)b_j+1}^{i-1b_j+l_j} X_{j,i}\right) + \sum_{t=r_j b_j+1}^{L_j} X_{p_j,t}.$$

Further, it should be noted that, under Assumptions $(\mathcal{A}_4)$ and $(\mathcal{A}_5)$, T tends to infinity if and only if $L_{\min} = \min_{1\le j\le m+1}(L_j)$ tends to infinity.

By using Lemma A.1 along with some algebraic computations, we have

$$\begin{gathered}\left(\Xi_1^{*\prime}, \Xi_2^{*\prime}, \ldots, \Xi_{m+1}^{*\prime}\right)' \xrightarrow[L_{\min}\to\infty]{\mathrm{P}} 0, \\ \left(\sum_{i=r_{\min}}^{r_1}\sum_{t=(i-1)b_1+1}^{ib_1} X'_{1,i}, \ldots, \sum_{j=r_{\min}}^{r_{m+1}}\sum_{t=(j-1)b_{m+1}+1}^{jb_{m+1}} X'_{m+1,j}\right)' \xrightarrow[L_{\min}\to\infty]{\mathrm{P}} 0.\end{gathered} \quad \text{(B.2)}$$

Therefore, the proof follows from the relations (B.1) and (B.2) along with the martingale difference sequence central limit theorem along with Slutsky's theorem. □

Proposition B.1. *Under* $(\mathcal{A}_1)$–$(\mathcal{A}_6)$, *we have* $\sqrt{T}(\hat{\delta}-\delta^0) \xrightarrow[T\to\infty]{d} \varepsilon_1 \sim \mathcal{N}_{q(m+1)}(0, \Gamma^{-1}\Omega\Gamma^{-1})$.

The proof follows by combining Lemma 3.1 and Slutsky's theorem.

Proof of Proposition 3.2. Let $J = (\bar{Z}^{0\prime}\bar{Z}^0)^{-1}R'(R(\bar{Z}^{0\prime}\bar{Z}^0)^{-1}R')^{-1}$, we have

$$\big(\sqrt{T}\big(\hat{\delta}-\delta^0\big)', \sqrt{T}\big(\tilde{\delta}-\delta^0\big)'\big)' \doteq \big(I_{(m+1)q}, I_{(m+1)q} - R'J'\big)'\sqrt{T}\big(\hat{\delta}-\delta^0\big) + \big(0, -\mu'J'\big)'.$$

Then, the first statement follows directly from Proposition B.1 and Slutsky's theorem, along with some algebraic computations. For the second statement, obviously

$$\big((\hat{\delta}-\tilde{\delta})', \big(\tilde{\delta}-\delta^0\big)'\big)' = \big((I_{q(m+1)}, 0)', (-I_{q(m+1)}, I_{q(m+1)})'\big)'\big(\big(\hat{\delta}-\delta^0\big)', \big(\tilde{\delta}-\delta^0\big)'\big)'.$$

Then, the rest of the proof follows directly from the first statement of the proposition along with Slutsky's theorem. □

Appendix C: Some results for the derivation of risk functions

Theorem C.1. *Let* h *be Borel measurable and real-valued integrable function, let* $X \sim \mathcal{N}_p(\mu, \Sigma)$, *where* Σ *is a nonnegative definite matrix with rank* $k \le p$. *Let* A *be a* $p \times p$*-nonnegative definite matrix with rank* k *such that* ΣA *is an idempotent matrix,* $A\Sigma A = A$; $\Sigma A \Sigma = \Sigma$; *and* $\Sigma A \mu = \mu$, *and let* $W = A^{1/2}W^*A^{1/2}$ *where* W^* *is a nonnegative definite matrix. Then,* $\mathrm{E}[h(X'AX)WX] = \mathrm{E}[h(\chi^2_{k+2}(\mu'A\mu))]W\mu$.

