2022

On Estimation Methods in Tensor Regression Models

Mai Ghannam
University of Windsor

On estimation methods in tensor regression models

by

Mai Ghannam

A Dissertation

Submitted to the Faculty of Graduate Studies

through the Department of Mathematics and Statistics

in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy at the

University of Windsor

Windsor, Ontario, Canada

On estimation methods in tensor regression models

by

Mai Ghannam

APPROVED BY:

P. Song, External Examiner
University of Michigan

D. Li
Department of Economics

A. Hussein
Department of Mathematics and Statistics

M. Belalia
Department of Mathematics and Statistics

S. Nkurunziza, Advisor
Department of Mathematics and Statistics

September 9, 2022

Declaration of Co-Authorship/Previous Publication

I. Co-Authorship Declaration

I hereby certify that this dissertation incorporates the outcome of joint research undertaken in collaboration with my supervisor, Dr. Sévérien Nkurunziza. In all cases, the primary contributions, simulations, data analysis and interpretation, were performed by the author, and the contribution of co-authors was primarily through the provision of some theoretical results. I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my dissertation, and have obtained written permission from each of the co-author(s) to include the above material(s) in my dissertation. I certify that, with the above qualification, this dissertation, and the research to which it refers, is the product of my own work.

II. Declaration of Previous Publication

This dissertation includes three original papers that have been previously published / submitted for publication in peer reviewed journals, as follows:

Dissertation Chapter	Publication title/full citation	Publication Status
Chapter 2, Appendix C	Ghannam, M., and Nkurunziza, S., (2022). The risk of tensor Stein-rules in elliptically contoured distributions, *Statistics*, 56:2, 421-254.	Published
Chapter 3	Ghannam, M., and Nkurunziza, S., (2022). Improved estimation in tensor regression with multiple change-points, *Electronic Journal of Statistics*, 16:2, 4162-4221.	Published
Chapter 4	Ghannam, M., and Nkurunziza, S., (2022). Tensor Stein-rules in a generalized tensor regression model, *Bernoulli*.	Under Review

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my dissertation. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor. I certify that, to the best of my knowledge, this dissertation does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in this dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in this dissertation. I declare that this is a true copy of my dissertation, including any final revisions, as approved by the committee of my dissertation and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.

Abstract

In this dissertation, we consider two estimation problems in some tensor regression models. The first estimation problem is about the tensor coefficient in a tensor regression model with multiple and unknown change-points. We generalize some recent findings in five ways. First, the problem studied is more general than the one in context of a matrix parameter with multiple change-points. Second, we develop asymptotic results of the tensor estimators in the context of a tensor regression with unknown change-points. Third, we construct a class of shrinkage tensor estimators that encompasses the unrestricted estimator (UE) and the restricted estimator (RE). Fourth, we generalize some identities which are crucial in deriving the asymptotic distributional risk (ADR) of the tensor estimators. Fifth, we show that the proposed shrinkage estimators (SEs) perform better than the UE. Finally, the theoretical results are corroborated by the simulation findings and by applying our methods to a real data analysis of MRI and fMRI datasets.

The second estimation problem is about the tensor regression coefficient in the context of a generalized tensor regression model with multi-mode covariates. We generalize the main results in recent literature in four ways. First, we weaken assumptions underlying the main results of the previous works. In particular, the dependence structure of the error and covariates are as weak as an $\mathcal{L}^2-$mixingale array, and the error term does not need to be uncorrelated with regressors. Second, we consider a more general constraint than the

one considered in previous works. Third, we establish the asymptotic properties of the tensor estimators. Specifically, we derive the joint asymptotic distribution of the unrestricted tensor estimator (UE) and the restricted tensor estimator (RE). Fourth, we propose a class of shrinkage-type estimators in the context of tensor regression, and under a general loss function, we derive sufficient conditions for which the shrinkage estimators dominate the UE. In addition to these interesting contributions, we derive a kind of functional central limit theorem for vector-valued mixingales and we establish some identities which are useful in studying the risk dominance of shrinkage-type tensor estimators. Finally, to illustrate the application of the proposed methods, we corroborate the results by some simulation studies of binary, Normal and Poisson data and we analyze a multi-relational network and neuro-imaging datasets.

To my mother.

This would not have been possible without your immense support.

Thank you!

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Nkurunziza, who inspired me to pursue graduate studies in statistics. Thank you for all the support and time you have invested in my academic work and for always being available for help. I sincerely appreciate your ongoing guidance and advice. You are an incredible teacher and role model.

I would also like to thank the external examiner, Dr. Peter Song from the University of Michigan and my advisory committee members, Dr. Hussein, Dr. Belalia, and Dr. Li. I appreciate the time you have spent reading my dissertation and I am grateful for your useful comments and suggestions.

I am also very grateful to my wonderful department for providing such a welcoming environment. Over the years, my peers, faculty and staff at the Department of Mathematics and Statistics have become my second family. Thank you for your encouragement and support.

I would also like to express my gratitude to my wonderful husband. Thank you for your understanding and patience during my busiest and most stressful times. Your help with the kids and comforting support made this journey much easier.

Last but not least, I would like to express my deepest gratitude to my mother. You have been through so much in life and yet you always find time to support my siblings and I in all of our endeavours. Thank you for believing in me and for pushing me to be the best

version of myself. I am especially grateful to you for instilling in me my love for God and for reminding me to always do good. Thank you for being such an incredible grandmother to my children. My children and I have become much better people because you are in our lives.

Contents

List of Tables

List of Figures

Nomenclature

$\mathbb{X}$	tensor notation
$\boldsymbol{X}$	matrix notation
x	vector notation
$\boxplus_{(d)}$	operator that stacks two equal sized tensors along d-th dimension
$\mathrm{Vec}(\cdot)$	operator that stacks all columns of a tensor into one column vector
$\times_d$	mode-(d) matrix product of a tensor by a matrix
$\mathbb{X}_{(n)}$	mode-n matrix of the tensor $\mathbb{X}$
$\boldsymbol{A} \otimes \boldsymbol{B}$	Kronecker product of two matrices
$\mathbb{A}(\bigotimes_{j=1}^{d})_j C_j$	$= \mathbb{A} \times_1 C_1 \times_2 C_2 \times_3 \cdots \times_d C_d$
$\bigotimes_{i=d}^{1} C_i$	$= C_d \otimes C_{d-1} \otimes \cdots \otimes C_1$
$\mathbb{A} \circ \mathbb{B}$	tensor product of two tensors $\mathbb{A}$ and $\mathbb{B}$
$\mathcal{N}_{q_1 \times \cdots \times q_d}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$q_1 \times \cdots \times q_d$ normal tensor distribution with tensor mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\chi_l^2(\Delta)$	non-central chi-squared distribution with l degrees of freedom and non-centrality parameter Δ

χ_l^2	central chi-squared distribution with l degrees of freedom
$\mathscr{E}_{q_1\times q_2\times\cdots\times q_d}(\boldsymbol{\mu},\boldsymbol{\Sigma};g)$	elliptically contoured distribution where $\boldsymbol{\mu}$ is a $q_1\times q_2\times\cdots\times q_d$ non-random tensor, $\boldsymbol{\Sigma}$ is a positive definite matrix and the pdf generator g
$\xrightarrow[T\to\infty]{\mathcal{L}^2}$	convergence in $\mathcal{L}^2$
$\xrightarrow[T\to\infty]{a.s}$	convergence almost surely
$\xrightarrow[T\to\infty]{\mathrm{P}}$	convergence in probability
$\xrightarrow[T\to\infty]{d}$	convergence in distribution
$o_p(a)$	$o_p(a)/a$ converges in probability to 0
$O_p(a)$	$O_p(a)/a$ is bounded in probability
$o(a)$	$o(a)/a$ converges to 0
$O(a)$	$O(a)/a$ is bounded
$\lVert\mathbb{X}\rVert_r$	$\mathscr{L}^r-$norm of random tensor variate
$\lVert\mathbb{X}\rVert$	$\mathscr{L}^2-$norm of random tensor variate
$\lVert\mathbb{X}\rVert_{\mathrm{F}}$	Frobenius norm of tensor

Chapter 1

Introduction

1.1 Introduction

Today, a growing field in statistical modeling is the analysis of multi-dimensional observations known as tensors or arrays. Examples of tensor data include three-dimensional neuro-images of subjects called magnetic resonance images (MRIs), four-dimensional neuro-imaging time series known as functional magnetic resonance images (fMRIs) and many relational networks and nodal connections are often summarized as array structures. An important aspect in most array data includes the spatial structure of the tensors and the relationships among neighbouring components. These relationships and spatial interpretation may be lost if the tensor array structure is broken through vectorization. For example, suppose we are interested in finding associations between particular brain regions and a disease such as schizophrenia. Breaking the tensor structures of the neuro-imaging data would result in a loss of locations of interest in which abnormalities may somehow explain the development of the disease. As such, classical regression fails to offer crucial information about neighbouring voxels and their relationships. Consequently, developing regression methods that preserve the spatial structure of the tensor is of great interest in tensor analysis.

To give some related references on tensor regression, we quote Penny et al. (2011) who regressed individual voxels of images. However, this removes the ability to intuitively understand the spatial structure of the image and ignores any possible correlation among neighbouring voxels. Other authors have tried to incorporate some correlation among neighbouring voxels in their analysis. To give some examples, we quote Li et al. (2010) and Skup et al. (2012). In particular, rather than jointly analyzing all the voxels, the smallest distinguishable 3D components of an image tensor, these authors estimated parameters by iteratively increasing the sphere of neighbouring voxels around each voxel and combined the responses of those neighbours using weights. Nevertheless, none of these authors have successfully analyzed the entire image.

Some other authors have attempted to keep the structure of the predictor tensor. We quote for example Zhang and Li (2017) who proposed a tensor envelope partial least squares algorithm to tackle the high-dimensional problem in tensor regression models. Further, Li and Zhang (2017) proposed a parsimonious tensor response regression model with a dimension reducing envelope method. We also quote Li et al. (2018) who proposed a tensor regression model based on the Tucker decomposition to reduce the dimensionality of the tensor coefficient. Another related good reference is the work of Zhou et al. (2013) who proposed a tensor regression model based on CP decomposition and a block-relaxation algorithm to overcome the problem of the ultra-high dimensional setting. We also quote Guhaniyogi et al. (2017) who proposed a Bayesian approach to study the tensor regression using multiway shrinkage priors.

Although the above works offered tremendous advancements in the analysis of tensor data, none of the quoted papers have considered the problem of change-points. In time series analysis, the change-point problem is a well known issue in statistical modelling as some external/internal phenomena might change the data substantially in a certain time spot due to unconventional shocks. Ignoring the existence of change-points in a model may lead to wrong statistical conclusions. To give some recent references about change-point problems and related issues, we quote for example Qu and Perron (2007), Gombay (2010), Robbins et al. (2011), Gallagher et al. (2012), Woody and Lund (2014), Chen and Nkurunziza (2015), Chen and Nkurunziza (2016), Roy et al. (2017), and references therein. The first estimation problem of the dissertation generalizes the concepts of the quoted papers in the context of a tensor regression model with multiple and unknown number of change-points.

In addition, the vast majority of literature available on tensor model estimation assume independent and identically distributed Gaussian observations. This assumption is not realistic as many tensor data are not normally distributed nor independent. For example, brain connectivity networks of different regions of the brain and relational networks are summarized as dependent, binary tensor data. As such, it is of interest to develop methods that can include these types of data. Thus, in the second estimation problem of this dissertation, we consider a generalized tensor model with an arbitrary link function that includes the Gaussian assumption within its framework. As a result, the proposed model allows for other different distributions to be studied such as binomial and Poisson tensor data.