Proof. Let $A^{1/2\dagger}$ be the Moore–Penrose pseudoinverse of $A^{1/2}$. By the definition of Moore–Penrose pseudo-inverse, we have $WX = A^{1/2}W^*A^{1/2}A^{1/2\dagger}A^{1/2}X = WA^{1/2\dagger}A^{1/2}X$, and then,

$$\mathrm{E}\big[h\big(X'AX\big)X'WX\big] = \mathrm{E}\big[h\big(X'AX\big)X'A^{1/2}A^{1/2\dagger}WA^{1/2\dagger}A^{1/2}X\big]. \tag{C.1}$$

Further, since $A^{1/2}\Sigma A^{1/2}$ is a symmetric and idempotent matrix, there exists an orthogonal matrix G such that $GA^{1/2}\Sigma A^{1/2}G' = ([I_k, 0]\vdots[0, 0])'$. Define $V = GA^{1/2}X$. Then, $\mathrm{E}[h(X'AX)WA^{1/2\dagger}A^{1/2}X] = \mathrm{E}[h(V_1'V_1)WA^{1/2\dagger}G'[I_k, 0]'V_1]$ with $V_1 = [I_k, 0]GA^{1/2}V$, and then, the rest of the proof follows from Theorem 1 in Judge and Bock [9] along with some algebraic computations. □

Remark C.1. For the special case where Σ is the p-dimensional matrix I_p, Theorem C.1 gives Theorem 1 in Judge and Bock [9] with $A = W^* = I_p$. This shows that the provided theorem generalizes the quoted classical result.

By using Theorem C.1, we establish the following corollary.

Corollary C.1. *Set $\mu_2 = -\mu_1$ and let ε_5 be as defined in Lemma 3.2. Let h be a Borel measurable and real-valued integrable function, let $W = A^{1/2}W^*A^{1/2}$, W^* is a nonnegative definite matrix. Then, we have* $\mathrm{E}[h(\varepsilon_5' A\varepsilon_5)W\varepsilon_5] = \mathrm{E}[h(\chi^2_{k+2}(\mu_2' A\mu_2))]W\mu_2$.

Theorem C.2. *Let* $D_1 = \mathrm{trace}(W\Sigma)$, $D_2 = \mu' W\mu$ *and assume the conditions of Theorem C.1 hold. Then,* $\mathrm{E}[h(X'AX)X'WX] = \mathrm{E}[h(\chi^2_{k+2}(\mu' A\mu))]D_1 + \mathrm{E}[h(\chi^2_{k+4}(\mu' A\mu))]D_2$.

Proof. By using the same transformation methods as in the proof of Theorem C.1, we have

$$\mathrm{E}\big[h\big(X'AX\big)X'WX\big] = \mathrm{E}\big[h\big(V_1'V_1\big)V_1'[I_k, 0]GA^{1/2\dagger}WA^{1/2\dagger}G'[I_k, 0]'V_1\big].$$

Therefore, the proof is completed by combining Theorem 2 in Judge and Bock [9] along with some algebraic computations. □

Remark C.2. Note that Theorem C.2 generalizes Theorem 2 in Judge and Bock [9]. Indeed, if $\Sigma = I_p$, the quoted result is obtained by taking $A = I_p$.

By using Theorem C.2, we establish the following corollary.

Corollary C.2. *Let* $D_1 = \mathrm{trace}(W\Lambda_{11})$, $D_2 = \mu_2' W\mu_2$ *and suppose that the conditions of Corollary C.1 hold. Then,* $\mathrm{E}[h(\varepsilon_5' A\varepsilon_5)\varepsilon_5' W\varepsilon_5] = \mathrm{E}[h(\chi^2_{k+2}(\mu_2' A\mu_2))]D_1 + \mathrm{E}[h(\chi^2_{k+4}(\mu_2' A\mu_2))]D_2$.

Proof. This corollary directly follows from Theorem C.2. □

Theorem C.3. *Let*

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_{2p}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right),$$

where the rank of Σ_{11} is $k \le p$, with $\mu_Y = -\mu_X$, $A\Sigma_{11}A = A$; $\Sigma_{11}A\Sigma_{11} = \Sigma_{11}$; $\Sigma_{11}A\mu_X = \mu_X$. Further, we assume that $W = A^{1/2}W^*A^{1/2}$, where W^* is a nonnegative definite matrix. Then,

$$\begin{aligned}
&\mathrm{E}\big[h\big(X'AX\big)Y'WX\big] \\
&\quad = -\mathrm{E}\big[h\big(\chi^2_{k+2}\big(\mu_X' A\mu_X\big)\big)\big]\mu_X' W\mu_X - \mathrm{E}\big[h\big(\chi^2_{k+2}\big(\mu_X' A\mu_X\big)\big)\big]\mu_X' A\Sigma_{12}W\mu_X \\
&\qquad + \mathrm{E}\big[h\big(\chi^2_{k+2}\big(\mu_X' A\mu_X\big)\big)\big]\,\mathrm{trace}(\Lambda_{12}W\Lambda_{11}A) + \mathrm{E}\big[h\big(\chi^2_{k+4}\big(\mu_X' A\mu_X\big)\big)\big]\mu_X' A\Lambda_{12}W\mu_X.
\end{aligned}$$

Proof. Using the similar transformation methods as in proof of Theorem C.1, we have

$$\mathrm{E}\big[h\big(X'AX\big)Y'WX\big] = \mathrm{E}\big[h\big(V_1'V_1\big)\mathrm{E}[Y|V_1]'WA^{1/2\dagger}G'[I_k, 0]'V_1\big],$$

where $\mathrm{E}[Y|V_1] = -\mu_X + \Sigma_{21}A^{1/2}G'[I_k, 0]'(V_1 - \mu_v)$. Further, from Theorem C.1,

$$\mathrm{E}\big[h\big(V_1'V_1\big)\mu_2' WA^{1/2\dagger}G'[I_k, 0]'V_1\big] = \mathrm{E}\big[h\big(\chi^2_{k+2}\big(\mu_X' A\mu_X\big)\big)\big]\mu_X' W\mu_X$$

and

$$\begin{aligned} &\mathrm{E}\big[h\big(V_1'V_1\big)\mu_v'[I_k,0]GA^{1/2}\Sigma_{12}WA^{1/2\dagger}G'[I_k,0]'V_1\big] \\ &\quad = \mathrm{E}\big[h\big(\chi^2_{k+2}\big(\mu_X'A\mu_X\big)\big)\big]\mu_X'A\Sigma_{12}WA^{1/2\dagger}\mu_X, \end{aligned}$$

and the proof is completed by some algebraic computations. □

By using this theorem, we establish the following corollary.

Corollary C.3. *With* ε_5 *and* ε_4 *defined in Lemma 3.2, and let* $\mu_2=-\mu_1$. *Then, we have*

$$\begin{aligned} &\mathrm{E}\big[h\big(\varepsilon_5'A\varepsilon_5\big)\varepsilon_4'W\varepsilon_5\big] \\ &\quad = -\mathrm{E}\big[h\big(\chi^2_{k+2}\big(\mu_2'A\mu_2\big)\big)\big]\mu_2'W\Lambda_{11}A\mu_2 - \mathrm{E}\big[h\big(\chi^2_{k+2}\big(\mu_2'A\mu_2\big)\big)\big]\mu_2'A\Lambda_{12}W\Lambda_{11}A\mu_2 \\ &\qquad + \mathrm{E}\big[h\big(\chi^2_{k+2}\big(\mu_2'A\mu_2\big)\big)\big]\,\mathrm{trace}(\Lambda_{12}W\Lambda_{11}A) \\ &\qquad + \mathrm{E}\big[h\big(\chi^2_{k+4}\big(\mu_2'A\mu_2\big)\big)\big]\mu_2'A\Lambda_{12}W\Lambda_{11}A\mu_2. \end{aligned}$$

Proof of Corollary 4.2. By some algebraic computations, we have,

$$\begin{aligned} &\mathrm{ADR}\big(\hat{\delta}^s,\delta^0,W\big) - \mathrm{ADR}\big(\hat{\delta},\delta^0,W\big) \\ &\quad = -(k-2)^2\,\mathrm{trace}\big(W(\Lambda_{11}+2\Lambda_{12})\big)\mathrm{E}\big[\chi^{-4}_{k+2}(\Delta)\big] \\ &\qquad - (k-2)\big(4\Delta C_1-(k+2)C_2\big)\mathrm{E}\big[\chi^{-4}_{k+4}(\Delta)\big], \end{aligned}$$

where $C_1=\mathrm{trace}(W(\Lambda_{11}+\Lambda_{12}))$, $C_2=\mu_1'A(\Lambda_{11}+4\Lambda_{12}/(k+2))W\mu_1$, and $C_3=\mathrm{trace}(W\Lambda_{11})$. Then, since $k\ge 2$, $\mathrm{ADR}(\hat{\delta}^s,\delta^0,W)\le \mathrm{ADR}(\hat{\delta},\delta^0,W)$ provided that $\mathrm{trace}(W(\Lambda_{11}+2\Lambda_{12}))\ge 0$ and $4\Delta C_1-(k+2)C_2\ge 0$. Note that if $C_2=0$, $4\Delta C_1-(k+2)C_2\ge 0$ holds for any $\Delta\ge 0$, and if $C_2>0$, $4\Delta C_1-(k+2)C_2\ge 0$ holds for $\Delta C_1\ge (k+2)C_2/4$, which is equivalent to $C_1\ge (k+2)C_2/(4\Delta)$.