A key difference of the methods proposed in this dissertation and tensor models in literature lies in the dependence structure of tensor error terms and the regressors. Specifically, for both of the models considered in this dissertation, we take assumptions that allow for the dependence structure of the error terms to be as weak as that of $\mathcal{L}^2-$mixingales. Recall that mixingales are the generalizations of martingale sequences. The concept was first introduced by McLeish (1977) and extended by Andrews (1988). This structure can be taken to cover many types of data, including those that are identically and independently distributed and those that are neither identically nor independently distributed. Moreover, this structure allows for the scenario where the tensor error terms and the matrix of covariates are dependent. Hence, the mixingale assumption admits a vast array of possible applications, including many auto-correlated and heteroscedastic models. For more details on the mixingale concept, we refer to McLeish (1977) and Davidson (1992), and references therein.

Moreover, we propose a class of tensor shrinkage estimators for the tensor regression parameter. To the best of our knowledge, such estimators have never been proposed in the context of tensor regression models. We establish that the shrinkage estimators are robust and more efficient than the unrestricted estimator even when the restriction fails. This class of shrinkage estimators encloses the unrestricted, restricted and James-Stein shrinkage estimators as special cases. For some references about shrinkage estimation, we quote Saleh (2006), Sen and Saleh (1985), Nkurunziza (2012) and references therein.

In this dissertation, under the weak dependence structures of mixingales, we establish the asymptotic properties of the UE and RE. Subsequently, for both of the models considered, we propose some sufficient conditions for which the shrinkage estimators dominate the unrestricted estimators under some prior information. This prior information is taken as a series of multi-mode restrictions imposed on the tensor parameter of the first model and a general constraint on the tensor parameter of the second model. Moreover, in order to derive the sufficient conditions for the dominance of the shrinkage estimators, we establish some identities on quadratic forms of tensor variates that extend some recent results in literature. The established identities are useful in deriving the asymptotic optimality of the proposed class of shrinkage-type tensor estimators.

1.2 Dissertation Organization and Highlights of Contributions

In this section, we highlight the contributions of this dissertation as summarized below.

1. In Chapter 2, we generalize the identities in Judge and Bock (1978), Nkurunziza (2013) and Chen and Nkurunziza (2015, 2016) in the context of normal tensor variates and elliptically contoured tensor variates.

2. In Chapter 3, we consider the condition of $\mathcal{L}^2-$mixingale of size $-1/2$ and we generalize Lemma 3.1 and Lemma 3.2 of Chen and Nkurunziza (2016) in the context of tensors. Moreover, we establish the joint asymptotic normality for the tensor UE and RE under a sequence of local alternative multi-mode restrictions.

3. We consider a more general restriction than in Chen and Nkurunziza (2016) and incorporate the multi-mode properties of tensors as discussed in Kolda and Bader (2009).

4. We propose a class of tensor James-Stein type of shrinkage estimators which includes as special cases the tensor UE, RE and SEs.

5. Using the tensor quadratic forms identities established in Chapter 2, we derive a condition for the SEs to dominate the UE and we also derive the condition for the REs to dominate the UE.

6. In Chapter 4, we generalize the tensor model of Chapter 3 to that of a generalized tensor regression model with an arbitrary link function and multi-mode covariate matrices.

7. Moreover, in Chapter 4, we weaken some assumptions of Chapter 3. In particular, we show that weaker conditions on the dependence structure of the tensor error terms and regressors give the condition of $\mathcal{L}^2-$ mixingale of size $\max -1/2 - \varsigma$, $0 \leqslant \varsigma \leqslant 1$. In addition, we define the estimating score function of the model and prove the existence and consistency of the UE and RE and we derive their joint asymptotic distribution.

8. We define the asymptotic distributional risk with respect to a more general loss function that includes the quadratic loss function of Chapter 3 as a special case.

9. Using the identities of Chapter 2, we derive the asymptotic distributional risk of the elliptically contoured tensor distribution.

10. We show that the SEs dominate the UE with respect to the more general loss function under some sufficient conditions.

The rest of this dissertation is organized as follows. In Chapter 2, we establish some essential identities that are crucial in deriving the asymptotic risk of the proposed estimators in subsequent chapters. Chapter 3 establishes some shrinkage methods in a tensor regression model with unknown change-points. In Chapter 4, we propose a generalized tensor regression model that generalizes the model of Chapter 3 for the case where the model has no change-points. Chapter 5 presents the conclusion and future research. In addition, we give some properties of tensors and we present some technical results in Appendix B and Appendix C.

Chapter 2

Some Useful Identities

In this chapter, we establish some useful identities about functions of quadratic forms of random tensor variates. Such identities are useful in computing the asymptotic distributional risk of the proposed shrinkage estimators presented in Chapter 3 and Chapter 4. Specifically, in Section 2.1, we establish identities for normal random tensors and in Section 2.2, we expand the results to the case of the family of elliptically contoured distributions.

To set up some notations, let $\mathbb{A} \boxplus_{(d)} \mathbb{B}$ represent the concatenation/stacking of the equal-sized tensors $\mathbb{A}$ and $\mathbb{B}$ along the d-th dimension. Let $\text{Vec}(\cdot)$ be the operator that stacks all columns of a tensor into one column vector and let $\times_d$ represent the mode-(d) matrix product of a tensor by a matrix. Let $\mathbb{A}_{(n)}$ denote the mode-n matrix of the tensor $\mathbb{A}$. For more details about mode-(d) tensor-matrix product, we refer to Kolda and Bader (2009) and Kolda (2006). Further, let $A \otimes B$ denote the Kronecker product of two matrices A and B. For a tensor $\mathbb{A}$ and matrices C_j, $j = 1, \cdots, d$, let $\mathbb{A}(\bigtimes_{j=1}^{d})_j C_j = \mathbb{A} \times_1 C_1 \times_2 C_2 \times_3 \cdots \times_d C_d$ and let $\bigotimes_{i=d}^{1} C_i = C_d \otimes C_{d-1} \otimes \cdots \otimes C_1$. Let $\mathbb{A} \circ \mathbb{B}$ be the tensor product of two tensors $\mathbb{A}$ and $\mathbb{B}$. Note that for the special case where $\mathbb{A}$ and $\mathbb{B}$ are vectors, this tensor product becomes the vector outer product. For more information on the tensor product, we refer to Kolda (2006) and Kolda and Bader (2009). We let $\mathbb{X} \sim \mathcal{N}_{q_1\times\cdots\times q_d}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a $q_1 \times \cdots \times q_d$ normal tensor variate with tensor mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In particular, we let $x \sim \mathcal{N}_q(\mu, \Sigma)$ denote a $q-$column normal vector with mean vector μ and covariance matrix Σ. Furthermore, we denote $\chi^2_l(\Delta)$ to be a non-central chi-squared random variable with l degrees of freedom and non-centrality parameter Δ and we let χ^2_l denote a central chi-squared random variable with l degrees of freedom. To simplify some notations, let

$$\psi^{(1)}_{i,n}(x) = \int_0^\infty \mathrm{E}[h^i(t^{-1}\chi^2_n(tx))]\omega(t)dt, \quad \psi^{(2)}_{i,n}(x) = \int_0^\infty t^{-1}\mathrm{E}[h^i(t^{-1}\chi^2_n(tx))]\omega(t)dt, \ x \geq 0, \quad (2.1)$$

and set $c = \psi^{(2)}_{0,1}(x) = \psi^{(2)}_{0,n}(x), x \geq 0$. Further, let $\mathbb{X}^* = \mathbb{X}(\bigtimes_{j=1}^{d})_j \Xi_j^{1/2}$ and let $q = \prod_{i=1}^{d} q_i$, $p = \prod_{i=1}^{d} p_i$.

2.1 The Gaussian Case

In this section, we present three identities about quadratic forms of Gaussian tensor variates which are useful in deriving the risk functions of the proposed estimators in Chapter 3 and Chapter 4. These identities are the tensor extensions of Theorems 1 and 2 of Judge and Bock (1978) and they are generalizations of Theorems B.1-B.3 of Chen and Nkurunziza (2016).

We first establish Theorem 2.1.1 which generalizes Theorem B.1 of Chen and Nkurunziza (2016).

Theorem 2.1.1. *Let h be a Borel measurable and real-valued integrable function. Let $\mathbb{X}$ be a $q_1 \times \cdots \times q_{d+1}$ random tensor such that $\mathbb{X} \sim \mathcal{N}_{q_1\times\cdots\times q_{d+1}}\left(\mathbb{M}, \bigotimes_{i=d+1}^{1} \Lambda_i\right)$, where $\mathbb{M}$ is a non-random tensor and Λ_i, $i = 1, \cdots, d+1$, are positive definite matrices. Suppose that there* *exist* $d+1$ *symmetric and nonnegative definite matrices,* $\mathbb{W}_j = \Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2}$, $j = 1, \cdots, d+1$, *where* $\mathbb{W}_j^*$ *are symmetric and nonnegative definite matrices,* Ξ_j *are nonnegative definite matrices with rank* l_j *for* $j = 1, \cdots, d+1$ *and suppose that for* $j = 1, \cdots, d+1$, $\Lambda_j\Xi_j$ *are idempotent;* $\Xi_j\Lambda_j\Xi_j = \Xi_j; \Lambda_j\Xi_j\Lambda_j = \Lambda_j; \mathbb{M} \times_j (\Lambda_j\Xi_j) = \mathbb{M}$. *Then,*

$$\mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^*\right)\right)\mathbb{X}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j\right] = \mathrm{E}\left[h\left(\chi^2_{l_1\cdots l_{d+1}+2}\left(\mathrm{trace}\left(\mathbb{M}_{(d)}^{*\prime}\mathbb{M}_{(d)}^*\right)\right)\right)\right]\left(\mathbb{M}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j\right).$$

The proof of Theorem 2.1.1 is given in Appendix A. From Theorem 2.1.1 we establish the next theorem which extends Theorem B.2 of Chen and Nkurunziza (2016).

Theorem 2.1.2. *Let* $\mathbb{X}^{**} = \mathbb{X}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}$, $\mathbb{M}_{11}^* = \mathbb{M}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}$, $D_1 = \prod_{j=1}^{d+1}\mathrm{trace}\left(\mathbb{W}_j\Lambda_j\right)$, *and let* $D_2 = \mathrm{trace}\left(\mathbb{M}_{11_{(d)}}^{*\prime}\mathbb{M}_{11_{(d)}}^*\right)$. *Under the conditions of Theorem 2.1.1, we have*

$$\begin{aligned}\mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^*\right)\right)\mathrm{trace}\left(\mathbb{X}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**}\right)\right] &= \mathrm{E}\left[h(\chi^2_{l_1\cdots l_{d+1}+2}\left(\mathrm{trace}\left(\mathbb{M}_{(d)}^{*\prime}\mathbb{M}_{(d)}^*\right)\right)\right]D_1 \\ &+ \mathrm{E}\left[h(\chi^2_{l_1\cdots l_{d+1}+4}\left(\mathrm{trace}\left(\mathbb{M}_{(d)}^{*\prime}\mathbb{M}_{(d)}^*\right)\right)\right]D_2.\end{aligned}$$

The proof of this theorem is given in Appendix A. Next, we establish Theorem 2.1.3 which generalizes Theorem B.3 of Chen and Nkurunziza (2016).