Since $C_2=\mu_1'A(\Lambda_{11}+4\Lambda_{12}/(k+2))W\Lambda_{11}A\mu_1$, and by Courant's theorem, we have

$$\mathrm{Ch}_{\min}\big(\Pi^*\big)\le \frac{\mu_1'A(\Lambda_{11}+4\Lambda_{12}/(k+2))W\Lambda_{11}A\mu_1}{\mu_1'A\mu_1}\le \mathrm{Ch}_{\max}\big(\Pi^*\big),$$

where $\Pi^*=(\Pi_0+\Pi_0')/2$, $\Pi_0=A^{1/2}(\Lambda_{11}+4\Lambda_{12}/(k+2))W\Lambda_{11}A^{1/2}$ and $\mathrm{Ch}_{\min}(\Pi^*)$, $\mathrm{Ch}_{\max}(\Pi^*)$ are denoted as the smallest and largest eigenvalue of Π^*, respectively. Then, $4\Delta C_1-(k+2)C_2\ge 0$ holds if $C_1\ge (k+2)\,\mathrm{Ch}_{\max}(\Pi^*)/4$. In addition, since $\mathrm{trace}(W(\Lambda_{11}+2\Lambda_{12}))\ge 0$ is equivalent to $C_1\ge -\mathrm{trace}(W\Lambda_{12})$, it follows that

$$\mathrm{ADR}\big(\hat{\delta}^s,\delta^0,W\big)\le \mathrm{ADR}\big(\hat{\delta},\delta^0,W\big)$$

if $\mathrm{trace}(W(\Lambda_{11}+\Lambda_{12}))\ge \max(-\mathrm{trace}(W\Lambda_{12}),(k+2)\,\mathrm{Ch}_{\max}(\Pi^*)/4)$. Further, by similar algebraic computations, we prove that $\mathrm{ADR}(\hat{\delta}^{s+},\delta^0,W)\le \mathrm{ADR}(\hat{\delta}^s,\delta^0,W)$, this completes the proof. □

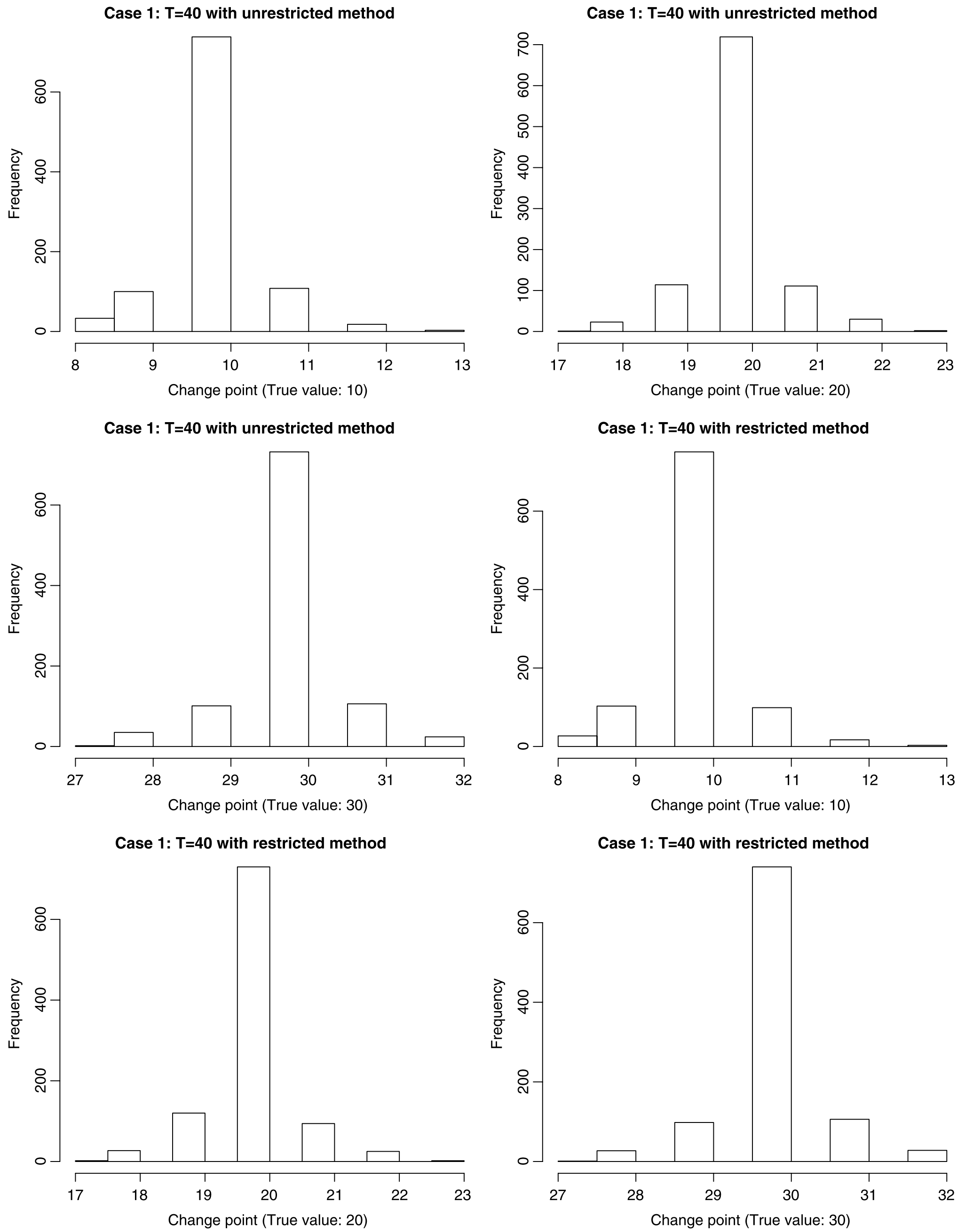

Figure 3. Histograms of the UE and RE of change points (case 1 with $T = 40$).