Theorem 2.1.3. *Let* $\mathbb{X} \boxplus_{(d+1)} \mathbb{Y} \sim \mathcal{N}_{q_1\times\cdots\times q_d\times 2q_{d+1}}\left(\mathbb{M}_X \boxplus_{(d+1)} -\mathbb{M}_X, \begin{pmatrix}\Pi_{11} & \Pi_{21}' \\ \Pi_{21} & \Pi_{22}\end{pmatrix}\right)$, *with* $\Pi_{11} = \bigotimes_{j=d+1}^{1}\Lambda_j$, $\Pi_{21} = \bigotimes_{j=d+1}^{1}B_j - \Pi_{11}$, *and* $\Pi_{22} = \bigotimes_{j=d+1}^{1}C_j - \bigotimes_{j=d+1}^{1}D_j - \Pi_{21}$. *Let* $\mathbb{Y}^{**} = \mathbb{Y}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}$ *and, for* $j = 1, \cdots, d+1$, *let* $\mathbb{W}_j = \Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2}$ *where* $\mathbb{W}_j^*$ *and* Ξ_j *are non-negative definite matrices such that* $\Xi_j\Lambda_j\Xi_j = \Xi_j$, $\Lambda_j\Xi_j\Lambda_j = \Lambda_j$ *and* $\mathbb{M}_X \times_j \Lambda_j\Xi_j =$

$\mathbb{M}_X$. *Let h be as in Theorem 2.1.2. Then,*

$$\begin{aligned}
& \mathrm{E}\left[h\left(\operatorname{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right)\right)\operatorname{trace}\left(\mathbb{Y}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**}\right)\right] \\
= \; & \mathrm{E}\left[h\left(\chi_{l+2}^{2}\left(\operatorname{trace}\left(\mathbb{M}_{X_{(d)}}^{**\prime}\mathbb{M}_{X_{(d)}}^{**}\right)\right)\right)\right]\left(\prod_{j=1}^{d+1}\operatorname{trace}(\mathbb{W}_j B_j)-\prod_{j=1}^{d+1}\operatorname{trace}(\mathbb{W}_j \Lambda_j)\right) \\
- \; & \mathrm{E}\left[h\left(\chi_{l+2}^{2}\left(\operatorname{trace}\left(\mathbb{M}_{X_{(d)}}^{**\prime}\mathbb{M}_{X_{(d)}}^{**}\right)\right)\right)\right]\operatorname{trace}\left(\mathbb{M}_{X_{(d)}}'\left(\mathbb{M}_X(\bigtimes_{j=1}^{d+1})_j\Xi_j B_j\mathbb{W}_j\right)_{(d)}\right) \\
+ \; & \mathrm{E}\left[h\left(\chi_{l+4}^{2}\left(\operatorname{trace}\left(\mathbb{M}_{X_{(d)}}^{**\prime}\mathbb{M}_{X_{(d)}}^{**}\right)\right)\right)\right]\operatorname{trace}\left(\mathbb{M}_{X_{(d)}}'\left(\mathbb{M}_X(\bigtimes_{j=1}^{d+1})_j\Xi_j B_j\mathbb{W}_j\right)_{(d)}\right) \\
- \; & \mathrm{E}\left[h\left(\chi_{l+4}^{2}\left(\operatorname{trace}\left(\mathbb{M}_{X_{(d)}}^{**\prime}\mathbb{M}_{X_{(d)}}^{**}\right)\right)\right)\right]\operatorname{trace}\left(\left(\mathbb{M}_X(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\right)_{(d)}'\left(\mathbb{M}_X(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\right)_{(d)}\right).
\end{aligned}$$

The proof of Theorem 2.1.3 is given in Appendix A.

2.2 The Elliptically Contoured Distribution Case

In this section, we establish identities about quadratic forms of the tensor variates that are distributed as elliptically contoured family of distributions. To this end, in Section 2.2.1, we first introduce some notations and define the elliptically contoured family of distributions and in Section 2.2.2, we derive corresponding identities. We note that establishing results for the class of elliptically contoured distributions is motivated by the fact that the assumption of normal distribution is not always practical or realistic. Further, elliptically contoured distributions include the Gaussian distribution and maintain similar properties. For more details on elliptically contoured distributions, we refer for example to Furman and Landsman (2006), Liu et al. (2009), Landsman and Valdez (2003) and Bingham et al. (2002). The established identities are successfully used in order to derive the risk of shrinkage type tensor estimators of Chapter 4. Moreover, the established results generalize some results in Chen and Nkurunziza (2016), Chen and Nkurunziza (2015), Nkurunziza (2013), Nkurunziza (2012) as well as the classical identities in Judge and Bock (1978). For more details about the tensor elliptically contoured distribution and some practical examples, we refer to Ghannam and Nkurunziza (2022).

2.2.1 The Elliptically Contoured Family of Distributions

In this subsection, we define the elliptically contoured random tensor and provide some distributions that belong to this class of tensor distributions. To that end, we first recall the definition of an elliptically contoured distribution in the particular case of scale mixtures of normal tensor variates.

Definition 2.2.1. *A $q_1 \times q_2 \times \cdots \times q_d -$ random tensor $\mathbb{X}$ is said to be an elliptically contoured tensor variate, denoted by $\mathbb{X} \sim \mathscr{E}_{q_1 \times q_2 \times \cdots \times q_d}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$, where $\boldsymbol{\mu}$ is a $q_1 \times q_2 \times \cdots \times q_d$ non-random tensor, $\boldsymbol{\Sigma}$, is a positive definite matrix and the pdf generator g, if the pdf of* $\mathrm{Vec}(\mathbb{X})$ *can be written as*

$$\kappa(x) = \int_0^{+\infty} f_{\mathcal{N}_{q_1 \cdots q_d}}\left(\mathrm{Vec}(\boldsymbol{\mu}), z^{-1}\boldsymbol{\Sigma}\right)(x)\omega(z)dz, \tag{2.2}$$

where $f_{\mathcal{N}_n(\boldsymbol{\mu}, z^{-1}\boldsymbol{\Sigma})}$ denotes the pdf of an $n-$ dimensional Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance-variance $z^{-1}\boldsymbol{\Sigma}$, $0 < z < \infty$, $\omega(\cdot)$ is the weight function associated with the random variate elliptically contoured distribution, as defined in Gupta and Varga (1995) and Chu (1973), for example. For more details on the generator density g, we refer the reader to Gupta and Varga (1995) and Batsidis (2010). Note that the condition in (2.2) *is equivalent to assuming that the pdf of the random tensor $\mathbb{X}$ is given by*

$$\kappa(x) = \int_0^{+\infty} f_{\mathcal{N}_{q_1 \times \cdots \times q_{d-1} \times q_d}}\left(\boldsymbol{\mu}, z^{-1}\boldsymbol{\Sigma}\right)(x)\omega(z)dz. \tag{2.3}$$

Note that a $q_1 \times q_2 \times \cdots \times q_d$ random tensor $\mathbb{X}$ follows an elliptically contoured distribution if and only if $\mathrm{Vec}(\mathbb{X})$ follows elliptically contoured distribution. Thus, we denote $\mathbb{X} \sim$

$\mathscr{E}_{q_1\times\cdots\times q_d}(\boldsymbol{\mu},\boldsymbol{\Sigma};g)$ if and only if $\mathrm{Vec}(\mathbb{X}) \sim \mathscr{E}_{q_1q_2\cdots q_d}(\mathrm{Vec}(\boldsymbol{\mu}),\boldsymbol{\Sigma};g)$. Table 2.1 gives some examples of the pdf of elliptically contoured distributions with their respective weighting functions.

In order to account for the cases of degenerate random tensors obtained from mode-j linear transformations of $\mathbb{X}$, we introduce the following notation. For the sake of simplicity, let us first consider the case of random vectors and recall that a linear transformation of an elliptically contoured random vector is an elliptically contoured random vector. In other words, let Z be an n-column random vector such that $Z \sim \mathscr{E}_n(\boldsymbol{\mu},\boldsymbol{\Sigma};g)$, with $\boldsymbol{\Sigma}$ being a positive definite matrix, let $\boldsymbol{A}$ be an $m\times n$-matrix with $\mathrm{rank}(\boldsymbol{A}) = n_0 \leq \min(m,n)$, let C be an $m-$column non-random vector and let $Y = \boldsymbol{A}Z + C$, then Y is an m-column elliptically contoured random vector. Nevertheless, if $n_0 < m$, the random vector Y is degenerated and using the notation of Garcia (2005), we denote $Y \sim \mathscr{E}_m^{n_0}(\boldsymbol{A}\boldsymbol{\mu}+C, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}';g)$. If $n_0 = m$, the random vector Y has a pdf and thus, to simplify the notation, we remove the superscript i.e. $\mathscr{E}_m^{m}(\boldsymbol{A}\boldsymbol{\mu}+C, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}';g) = \mathscr{E}_m(\boldsymbol{A}\boldsymbol{\mu}+C, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}';g)$. Similarly, for a $q_1\times\cdots\times q_d$ random tensor $\mathbb{Y}$, we denote $\mathbb{Y} \sim \mathscr{E}_{q_1\times\cdots\times q_d}^{r_1\times\cdots\times r_d}(\boldsymbol{\mu}_Y,\boldsymbol{\Sigma}_Y;g)$ if and only if $\mathrm{Vec}(\mathbb{Y}) \sim \mathscr{E}_{q_1q_2\cdots q_d}^{r_1r_2\cdots r_d}(\mathrm{Vec}(\boldsymbol{\mu}_Y),\boldsymbol{\Sigma}_Y;g)$, where $\mathrm{rank}(\boldsymbol{\Sigma}_Y) = \prod_{i=1}^{d} r_i \leq \prod_{i=1}^{d} q_i$. In particular, let $\mathbb{Z} \sim \mathscr{E}_{n_1\times\cdots\times n_d}\left(\boldsymbol{\mu}, \bigotimes_{i=d}^{1}\boldsymbol{\Lambda}_i;g\right)$, then for $q_i\times n_i$ matrices $\boldsymbol{A}_i$ with $\mathrm{rank}(\boldsymbol{A}_i) = r_i$, $i = 1,2,\ldots,d$, by some algebraic computations, one can verify that $\mathbb{Y} = \mathbb{Z}\left(\bigtimes_{i=1}^{d}\right)_i\boldsymbol{A}_i + \mathbb{C} \sim \mathscr{E}_{q_1\times\cdots\times q_d}^{r_1\times\cdots\times r_d}\left(\mathrm{Vec}\left(\boldsymbol{\mu}\left(\bigtimes_{i=1}^{d}\right)_i\boldsymbol{A}_i\right), \bigotimes_{i=d}^{1}\boldsymbol{A}_i\boldsymbol{\Lambda}_i\boldsymbol{A}_i';g\right)$. In the special case where $r_i = q_i$, we get $\mathbb{Y} \sim \mathscr{E}_{q_1\times\cdots\times q_d}\left(\mathrm{Vec}\left(\boldsymbol{\mu}\left(\bigtimes_{i=1}^{d}\right)_i\boldsymbol{A}_i\right), \bigotimes_{i=d}^{1}\boldsymbol{A}_i\boldsymbol{\Lambda}_i\boldsymbol{A}_i';g\right)$.

Table 2.1: Examples of p.d.f of elliptically contoured distributions with the respective weighting functions

Distribution	$\kappa(\boldsymbol{x})$	"$\omega(z)$"
Gaussian	$(2\pi)^{-q/2}\prod_{i=1}^{d-1}\lvert\boldsymbol{\Lambda}_i\rvert^{-\frac{q_i}{2}}\lvert\boldsymbol{\Lambda}_d\rvert^{-\frac{q_d}{2}}\times \exp\left[-\frac{1}{2}\,\mathrm{trace}(g_0(\boldsymbol{x}))\right]$	$\delta(z-1)$
t with q_0 d.f.	$\frac{\Gamma\left(\frac{q_0+q}{2}\right)}{(q_0\pi)^{q/2}\Gamma\left(\frac{q_0}{2}\right)}\prod_{i=1}^{d-1}\lvert\boldsymbol{\Lambda}_i\rvert^{-\frac{q_i}{2}}\lvert\boldsymbol{\Lambda}_d\rvert^{-\frac{q_d}{2}}\times \left[1+\mathrm{trace}\left(g_0(\boldsymbol{x})\right)/q_0\right]^{-\frac{q_0+q}{2}}$	$\frac{\left(\frac{q_0 z}{2}\right)^{\frac{q_0}{2}}e^{-\frac{q_0 z}{2}}}{z\Gamma\left(\frac{q_0}{2}\right)},\ z>0$
Pearson type VII	$\prod_{i=1}^{d-1}\lvert\boldsymbol{\Lambda}_i\rvert^{-\frac{q_i}{2}}\lvert\boldsymbol{\Lambda}_d\rvert^{-\frac{q_d}{2}}\frac{\Gamma(m)}{(q_0\pi)^{q/2}\Gamma\left(m-\frac{q}{2}\right)}\times \left[1+\mathrm{trace}\left(g_0(\boldsymbol{x})\right)/q_0\right]^{-m},\ m>q/2$	$\frac{z^{m-q/2-1}\exp(-q_0 z/2)}{(q_0/2)^{q/2-m}\Gamma(m-q/2)},\ z>0$
Laplace	$\frac{\Gamma(q/2)}{(2\pi)^{q/2}\Gamma(q)}\prod_{i=1}^{d-1}\lvert\boldsymbol{\Lambda}_i\rvert^{-\frac{q_i}{2}}\lvert\boldsymbol{\Lambda}_d\rvert^{-\frac{q_d}{2}}\times \exp\left[-\frac{1}{\sqrt{2}}\left(\mathrm{trace}(g_0(\boldsymbol{x}))\right)^{\frac{1}{2}}\right]$	$\frac{\sqrt{\pi}\Gamma(q/2)}{\sqrt{2}\Gamma(q)}z^{-q-1}\exp(-(4z)^{-1}),\ z>0$

2.2.2 Identities

Remark 2.2.1. *For the Gaussian tensor-variate case, $\psi_{i,j}^{(1)}$ and $\psi_{i,j}^{(2)}$ in* (2.1) *become*

$$\psi_{i,p+j}^{(1)}(x)=\psi_{i,p+j}^{(2)}(x)=\mathrm{E}\left[h^i(\chi_{p+j}^2(x))\right],\quad x\geq 0. \tag{2.4}$$

Theorem 2.2.1. *Let $\boldsymbol{\Lambda}_i$ be $q_i\times q_i$ positive semi-definite matrices with ranks $p_i\leq q_i$, $i=1,\cdots,d$. Let $\mathbb{X}\sim\mathscr{E}_{q_1\times q_2\times\cdots\times q_d}^{p_1\times p_2\times\cdots\times p_d}\left(\mathbb{M},\bigotimes_{j=d}^{1}\boldsymbol{\Lambda}_j;g\right)$. Suppose that there exist d symmetric and nonnegative definite matrices $\mathbb{W}_j=\boldsymbol{\Xi}_j^{1/2}\mathbb{W}_j^*\boldsymbol{\Xi}_j^{1/2}$, $j=1,\cdots,d$, where $\mathbb{W}_j^*$ are symmetric and nonnegative definite matrices. Let $\boldsymbol{\Xi}_i$ be nonnegative definite matrices with rank $p_i, i=1,\cdots,d$ and suppose that for $j=1,\cdots,d$, $\boldsymbol{\Lambda}_j\boldsymbol{\Xi}_j$ are idempotent; $\boldsymbol{\Xi}_j\boldsymbol{\Lambda}_j\boldsymbol{\Xi}_j=\boldsymbol{\Xi}_j$; $\boldsymbol{\Lambda}_j\boldsymbol{\Xi}_j\boldsymbol{\Lambda}_j=\boldsymbol{\Lambda}_j$; $\mathbb{M}\times_j\left(\boldsymbol{\Lambda}_j\boldsymbol{\Xi}_j\right)=\mathbb{M}$. Then, for any Borel measurable, real-valued square-*

integrable function h, we have

$$\mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}^{*\prime}_{(d)}\mathbb{X}^{*}_{(d)}\right)\right)\mathbb{X}(\bigtimes_{j=1}^{d})_j\mathbb{W}_j\right] = \psi^{(1)}_{1,p+2}\left(\|\mathbb{M}\|^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\left(\mathbb{M}(\bigtimes_{j=1}^{d})_j\mathbb{W}_j\right).$$

The proof of Theorem 2.2.1 is given in Appendix A. Note that Theorem 2.2.1 generalizes Theorem 2.1.1. By using Theorem 2.2.1, we derive the following corollary which generalizes Theorem 3.1 in Nkurunziza (2013).

Corollary 2.2.1. *Let $\mathbf{\Lambda}_i$ be $q_i \times q_i$ positive semi-definite matrices with ranks $p_i \le q_i$, $i = 1, \cdots, d$. Let $\mathbb{X} \sim \mathcal{E}^{p_1\times\cdots\times p_d}_{q_1\times\cdots\times q_d}\left(\mathbb{M}, \bigotimes_{j=d}^{1}\mathbf{\Lambda}_j; g\right)$. Let Ξ_i be symmetric, positive definite matrices such that $\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j^{1/2}$ are idempotent matrices and $\mathbb{M}\times_d(\Xi_d\mathbf{\Lambda}_d\Xi_d) = \mathbb{M}\times_d\Xi_d$. Then, for any Borel measurable, real-valued square-integrable function h, we have*

$$\mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}^{*\prime}_{(d)}\mathbb{X}^{*}_{(d)}\right)\right)\mathbb{X}\right] = \psi^{(1)}_{1,p+2}\left(\|\mathbb{M}\|^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathbb{M}.$$

Proof. The proof follows from Theorem 2.2.1 by taking $\mathbb{W}_j = I_{p_j};\ j = 1, \cdots, d$. □

Theorem 2.2.2. *Let h be a measurable and real-valued square-integrable function and let $\mathbb{Y}^{**} = \mathbb{Y}(\bigtimes_{j=1}^{d})_j\mathbb{W}_j^{1/2}$ where $\mathbb{X} \boxplus_{(d)} \mathbb{Y} \sim \mathcal{E}^{p_1\times\cdots\times p_d}_{q_1\times\cdots\times q_{d-1}\times 2q_d}\left(\mathbb{M}\boxplus_{(d)}\mathbb{M}_2, \begin{pmatrix}\mathbf{\Pi}_{11} & \mathbf{\Pi}'_{21}\\ \mathbf{\Pi}_{21} & \mathbf{\Pi}_{22}\end{pmatrix}\right)$, with $\mathbf{\Pi}_{11} = \bigotimes_{j=d}^{1}\mathbf{\Lambda}_j$, $\mathbf{\Pi}_{21} = \bigotimes_{j=d}^{1}\boldsymbol{B}_j - \mathbf{\Pi}_{11}$, and $\mathbf{\Pi}_{22} = \bigotimes_{j=d}^{1}\boldsymbol{C}_j - \bigotimes_{j=d}^{1}\boldsymbol{D}_j - \mathbf{\Pi}_{21}$. Also, for $j = 1, \cdots, d$, let $\mathbb{W}_j = \Xi_j^{1/2}\mathbb{W}_j^{*}\Xi_j^{1/2}$ where $\mathbb{W}_j^{*}$ and Ξ_j are non-negative definite matrices such that*

$\Xi_j \Lambda_j \Xi_j = \Xi_j$, $\quad \Lambda_j \Xi_j \Lambda_j = \Lambda_j$ *and* $\mathbb{M} \times_j \Lambda_j \Xi_j = \mathbb{M}$. *Then,*

$$\begin{aligned}
&\mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}^{*\prime}_{(d)}\mathbb{X}^{*}_{(d)}\right)\right)\mathrm{trace}\left(\mathbb{Y}^{**\prime}_{(d)}\mathbb{X}^{**}_{(d)}\right)\right] \\
&= \psi^{(1)}_{1,p+2}\left(\|\mathbb{M}\|^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\left(\mathbb{M}_2\left(\bigotimes_{j=1}^{d}\right)_j \mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mathbb{M}\left(\bigotimes_{j=1}^{d}\right)_j \mathbb{W}_j^{1/2}\right)_{(d)}\right) \\
&-\psi^{(1)}_{1,p+2}\left(\|\mathbb{M}\|^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\mathbb{M}'_{(d)}\left(\mathbb{M}\left(\bigotimes_{j=1}^{d}\right)_j \Xi_j B_j \mathbb{W}_j\right)_{(d)}\right) \\
&+\psi^{(2)}_{1,p+2}\left(\|\mathbb{M}\|^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\prod_{j=1}^{d}\mathrm{trace}(\mathbb{W}_j \boldsymbol{B}_j) - \psi^{(2)}_{1,p+2}\left(\|\mathbb{M}\|^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\prod_{j=1}^{d}\mathrm{trace}(\mathbb{W}_j \boldsymbol{\Lambda}_j) \\
&-\psi^{(1)}_{1,p+4}\left(\|\mathbb{M}\|^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\left(\mathbb{M}\left(\bigotimes_{j=1}^{d}\right)_j \mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mathbb{M}\left(\bigotimes_{j=1}^{d}\right)_j \mathbb{W}_j^{1/2}\right)_{(d)}\right) \\
&+\psi^{(1)}_{1,p+2}\left(\|\mathbb{M}\|^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\left(\mathbb{M}\left(\bigotimes_{j=1}^{d}\right)_j \mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mathbb{M}\left(\bigotimes_{j=1}^{d}\right)_j \mathbb{W}_j^{1/2}\right)_{(d)}\right) \\
&+\psi^{(1)}_{1,p+4}\left(\|\mathbb{M}\|^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\mathbb{M}'_{(d)}\left(\mathbb{M}\left(\bigotimes_{j=1}^{d}\right)_j \Xi_j B_j \mathbb{W}_j\right)_{(d)}\right).
\end{aligned}$$

The proof of this theorem is given in Appendix A. Note that Theorem 2.2.2 generalizes Theorem 2.1.2. From Theorem 2.2.2, we derive the following corollary which generalizes Theorem 3.2 of Nkurunziza (2013).

Corollary 2.2.2. *Let* $\mathbb{X} \sim \mathscr{E}^{p_1\times\cdots\times p_d}_{q_1\times\cdots\times q_d}\left(\mathbb{M}, \bigotimes_{j=d}^{1}\boldsymbol{\Lambda}_j; g\right)$ *where* $\boldsymbol{\Lambda}_i$ *are the same as in Corollary 2.2.1. Let* $\mathbb{Y}$ *be* $q_1\times\cdots\times q_d$ *random tensor with elliptically contoured distribution such that* $\mathrm{E}[\mathbb{Y}|\mathbb{X}] = \mathrm{E}[\mathbb{Y}] = \mathbb{M}_2$. *Then, for any Borel-measurable real-valued square-integrable function h and any positive definite matrix* $\boldsymbol{A}$, *we have*

$$\mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}^{*\prime}_{(d)}\mathbb{X}^{*}_{(d)}\right)\right)\mathrm{trace}\left(\mathbb{Y}'_{(d)}\boldsymbol{A}\mathbb{X}_{(d)}\right)\right] = \psi^{(1)}_{1,p+2}\left(\|\mathbb{M}\|^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\mathbb{M}'_{2_{(d)}}\boldsymbol{A}\mathbb{M}_{(d)}\right).$$

Proof. The proof follows from Theorem 2.2.2 by taking $\boldsymbol{B}_j = \boldsymbol{\Lambda}_j$, $\mathbb{W}_d = \boldsymbol{A}$ and $\mathbb{W}_i = I_{p_i}$, for $i = 1, \cdots, d-1$. □

Theorem 2.2.3. *Let* $\mathbb{X}^{**} = \mathbb{X}(\bigtimes_{j=1}^{d})_j \mathbb{W}_j^{1/2}$, $\mathbb{M}^* = \mathbb{M}(\bigtimes_{j=1}^{d})_j \mathbb{W}_j^{1/2}$, $D_1 = \prod_{j=1}^{d} \text{trace}\left(\mathbb{W}_j \mathbf{\Lambda}_j\right)$, *and let* $D_2 = \text{trace}\left(\mathbb{M}^{*\prime}_{(d)} \mathbb{M}^*_{(d)}\right)$. *Under the conditions of Theorem 2.2.1, we have*

$$\mathrm{E}\left[h\left(\text{trace}\left(\mathbb{X}^{*\prime}_{(d)}\mathbb{X}^{*}_{(d)}\right)\right)\text{trace}\left(\mathbb{X}^{**\prime}_{(d)}\mathbb{X}^{**}_{(d)}\right)\right] = \psi^{(2)}_{1,p+2}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)D_1 + \psi^{(1)}_{1,p+4}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)D_2.$$

The proof of Theorem 2.2.3 is given in Appendix A. It should be noted that Theorem 2.2.3 generalizes Theorem 2.1.3. From Theorem 2.2.3, we derive the following corollary which extends Theorem 3.3 in Nkurunziza (2013).