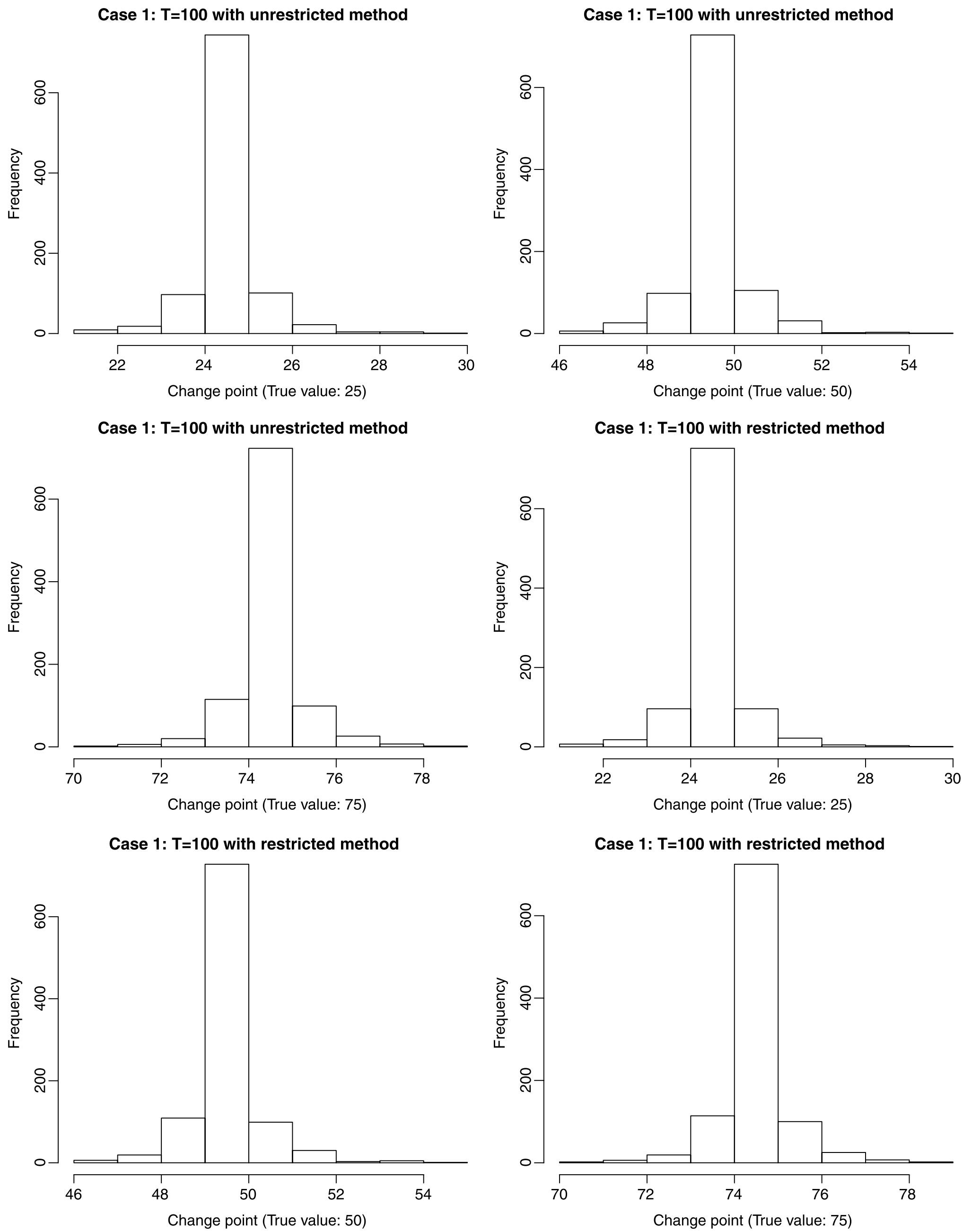

Figure 4. Histograms of the UE and RE of change points (case 1 with $T = 100$).