Corollary 2.2.3. *Let* Ξ_i *and* $\mathbf{\Lambda}_i$ *be as in Corollary 2.2.1, and let* $\mathbb{X} \sim \mathscr{E}_q^p\left(\mathbb{M}, \bigotimes_{j=d}^{1} \mathbf{\Lambda}_j; g\right)$. *Then, for any Borel-measurable real-valued square-integrable function h and any symmetric positive definite matrix* $\boldsymbol{A}$,

$$\mathrm{E}\left[h\left(\text{trace}\left(\mathbb{X}^{*\prime}_{(d)}\mathbb{X}^{*}_{(d)}\right)\right)\text{trace}\left(\mathbb{X}'_{(d)}\boldsymbol{A}\mathbb{X}_{(d)}\right)\right] = \psi^{(1)}_{1,p+4}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\text{trace}\left(\mathbb{M}'_{(d)}\boldsymbol{A}\mathbb{M}_{(d)}\right)$$
$$+ \psi^{(2)}_{1,p+2}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\text{trace}(\boldsymbol{A}\mathbf{\Lambda}_d)\prod_{j=1}^{d-1}\text{trace}\left(\mathbf{\Lambda}_j\right).$$

Proof. The proof follows from Theorem 2.2.3 by taking $\mathbb{W}_d = \boldsymbol{A}$ and $\mathbb{W}_i = I_{p_i}$, for $i = 1, \cdots, d-1$. □

Remark 2.2.2. *Note that Theorems 2.2.1-2.2.3 generalize Theorems 3.1-3.3 of Nkurunziza (2013) which become special cases with* $d = 2$.

2.3 Conclusion

In summary, in this chapter, we established some useful identities about functions of quadratic forms of normal and elliptically contoured tensor variates. These identities are crucial for deriving the asymptotic distributional risks of the proposed estimators of Chapter 3 and Chapter 4.

Chapter 3

Tensor Regression With Multiple Change-points

In this chapter, we introduce a tensor regression model with multiple change-points. This model is a generalization of the models in Chen and Nkurunziza (2015) and Chen and Nkurunziza (2016) as it involves tensor observations, parameters, error terms and the number of change-points may be unknown. Moreover, the restriction incorporated is of a much more general form. We also propose a class of tensor shrinkage estimators to include possible prior knowledge through the restriction. The special cases of shrinkage estimators known as the James-Stein estimators are then shown to be more efficient estimators than the unrestricted estimator irrespective on the accuracy of the prior information. This is established both theoretically and through simulation studies and through the analysis of real neuroimaging datasets.

To give some references on tensor regression, we quote Liu et al. (2019) who proposed ridge regression for tensor labels for hyperspectral image classification using CP (Candecomp/Parafac) tensor decomposition, Raskutti et al. (2015) who considered a general class of convex regularization techniques to exploit sparsity and low-rankness of a coefficient tensor and Hoff (2015) extended bilinear regression to predicting a tensor from another tensor using Tucker product. As other interesting references, Xu et al. (2019) developed a likelihood procedure to estimate tensor coefficients in a classical generalized linear model with multi-mode covariates, and Li and Zhang (2017) used a technique known as the envelope method that identifies immaterial information. We also quote Zhou et al. (2013) who implemented a block relaxation algorithm involving CP decomposition of a tensor coefficient, and we cite Lock (2018) who proposed a method where a tensor can be predicted from a tensor covariate by solving a least squares penalty function minimization problem. Further, Li et al. (2018) considered tensor regression using Tucker decomposition on scalar response of a exponential family of distributions, and Zhang and Li (2017) developed a tensor envelope partial least squares regression.

Our work is different in several ways. First, our method incorporates the change-points framework and we consider a very general problem and a more general restriction than that, for example, in Chen and Nkurunziza (2016). The restriction studied is especially useful for taking into account for some prior knowledge about the tensor predictor and/or for testing the statistical significance of some coefficients. Second, we propose a class of tensor shrinkage estimators which, to the best of our knowledge, have never been proposed in the context of tensor regression model. The proposed estimation method is robust and flexible as it is shown to preserve a very good performance in the context of uncertainty about the restriction and/or the significance of some components of the tensor coefficient. To this end, we establish the asymptotic properties of the estimators of the tensor regression coefficient under weaker assumptions on the error terms and regressors. In particular, unlike the above quoted works on the tensor model, the established results hold in the general context where the observations are not necessarily independent nor identically distributed. We also show, theoretically and by simulations, that the proposed tensor shrinkage estimators (SEs) perform better than the unrestricted estimator (UE). Furthermore, we tackled an additional issue related to the fact that the dimensions of the proposed tensor estimators are random variables.

The rest of this chapter is organized as follows. In Section 3.1, we describe the statistical model and state some preliminary results. In Section 3.2, we present the joint asymptotic distribution of the unrestricted estimator (UE) and the restricted estimator (RE). In Section 3.3, we introduce a class of tensor shrinkage estimators (SE) and the conditions under which the SEs dominate the UE. In Section 3.4, we study the estimation problem in the case of an unknown number of change-points. In Section 3.5, we present some simulation studies and we analyse an MRI dataset as well as an fMRI dataset. In Section 3.6, we present some concluding remarks.

3.1 Statistical model and preliminary results

In this section, we present the statistical model and the main regularity conditions. Let $(\Omega, \mathscr{F}, P)$ be a probability space and let $\{\mathscr{F}^i_{T,k}, -\infty \leq k \leq i \leq \infty, T \geq 1\}$ be a complete filtration. We consider the tensor regression model with T observations and m_0 unknown change-points $1 < \tau_1 < \cdots < \tau_{m_0} < T$. In Sections 3.1-3.3, we focus on the case where the number of change-points m_0 is known. In Section 3.4, we outline the estimation method of m_0 and show that the method produces a consistent estimator. For convenience, let τ_0=1 and $\tau_{m_0+1} = T$. Then, the model of interest is

$$\mathbb{Y} = \boldsymbol{\delta} \times_{d+1} \bar{Z} + \mathbb{U}, \tag{3.1}$$

where $\mathbb{Y} = \mathbb{Y}_1 \boxplus_{(d+1)} \mathbb{Y}_2 \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \mathbb{Y}_T \in \mathbb{R}^{q_1 \times \cdots \times q_d \times T}$, $\mathbb{U} = \mathbb{U}_1 \boxplus_{(d+1)} \mathbb{U}_2 \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \mathbb{U}_T \in \mathbb{R}^{q_1 \times \cdots \times q_d \times T}$, $\boldsymbol{\delta} = \mathbb{B}_1 \boxplus_{(d+1)} \mathbb{B}_2 \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \mathbb{B}_{m_0+1} \in \mathbb{R}^{q_1 \times \cdots \times q_d \times (m_0+1) q_{d+1}}$, $\bar{Z} = \mathrm{diag}(Z_1, \cdots, Z_{m_0+1})$ with $Z_1 = (z_1, \cdots, z_{\tau_1})'$, and for $j = 2, 3, \cdots, m_0 + 1$, $Z_j = (z_{\tau_{j-1}+1}, \cdots, z_{\tau_j})'$, $z_{\tau_{i-1}+1}$ is a q_{d+1}-

column vector for $i = 1, \cdots, m_0 + 1$.

Here, $\mathbb{Y}$ is the stacked tensor of observations, $\mathbb{U}$ is the stacked tensor of error terms, $\boldsymbol{\delta}$ is the stacked parameter to be estimated and $\bar{\boldsymbol{Z}}$ is the corresponding covariate matrix. It is important to note that, unlike classical regression and other tensor regression models in literature, we do not make any assumptions about the distribution or dependence structure of the tensor error terms, $\mathbb{U}_i$. Moreover, the model introduces change-points with different possible regimes for different groups of observations.
We also consider the case where $\boldsymbol{\delta}$ may satisfy the following restriction

$$\boldsymbol{\delta} \times_1 R_1 \times_2 R_2 \times_3 \cdots \times_{d+1} R_{d+1} = \boldsymbol{r}, \tag{3.2}$$

where for $j = 1, \cdots, d$, R_j is a known $l_j \times q_j$ matrix with rank l_j, $l_j \leqslant q_j$ and R_{d+1} is a known $l_{d+1} \times (m_0 + 1)q_{d+1}$ known matrix with rank $l_{d+1} \leqslant (m_0 + 1)q_{d+1}$, and r is a known $l_1 \times \cdots \times l_{d+1}$ tensor. This restriction is used to incorporate some prior knowledge about the tensor parameter. In practice, the restriction in (3.2) can reflect the fact that previous statistical investigations or expert knowledge indicate that some components are not statistically significant or that there exists an association between the components of the tensor parameter. For instance, in the context of neuro-imaging data analysis, restrictions built on some prior knowledge could be useful in more efficient estimation and better understanding of the underlying brain structure for diseases. For example, in some preliminary studies, there is a suspicion that a certain brain region may be associated with having a disease. In that case, one can utilize the restrictions to pinpoint that specific region by setting up the appropriate restrictions for the first 3 dimensions, i.e., R_1, R_2 and R_3. If a brain region is suspected to have no effect on the disease, then the restriction of the fourth dimension, R_4, could be set up as the identity matrix and $\boldsymbol{r}_0$ could be set up as the zero tensor. Moreover, R_4 could be used to choose the covariates of interest such as the diagnosis indicator or age. A practical application of restrictions on a real neuro-imaging data can be found in

Subsection 3.5.2

Remark 3.1.1. *The restriction in* (3.2) *is a generalization of the restriction imposed in Chen and Nkurunziza (2016) for the matrix parameter case. Indeed, by letting* $R_1 = R$*,* $R_2 = L'$ *and* $\boldsymbol{\delta} \in \mathbb{R}^{q_1 \times (m+1) q_2}$*, the above condition reduces to* $\boldsymbol{\delta} \times_1 R_1 \times_2 R_2 = R\boldsymbol{\delta} L = \boldsymbol{r}$*, where* $\boldsymbol{r} \in \mathbb{R}^{l_1 \times l_2}$.

3.1.1 Preliminary results: the known change-points case

In this subsection, we give some preliminary results in the context where $\boldsymbol{\tau_1}, \cdots, \boldsymbol{\tau_{m_0}}$ are known. In particular, we derive the unrestricted estimator (UE) and the restricted estimator (RE) in the case where the change-points are known. Define $\boldsymbol{\tau} = (\boldsymbol{\tau_1}, \cdots, \boldsymbol{\tau_{m_0}})'$, let $\hat{\boldsymbol{\delta}}(\tau)$ denote the UE of $\boldsymbol{\delta}$ when τ is known. Similarly, let $\tilde{\boldsymbol{\delta}}(\tau)$ denote the RE of $\boldsymbol{\delta}$ when τ is known. Let $Q = \bar{Z} \otimes \bigotimes_{j=d}^{1} I_{q_j}$, $R = \bigotimes_{j=d+1}^{1} R_j$, let $\mathbb{J}_i = R_i'(R_i R_i')^{-1} R_i$, let $\mathbb{G}_i = R_i'(R_i R_i')^{-1}$ for $i = 1, \cdots, d$, let $\mathbb{J}_{d+1} = (\bar{Z}'\bar{Z})^{-1} R_{d+1}' (R_{d+1} (\bar{Z}'\bar{Z})^{-1} R_{d+1}')^{-1} R_{d+1}$, and let $\mathbb{G}_{d+1} = (\bar{Z}'\bar{Z})^{-1} R_{d+1}' (R_{d+1} (\bar{Z}'\bar{Z})^{-1} R_{d+1}')^{-1}$.

Proposition 3.1.1. *The UE and the RE are respectively given by*

$$\hat{\boldsymbol{\delta}}(\tau) = \mathbb{Y} \times_{d+1} (\bar{Z}'\bar{Z})^{-1} \bar{Z}', \quad \tilde{\boldsymbol{\delta}}(\tau) = \hat{\boldsymbol{\delta}}(\tau) - \hat{\boldsymbol{\delta}}(\tau) (\bigtimes_{i=1}^{d+1})_i \mathbb{J}_i + \boldsymbol{r} (\bigtimes_{j=1}^{d+1})_j \mathbb{G}_j.$$

The proof of Proposition 3.1.1 can be found in Appendix B. Note that for the special case where $d = 1$, the estimator in Proposition 3.1.1 yields the estimator given in Chen and Nkurunziza (2016).

3.1.2 Estimation in the case of unknown change-points

In this subsection, we outline the estimation method of $\boldsymbol{\delta}$ when the location of the change-points $\tau_1, \cdots, \tau_{m_0}$ are unknown. Thus, we also outline the estimation method for $\boldsymbol{\tau} =$

$(\tau_1, \cdots, \tau_{m_0})'$. In similar ways as in Chen and Nkurunziza (2016), one estimates the unknown parameter $\boldsymbol{\delta}$ and $\boldsymbol{\tau}$ by minimizing the least squares of objective function. This gives the UE of $\boldsymbol{\delta}$ and $\boldsymbol{\tau}$. Let $\hat{\boldsymbol{\tau}}$ and $\tilde{\boldsymbol{\tau}}$ denote the UE and RE of the true change-points from restricted least squares, respectively. Also, let $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}})$ and $\tilde{\boldsymbol{\delta}}(\tilde{\boldsymbol{\tau}})$ be the UE and RE for the regression coefficient tensor $\boldsymbol{\delta}$, respectively. Let $\mathrm{SSR}_T^U(\boldsymbol{\tau})$ and $\mathrm{SSR}_T^R(\boldsymbol{\tau})$ be the Frobenius-norm of residuals from the UE and RE least squares regression model evaluated at the partition $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_{m_0})$, respectively. We have

$$\hat{\boldsymbol{\tau}} = \arg\min_{\tau} \mathrm{SSR}_T^U(\boldsymbol{\tau}), \quad \text{and} \quad \tilde{\boldsymbol{\tau}} = \arg\min_{\tau} \mathrm{SSR}_T^R(\boldsymbol{\tau}). \tag{3.3}$$

The minimization of (3.3) needs to be done numerically by using the dynamic programming algorithm which is similar to the one used in Nkurunziza et al. (2019), Chen and Nkurunziza (2015) and references therein.

3.2 Asymptotic results

In this section, we derive some technical results underlying the proposed method. Specifically, we derive some fundamental results which are useful in generalizing the main results of Chen and Nkurunziza (2015) as well as that in Chen and Nkurunziza (2016). In the special case of matrix estimation problem, the established results are also useful in simplifying the proofs of the main results of the above quoted papers. In particular, we derive some asymptotic results for the UE and the RE.

3.2.1 Some fundamental results

Let $o_p(a)$ denote a random variate (r.v.) such that $o_p(a)/a$ converges in probability to 0, let $O_p(a)$ denote a r.v. such that $O_p(a)/a$ is bounded in probability. Similarly, let $o(a)$

denote a non-random quantity such that $o(a)/a$ converges to 0, and $O(a)$ denote a non-random quantity such that $O(a)/a$ is bounded. Further, the notations $\xrightarrow[T\to\infty]{\mathcal{L}^2}$, $\xrightarrow[T\to\infty]{a.s}$, $\xrightarrow[T\to\infty]{\mathrm{P}}$ and $\xrightarrow[T\to\infty]{d}$ stand for convergence in $\mathcal{L}^2$, convergence almost surely, convergence in probability and convergence in distribution, respectively. Let $f : \mathbb{E} \mapsto \mathbb{F}$ be a linear transformation where $\mathbb{E}$ and $\mathbb{F}$ are vector spaces. Define $\mathrm{Im}(f) = \{f(x) : x \in \mathbb{E}\}$ and $\mathrm{Ker}(f) = \{x :\in f(x) = 0\}$. Let $\left\{\mathscr{G}^i_{T,k}, \infty \leq k \leq i \leq \infty, T \geq 1\right\}$ be a filtration, let $\mathscr{H}_i$ be the Hilbert space of $\mathscr{G}^i_{T,-\infty}$−measurable and square integrable functions and let P_i be a projector onto $\mathscr{H}_i$. We suppose that the projector P_i satisfies the following assumption.

Assumption 3.2.1. *The sequence* $\{P_i\}_{i=-\infty}^{\infty}$ *is such that* $\|P_{i-j}(X)\| \leqslant c_i\varphi(j)$ *and* $\|X - P_{i+j}(X)\| \leqslant c_i\varphi(j+1)$ *where* $\{\varphi(j)\}_{j\geq 0}$ *is a decreasing function such that* $\varphi(j) = O\left(\frac{1}{j^{1/2}\boldsymbol{\kappa}(j)}\right)$, *with* $\boldsymbol{\kappa}(.)$ *an increasing and positive function such that* $\sum_{j=1}^{\infty} \frac{1}{j\boldsymbol{\kappa}(j)} < \infty$.

Under this assumption, we derive below several propositions and a lemma which play an important role in deriving the asymptotic normality of the UE.

Proposition 3.2.1. *If Assumption 3.2.1 holds, then,*

$$P_{i+m}(X) \xrightarrow[m\to\infty]{a.s \text{ and } \mathcal{L}^2} X \text{ and } P_{i-l-1}(X) \xrightarrow[l\to\infty]{a.s \text{ and } \mathcal{L}^2} 0.$$

Proof. Since $\|X - P_{i+m}(X)\|^2 \leq c_i^2\varphi^2(m+1)$, then we have

$$\sum_{m=0}^{\infty} \|X - P_{i+m}(X)\|^2 \leq c_i^2 \sum_{m=0}^{\infty} \varphi^2(m+1) = c_i^2 \sum_{m=1}^{\infty} \varphi^2(m).$$

Since $\varphi(m) = O\left(\frac{1}{m^{1/2}\boldsymbol{\kappa}(m)}\right)$, we get

$$\sum_{m=0}^{\infty} \|X - P_{i+m}(X)\|^2 \leq c_i^2 A \sum_{m=1}^{\infty} \frac{1}{m\boldsymbol{\kappa}^2(m)},$$

for some $A > 0$. Also, since $\boldsymbol{\kappa}$ is an increasing, positive function, we have $\boldsymbol{\kappa}^{-1}(1) \geq \boldsymbol{\kappa}^{-1}(m)$. Hence, we get

$$\sum_{m=0}^{\infty} \|X - P_{i+m}(X)\|^2 \leq c_i^2 A\boldsymbol{\kappa}^{-1}(1) \sum_{m=1}^{\infty} \frac{1}{m\boldsymbol{\kappa}(m)} < \infty.$$

As such, by Markov's inequality, for $\epsilon > 0$,

$$\sum_{m=0}^{\infty} \mathrm{P}\left(\|X - P_{i+m}(X)\|_{\mathrm{F}} \geq \epsilon\right) \leq \sum_{m=0}^{\infty} \mathrm{E}\left[\|X - P_{i+m}(X)\|_{\mathrm{F}}^2\right]/\epsilon^2 = \sum_{m=0}^{\infty} \|X - P_{i+m}(X)\|^2/\epsilon^2 < \infty.$$

Therefore, by the Borel-Cantelli Lemma, we have $X - P_{i+m}(X) \xrightarrow[m\to\infty]{a.s} 0$. This implies that $P_{i+m}(X) \xrightarrow[m\to\infty]{a.s} X$. Similarly, since $\sum_{l=0}^{\infty} \|P_{i-l-1}(X)\| \leq c_i^2 \sum_{l=1}^{\infty} \varphi^2(l) < \infty$, then $P_{i-l-1}(X) \xrightarrow[l\to\infty]{a.s} 0$. This completes the proof. □

Proposition 3.2.2. *If Assumption 3.2.1 holds, then* $\langle P_i(X), X\rangle = \|P_i(X)\|^2$.

Proof. By the definition of orthogonal projection, we have $X - P_i(X) \in \mathrm{Ker}(P_i)$. Since $\mathrm{Ker}(P_i) \subset (\mathrm{Im}(P_i))^{\perp}$, then $X - P_i(X) \in (\mathrm{Im}(P_i))^{\perp}$. Hence, $X - P_i(X)$ is orthogonal to any element of $\mathrm{Im}(P_i)$. In particular, $X - P_i(X)$ is orthogonal to $P_i(X)$. Hence, $\langle X - P_i(X), P_i(X)\rangle = \langle X, P_i(X)\rangle - \langle P_i(X), P_i(X)\rangle = 0$. This implies that $\langle X, P_i(X)\rangle = \|P_i(X)\|^2$. This completes the proof of the lemma. □

Proposition 3.2.3. *Suppose the conditions of Proposition 3.2.1 hold. Then,* $X = \sum_{k=-\infty}^{\infty} [P_{i+k}(X) - P_{i+k-1}(X)]$ *a.s and in* $\mathcal{L}^2$.

Proof. Note that $\sum_{k=-\infty}^{\infty} [P_{i+k}(X) - P_{i+k-1}(X)] = \lim_{l,m\to\infty} \sum_{k=-l}^{m} [P_{i+k}(X) - P_{i+k-1}(X)]$. Since this is a telescoping series, we have

$$\begin{aligned}\sum_{k=-\infty}^{\infty} [P_{i+k}(X) - P_{i+k-1}(X)] &= \lim_{l,m\to\infty} [P_{i+m}(X) - P_{i-l-1}(X)] \\ &= \lim_{m\to\infty} P_{i+m}(X) - \lim_{l\to\infty} P_{i-l-1}(X).\end{aligned}$$

By Proposition 3.2.1, we get

$$\sum_{k=-\infty}^{\infty} [P_{i+k}(X) - P_{i+k-1}(X)] = X - 0 = X, a.s \text{ and in } \mathcal{L}^2.$$

This completes the proof. □

Proposition 3.2.4. *Let $\bar{P}_i = \mathbf{1}_{\mathscr{H}_i} - P_i$, where $\mathbf{1}_{\mathscr{H}_i}$ denotes the projector identity onto $\mathscr{H}_i$. Suppose that the projectors P_i satisfy the conditions of Proposition 3.2.1 and suppose that $\langle P_i(X), P_j(X)\rangle = \|P_i(X)\|^2$, for all $i \leq j$. Then,*

1. $\|\bar{P}_{i+k-1}(X) - \bar{P}_{i+k}(X)\|^2 = \|P_{i+k}(X)\|^2 - \|P_{i+k-1}(X)\|^2$,

2. $\langle \bar{P}_{i+k-1}(X) - \bar{P}_{i+k}(X), \bar{P}_{j+k-1}(X) - \bar{P}_{j+k}(X)\rangle = 0$, *for all* $j \neq i$,

3. $\|\bar{P}_{i+k-1}(X)\|^2 - \|\bar{P}_{i+k}(X)\|^2 = \|P_{i+k}(X)\|^2 - \|P_{i+k-1}(X)\|^2$,

4. $\|X\|^2 = \sum_{k=-\infty}^{\infty} \|P_{i+k}(X) - P_{i+k-1}(X)\|^2$.