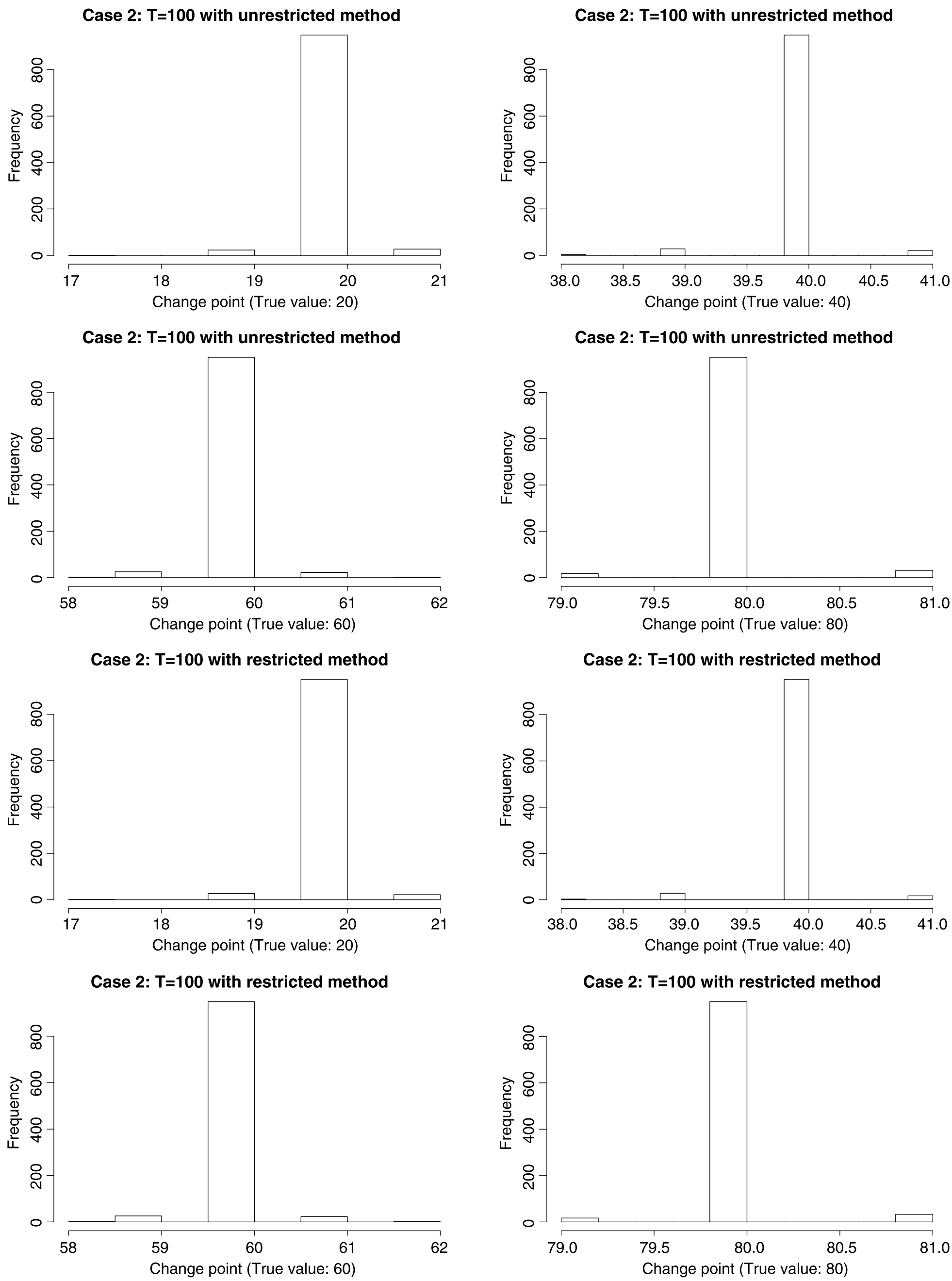

Figure 5. Histograms of the UE and RE of change points (case 2 with $T = 100$).

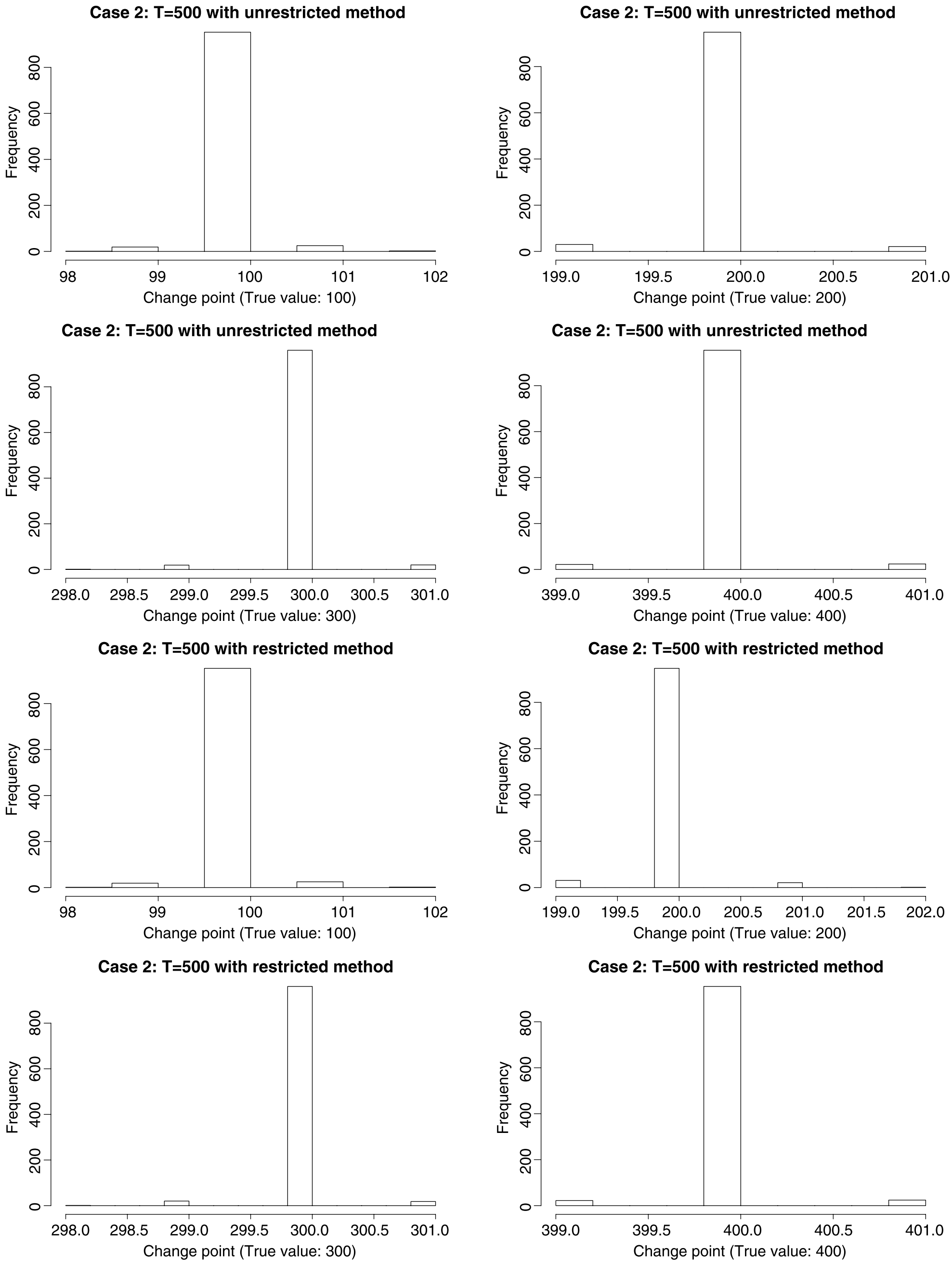

Figure 6. Histograms of the UE and RE of change points (case 2 with $T = 500$).

Acknowledgements

The authors would like to acknowledge the financial support received from Natural Sciences and Engineering Research Council of Canada. Further, the authors would like to thank anonymous referees for useful comments and suggestions.

References

[1] Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *J. Appl. Econometr.* **18** 1–22.
[2] Baranchick, A. (1964). Multiple regression and estimation of the mean of a multivariate normal distribution. Technical Report No. 51, Dept. Statistics, Stanford Univ.
[3] Braun, J.V. and Muller, H.G. (1998). Statistical methods for DNA sequence segmentation. *Statist. Sci.* **13** 142–162.
[4] Broemeling, L.D. and Tsurumi, H. (1987). *Econometrics and Structural Change. Statistics: Textbooks and Monographs* **74**. New York: Dekker, Inc. MR0922263
[5] Fu, Y.-X. and Curnow, R.N. (1990). Locating a changed segment in a sequence of Bernoulli variables. *Biometrika* **77** 295–304. MR1064801
[6] Fu, Y.-X. and Curnow, R.N. (1990). Maximum likelihood estimation of multiple change points. *Biometrika* **77** 563–573. MR1087847
[7] Hossain, S., Doksum, K.A. and Ahmed, S.E. (2009). Positive shrinkage, improved pretest and absolute penalty estimators in partially linear models. *Linear Algebra Appl.* **430** 2749–2761. MR2509855
[8] James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.* **I** 361–379. Berkeley, CA: Univ. California Press. MR0133191
[9] Judge, G.G. and Bock, M.E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. Amsterdam: North-Holland. MR0483199
[10] Judge, G.G. and Mittelhammer, R.C. (2004). A semiparametric basis for combining estimation problems under quadratic loss. *J. Amer. Statist. Assoc.* **99** 479–487. MR2062833
[11] Lombard, F. (1986). The change-point problem for angular data: A nonparametric approach. *Technometrics* **28** 391–397.
[12] McLeish, D.L. (1977). On the invariance principle for nonstationary mixingales. *Ann. Probab.* **5** 616–621. MR0445583
[13] Nkurunziza, S. (2011). Shrinkage strategy in stratified random sample subject to measurement error. *Statist. Probab. Lett.* **81** 317–325. MR2764300
[14] Nkurunziza, S. (2012). The risk of pretest and shrinkage estimators. *Statistics* **46** 305–312. MR2929155
[15] Nkurunziza, S. and Ahmed, S.E. (2010). Shrinkage drift parameter estimation for multi-factor Ornstein–Uhlenbeck processes. *Appl. Stoch. Models Bus. Ind.* **26** 103–124. MR2722886
[16] Perron, P. and Qu, Z. (2006). Estimating restricted structural change models. *J. Econometrics* **134** 373–399. MR2328414
[17] Perron, P. and Yabu, T. (2009). Testing for shifts in trend with an integrated or stationary noise component. *J. Bus. Econom. Statist.* **27** 369–396. MR2554242
[18] Saleh, A.K.Md.E. (2006). *Theory of Preliminary Test and Stein-Type Estimation with Applications. Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. MR2218139
[19] Tan, Z. (2014). Improved minimax estimation of a multivariate normal mean under heteroscedasticity. *Bernoulli*. To appear.

[20] Zeileis, A., Kleiber, C., Krämer, W. and Hornik, K. (2003). Testing and dating of structural changes in practice. *Comput*. *Statist*. *Data Anal*. **44** 109–123. MR2019790

Received October 2013 and revised May 2014