Proof. 1. We have

$$\|\bar{P}_{i+k-1}(X) - \bar{P}_{i+k}(X)\|^2 = \|P_{i+k}(X) - P_{i+k-1}(X)\|^2$$
$$= \langle P_{i+k}(X) - P_{i+k-1}(X), P_{i+k}(X) - P_{i+k-1}(X)\rangle$$
$$= \langle P_{i+k}(X), P_{i+k}(X)\rangle + \langle P_{i+k-1}(X), P_{i+k-1}(X)\rangle - 2\langle P_{i+k}(X), P_{i+k-1}(X)\rangle$$
$$= \langle P_{i+k}(X), P_{i+k}(X)\rangle + \langle P_{i+k-1}(X), P_{i+k-1}(X)\rangle - 2\langle P_{i+k-1}(X), P_{i+k-1}(X)\rangle$$
$$= \langle P_{i+k}(X), P_{i+k}(X)\rangle - \langle P_{i+k-1}(X), P_{i+k-1}(X)\rangle = \|P_{i+k}(X)\|^2 - \|P_{i+k-1}(X)\|^2.$$

This completes the proof of Part 1.

2. Without loss of generality, suppose that $j < i$.We have

$$\langle \bar{P}_{i+k-1}(X) - \bar{P}_{i+k}(X), \bar{P}_{j+k-1}(X) - \bar{P}_{j+k}(X)\rangle$$
$$= \langle P_{i+k}(X) - P_{i+k-1}(X), P_{j+k}(X) - P_{j+k-1}(X)\rangle$$
$$= \langle P_{i+k}(X), P_{j+k}(X)\rangle + \langle P_{i+k-1}(X), P_{j+k-1}(X)\rangle$$
$$- \langle P_{i+k}(X), P_{j+k-1}(X)\rangle - \langle P_{i+k-1}(X), P_{j+k}(X)\rangle$$
$$= \|P_{j+k}(X)\|^2 + \|P_{j+k-1}(X)\|^2 - \|P_{j+k-1}(X)\|^2 - \|P_{j+k}(X)\|^2 = 0.$$

This completes the proof of Part 2.

3. Note that

$$
\begin{aligned}
&||\bar{P}_{i+k-1}(X)||^2 - ||\bar{P}_{i+k}(X)||^2 \\
&= \langle X - P_{i+k-1}(X), X - P_{i+k-1}(X)\rangle - \langle X - P_{i+k}(X), X - P_{i+k}(X)\rangle \\
&= \langle X, X\rangle - 2\langle P_{i+k-1}(X), X\rangle + \langle P_{i+k-1}(X), P_{i+k-1}(X)\rangle \\
&- \langle X, X\rangle + 2\langle P_{i+k}(X), X\rangle - \langle P_{i+k}(X), P_{i+k}(X)\rangle.
\end{aligned}
$$

Then, by Proposition 3.2.2, we get

$$
\begin{aligned}
&||\bar{P}_{i+k-1}(X)||^2 - ||\bar{P}_{i+k}(X)||^2 \\
&= -2||P_{i+k-1}(X)||^2 + ||P_{i+k-1}(X)||^2 + 2||P_{i+k}(X)||^2 - ||P_{i+k}(X)||^2 \\
&= ||P_{i+k}(X)||^2 - ||P_{i+k-1}(X)||^2.
\end{aligned}
$$

This completes the proof of the third part.

4. Using Proposition 3.2.3, we have that

$$||X||^2 = \langle X, X\rangle = \langle X, \sum_{k=-\infty}^{\infty} [P_{i+k}(X) - P_{i+k-1}(X)]\rangle.$$

Hence, we have

$$
\begin{aligned}
||X||^2 = \langle X, X\rangle &= \langle X, \lim_{l,m\to\infty} \sum_{k=-l}^{m} [P_{i+k}(X) - P_{i+k-1}(X)]\rangle \\
&= \lim_{l,m\to\infty} \sum_{k=-l}^{m} \langle X, P_{i+k}(X) - P_{i+k-1}(X)\rangle.
\end{aligned}
$$

Then, $\langle X, P_{i+k}(X) - P_{i+k-1}(X)\rangle = \langle X, P_{i+k}(X)\rangle - \langle X, P_{i+k-1}(X)\rangle$.

By Proposition 3.2.2,

$\langle X, P_{i+k}(X)\rangle = ||P_{i+k}(X)||^2$ and $\langle X, P_{i+k-1}(X)\rangle = ||P_{i+k-1}(X)||^2$.

Hence,

$$\begin{aligned}\langle X, P_{i+k}(X) - P_{i+k-1}(X)\rangle &= \|P_{i+k}(X)\|^2 - \|P_{i+k-1}(X)\|^2\\ &= \langle P_{i+k}(X), P_{i+k}(X)\rangle - \langle P_{i+k-1}(X), P_{i+k-1}(X)\rangle\\ &= \langle P_{i+k}(X) - P_{i+k-1}(X), P_{i+k}(X) - P_{i+k-1}(X)\rangle\\ &= \|P_{i+k}(X) - P_{i+k-1}(X)\|^2.\end{aligned}$$

This completes the proof of the fourth part.

□

Below, we establish a lemma which is useful in deriving the asymptotic properties of the UE and the RE.

Lemma 3.2.1. *Let $\{a_k\}_{k=-\infty}^{\infty}$ and $\{b_k\}_{k=-\infty}^{\infty}$ be sequences of positive numbers such that $a_k = a_{-k}$ and suppose that $a_j^{-1} - a_{j-1}^{-1} = O(\boldsymbol{\kappa}(k))$, $b_j = O\left(\frac{1}{j\boldsymbol{\kappa}^2(j)}\right)$ where $\boldsymbol{\kappa}$ is as in Assumption 3.2.1. Then, $\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})b_k < \infty$.*

Proof. We have $\sum_{j=1}^{\infty}|a_j^{-1} - a_{j-1}^{-1}|b_j = \sum_{j=1}^{\infty} O(\boldsymbol{\kappa}(j))O\left(\frac{1}{j\boldsymbol{\kappa}^2(j)}\right) = \sum_{j=1}^{\infty} O\left(\frac{1}{j\boldsymbol{\kappa}(j)}\right)$. As such, we get for some $0 < A_0 < \infty$,

$$\sum_{j=1}^{\infty}(a_j^{-1} - a_{j-1}^{-1})b_j \leq A_0 \sum_{j=1}^{\infty} \frac{1}{j\boldsymbol{\kappa}(j)} < \infty.$$

This completes the proof. □

By using Lemma 3.2.1, we derive the following two propositions which play an important role in deriving the asymptotic properties of the UE.

Proposition 3.2.5. *Let $\{a_k\}_{k=-\infty}^{\infty}$ be as in Lemma 3.2.1. Under Assumption 3.2.1,* $\lim_{n\to\infty} a_{n+1}^{-1}\varphi^2(n+1) = 0$.

Proof. Since $a_j^{-1} - a_{j-1}^{-1} = O(\boldsymbol{\kappa}(j))$, we have

$$a_n^{-1} - a_0^{-1} = \sum_{j=1}^{n}\left(a_j^{-1} - a_{j-1}^{-1}\right) \leq B_0 \sum_{j=1}^{n} \boldsymbol{\kappa}(j),$$

for some $0 < B_0 < \infty$. Then, we have

$$a_n^{-1}\varphi^2(n) \le a_0^{-1}\varphi^2(n) + B_0\varphi^2(n)\sum_{j=1}^{n}\boldsymbol{\kappa}(k).$$

Since $\boldsymbol{\kappa}(j)$ is increasing, we get

$$a_n^{-1}\varphi^2(n) \le a_0^{-1}\varphi^2(n) + B_0\varphi^2(n)\sum_{j=1}^{n}\boldsymbol{\kappa}(n) = a_0^{-1}\varphi^2(n) + B_0\varphi^2(n)n\boldsymbol{\kappa}(n).$$

Since $\varphi^2(n) = O(\frac{1}{n\boldsymbol{\kappa}^2(n)})$, we have

$$a_n^{-1}\varphi^2(n) \le a_0^{-1}\varphi^2(n) + B_1\frac{1}{n\boldsymbol{\kappa}^2(n)}\boldsymbol{\kappa}(n) = a_0^{-1}\varphi^2(n) + B_1\frac{1}{\boldsymbol{\kappa}(n)},$$

for some $0 < B_1 < \infty$. Hence, since

$$a_{n+1}^{-1}\varphi^2(n+1) \le a_0^{-1}\varphi^2(n+1) + B_1\frac{1}{\boldsymbol{\kappa}(n+1)},$$

then

$$0 \le \lim_{n\to\infty} a_{n+1}^{-1}\varphi^2(n+1) \le a_0^{-1}\lim_{n\to\infty}\varphi^2(n+1) + B_1\lim_{n\to\infty}\frac{1}{\boldsymbol{\kappa}(n+1)} = 0.$$

This completes the proof. □

Proposition 3.2.6. *Let $\{a_k\}_{k=-\infty}^{\infty}$ be as in Lemma 3.2.1. Under Assumption 3.2.1,*

1. $\sum_{k=1}^{\infty}\sum_{i=1}^{L_p}\left(a_k^{-1} - a_{k-1}^{-1}\right)\|P_{i-k}(X)\|^2 < \infty,$
2. $\sum_{k=1}^{\infty}\sum_{i=1}^{L_p}\left(a_{k+1}^{-1} - a_k^{-1}\right)\|\bar{P}_{i+k}(X)\|^2 < \infty,$
3. $\sum_{k=1}^{\infty}\sum_{i=1}^{L_p}\left[a_k^{-1}\|\bar{P}_{i+k-1}(X)\|^2 - a_{k+1}^{-1}\|\bar{P}_{i+k}(X)\|^2\right] = \sum_{i=1}^{L_p} a_1^{-1}\|\bar{P}_i(X)\|^2 < \infty,$
4. $\sum_{k=1}^{\infty}\sum_{i=1}^{L_p}\left[a_{k-1}^{-1}\|P_{i-k}(X)\|^2 - a_k^{-1}\|P_{i-k-1}(X)\|^2\right] = \sum_{i=1}^{L_p} a_0^{-1}\|P_{i-1}(X)\|^2 < \infty.$

Proof. 1. We have

$$\begin{aligned}\sum_{k=1}^{\infty}\sum_{i=1}^{L_p}\left(a_k^{-1} - a_{k-1}^{-1}\right)\|P_{i-k}(X)\|^2 &\le \sum_{k=1}^{\infty}\sum_{i=1}^{L_p}\left(a_k^{-1} - a_{k-1}^{-1}\right)c_{p,i}^2\varphi^2(k)\\ &= \sum_{i=1}^{L_p} c_{p,i}^2\sum_{k=1}^{\infty}\left(a_k^{-1} - a_{k-1}^{-1}\right)\varphi^2(k).\end{aligned}$$

Thus, by taking $b_k = \varphi^2(k)$ and by applying Lemma 3.2.1,

$$\sum_{k=1}^{\infty} \left(a_k^{-1} - a_{k-1}^{-1}\right) \varphi^2(k) < \infty.$$

This completes the proof of Part 1.

2. We have

$$\begin{aligned}\sum_{k=1}^{\infty} \sum_{i=1}^{L_p} \left(a_{k+1}^{-1} - a_k^{-1}\right) \|\bar{P}_{i+k}(X)\|^2 &\leq \sum_{k=1}^{\infty} \sum_{i=1}^{L_p} \left(a_{k+1}^{-1} - a_k^{-1}\right) c_{p,i}^2 \varphi^2(k+1) \\ &= \sum_{i=1}^{L_p} c_{p,i}^2 \sum_{k=1}^{\infty} \left(a_{k+1}^{-1} - a_k^{-1}\right) \varphi^2(k+1).\end{aligned}$$

Hence, we have by Lemma 3.2.1,

$$\sum_{k=1}^{\infty} \sum_{i=1}^{L_p} \left(a_{k+1}^{-1} - a_k^{-1}\right) \|\bar{P}_{i+k}(X)\|^2 \leq \sum_{i=1}^{L_p} c_{p,i}^2 \sum_{k=2}^{\infty} \left(a_k^{-1} - a_{k-1}^{-1}\right) \varphi^2(k) < \infty.$$

This completes the proof of the second part.

3. Since the sum is a telescoping series,

$$\sum_{k=1}^{\infty} \sum_{i=1}^{L_p} \left[a_k^{-1} \|\bar{P}_{i+k-1}(X)\|^2 - a_{k+1}^{-1} \|\bar{P}_{i+k}(X)\|^2\right] = \sum_{i=1}^{L_p} a_1^{-1} \|\bar{P}_i(X)\|^2 - \lim_{n\to\infty} a_{n+1}^{-1} \sum_{i=1}^{L_p} \|\bar{P}_{i+n}(X)\|^2.$$

Then,

$$\lim_{n\to\infty} a_{n+1}^{-1} \sum_{i=1}^{L_p} \|\bar{P}_{i+n}(X)\|^2 \leq \lim_{n\to\infty} a_{n+1}^{-1} \sum_{i=1}^{L_p} c_{p,i}^2 \varphi^2(n+1) = \sum_{i=1}^{L_p} c_{p,i}^2 \lim_{n\to\infty} a_{n+1}^{-1} \varphi^2(n+1).$$

By Proposition 3.2.5, $\lim_{n\to\infty} a_{n+1}^{-1} \varphi^2(n+1) = 0$. This gives

$$\lim_{n\to\infty} a_{n+1}^{-1} \sum_{i=1}^{L_p} \|\bar{P}_{i+n}(X)\|^2 = 0.$$

This completes the proof of the third part.

4. Note that

$$\begin{aligned}&\sum_{k=1}^{\infty} \sum_{i=1}^{L_p} \left[a_{k-1}^{-1} \|P_{i-k}(X)\|^2 - a_k^{-1} \|P_{i-k-1}(X)\|^2\right] \\ &= \sum_{i=1}^{L_p} a_0^{-1} \|P_{i-1}(X)\|^2 - \lim_{n\to\infty} \sum_{i=1}^{L_p} a_n^{-1} \|P_{i-n-1}(X)\|^2.\end{aligned}$$

By Proposition 3.2.5, we get $\lim_{n\to\infty} \sum_{i=1}^{L_p} a_n^{-1}||P_{i-n-1}(X)||^2 = 0$. Therefore,

$$\sum_{k=1}^{\infty} \sum_{i=1}^{L_p} \left[a_{k-1}^{-1}||P_{i-k}(X)||^2 - a_k^{-1}||P_{i-k-1}(X)||^2\right] = \sum_{i=1}^{L_p} a_0^{-1}||P_{i-1}(X)||^2 < \infty.$$

This completes the proof.

□

Using Proposition 3.2.5 and Proposition 3.2.6, we derive the following corollary.

Corollary 3.2.1. *Suppose that the conditions of Proposition 3.2.6 hold and let*

$$U_{j,k}(l) = \sum_{i=l+1}^{l+j} \left[P_{i+k}(X) - P_{i+k-1}(X)\right],$$

$j = 1, \cdots, L_p, k = 1, 2, \cdots$. *Then, for* $s_1 = 1, \cdots, q_1, s_2 = 1, \cdots, q_2, \cdots, s_d = 1, \cdots, q_d, s_{d+1} = 1, \cdots, q_{d+1}$,

$$\sum_{k=1}^{\infty} a_k^{-1}||U_{L_p,k}^2(l)||^2 = \sum_{i=l+1}^{l+L_p} \left(\sum_{k=1}^{\infty}(a_{k+1}^{-1} - a_k^{-1})||\bar{P}_{i+k}(X)||^2 + a_1^{-1}||\bar{P}_i(X)||^2\right),$$

and

$$\sum_{k=-\infty}^{-1} a_k^{-1}||U_{L_p,k}^2(l)||^2 = \sum_{i=l+1}^{l+L_p} \left(\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})||P_{i-k}(X)||^2 + a_0^{-1}||P_{i-1}(X)||^2\right).$$

Proof. Note that

$$||U_{L_p,k}(l)||^2 = \left\|\left(\sum_{i=l+1}^{l+L_p}(\bar{P}_{i+k-1}(X) - \bar{P}_{i+k}(X))\right)\right\|^2$$

$$= \sum_{i=l+1}^{l+L_p} ||\bar{P}_{i+k-1}(X) - \bar{P}_{i+k}(X)||^2 + 2\sum_{i=l+2}^{l+L_p} \sum_{j=l+1}^{i-1} \langle \bar{P}_{i+k-1}(X) - \bar{P}_{i+k}(X), \bar{P}_{j+k-1}(X) - \bar{P}_{j+k}(X)\rangle.$$

Hence, it follows from Proposition 3.2.4 that

$$||U_{L_p,k}(l)||^2 = \sum_{i=l+1}^{l+L_p} \left[||P_{i+k}(X)||^2 - ||P_{i+k-1}(X)||^2\right] = \sum_{i=l+1}^{l+L_p} \left[||\bar{P}_{i+k-1}(X)||^2 - ||\bar{P}_{i+k}(X)||^2\right].$$

Using this, we have

$$a_0^{-1}\|U_{L_p,0}(l)\|^2 = \sum_{i=l+1}^{l+L_p} a_0^{-1}\left[\|P_i(X)\|^2 - \|P_{i-1}(X)\|^2\right]$$

and

$$\sum_{k=1}^{\infty} a_k^{-1}\|U_{L_p,k}(l)\|^2 = \sum_{k=1}^{\infty} a_k^{-1}\left(\sum_{i=l+1}^{l+L_p}\left[\|\bar{P}_{i+k-1}(X)\|^2 - \|\bar{P}_{i+k}(X)\|^2\right]\right).$$

Moreover, we have

$$\begin{aligned}\sum_{k=1}^{\infty} a_k^{-1}\|U_{L_p,k}(l)\|^2 &= \sum_{k=1}^{\infty}\left(\sum_{i=l+1}^{l+L_p}(a_{k+1}^{-1} - a_k^{-1})\|\bar{P}_{i+k}(X)\|^2\right)\\ &+ \sum_{k=1}^{\infty}\sum_{i=l+1}^{l+L_p}\left(a_k^{-1}\|\bar{P}_{i+k-1}(X)\|^2 - a_{k+1}^{-1}\|\bar{P}_{i+k}(X)\|^2\right).\end{aligned}$$

Hence, using the third statement in Proposition 3.2.6, we get

$$\sum_{k=1}^{\infty} a_k^{-1}\|U_{L_p,k}(l)\|^2 = \sum_{k=1}^{\infty}\left(\sum_{i=l+1}^{l+L_p}(a_{k+1}^{-1} - a_k^{-1})\|\bar{P}_{i+k}(X)\|^2\right) + \sum_{i=l+1}^{l+L_p} a_1^{-1}\|\bar{P}_i(X)\|^2 < \infty,$$

this proves the first statement. We prove the second statement by following similar steps and using the assumption that $a_k = a_{-k}$. Namely, we have

$$\begin{aligned}\sum_{k=-\infty}^{-1} a_k^{-1}\|U_{L_p,k}(l)\|^2 &= \sum_{k=-\infty}^{-1} a_k^{-1}\sum_{i=l+1}^{l+L_p}\left[\|P_{i+k}(X)\|^2 - \|P_{i+k-1}(X)\|^2\right]\\ &= \sum_{k=1}^{\infty}\sum_{i=l+1}^{l+L_p} a_k^{-1}\left[\|P_{i-k}(X)\|^2 - \|P_{i-k-1}(X)\|^2\right].\end{aligned}$$

Now, using the first and fourth statements in Proposition 3.2.6, we have

$$\begin{aligned}\sum_{k=-\infty}^{-1} a_k^{-1}\|U_{L_p,k}(l)\|^2 &= \sum_{k=1}^{\infty}\sum_{i=l+1}^{l+L_p}(a_k^{-1} - a_{k-1}^{-1})\|P_{i-k}(X)\|^2\\ &+ \sum_{k=1}^{\infty}\sum_{i=l+1}^{l+L_p}\left(a_{k-1}^{-1}\|P_{i-k}(X)\|^2 - a_k^{-1}\|P_{i-k-1}(X)\|^2\right) < \infty.\end{aligned}$$

Hence, by the fourth statement of Proposition 3.2.6,

$$\sum_{k=-\infty}^{-1} a_k^{-1}\|U_{L_p,k}(l)\|^2 = \sum_{i=l+1}^{l+L_p}\left(\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\|P_{i-k}\|^2 + a_0^{-1}\|P_{i-1}\|^2\right),$$

which completes the proof. □

To establish some main results, we consider the case where $\boldsymbol{k}(\cdot)$ satisfies the following assumption.

Assumption 3.2.2. *The function $\boldsymbol{\kappa}(.)$ is positive and increasing such that , for some $M > 0$, $\boldsymbol{\kappa}(x) \geqslant 1$ for all $x \geqslant M$ and* $\sum_{n=1}^{\infty}\left(\sum_{j=0}^{n}\boldsymbol{\kappa}(j)\right)^{-1/2} < \infty$.

Lemma 3.2.2. *Under Assumption 3.2.2, let $b_j = O\left(j^{-1}\boldsymbol{\kappa}^{-2}(j)\right)$ and let $a_0 = \boldsymbol{\kappa}(0)$, $a_{-j} = a_j$, $a_j = \left(-a_{j-1}^{-1} + \sqrt{a_{j-1}^{-2} + 4\boldsymbol{\kappa}(j)}\right)/(2\boldsymbol{\kappa}(j))$, $j = 1, 2, \ldots$ Then,*

1. $\sum_{j=-\infty}^{\infty} a_j < \infty$; **(2).** $\sum_{n=1}^{\infty}\left(\sum_{j=1}^{n} j\boldsymbol{\kappa}^2(j)\right)^{-1/2} < \infty$; 3. $\sum_{n=1}^{\infty}(a_n^{-1} - a_{n-1}^{-1})b_n < \infty$;

4. $\sum_{n=1}^{\infty} n^{-1}\boldsymbol{\kappa}^{-1}(n) < \infty$.

Proof. 1. Obviously, $a_j > 0$, for any integer j. Further, one can verify that $a_j^{-1} = \frac{\left(a_{j-1}^{-1} + \sqrt{a_{j-1}^{-2} + 4\boldsymbol{\kappa}(j)}\right)}{2}$, then, $a_j^{-1} - a_{j-1}^{-1} = a_j\boldsymbol{\kappa}(j)$ and then, $a_j^{-1} - a_{j-1}^{-1} > 0$, for all $j = 1, 2, \ldots$ This proves that a_j is an increasing sequence. We also have $\left(a_j^{-2} - a_{j-1}^{-2}\right) > \boldsymbol{\kappa}(j)$, $\quad j = 1, 2, \ldots$ This gives $a_n^{-2} > \sum_{j=0}^{n}\boldsymbol{\kappa}(j)$, $n = 1, 2, \ldots$ Hence, $a_n < \left(\sum_{j=0}^{n}\boldsymbol{\kappa}(j)\right)^{-1/2}$, $n = 1, 2, 3, \ldots$. Therefore, $\sum_{n=-\infty}^{\infty} a_j < 2\sum_{n=1}^{\infty}\left(\sum_{j=0}^{n}\boldsymbol{\kappa}(j)\right)^{-1/2} < \infty$, this proves Part 1.

2. Part 2 follows directly from Part 1.

3. We have, $\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})b_k \leqslant A_0\sum_{k=1}^{\infty}\frac{a_k}{k\boldsymbol{\kappa}(k)} \leqslant \frac{1}{\boldsymbol{\kappa}(1)}\sum_{k=1}^{\infty} a_k < \infty$.

4. Since $n\boldsymbol{\kappa}^2(n)$ is increasing, we have $\sum_{j=1}^{n} j\boldsymbol{\kappa}^2(j) \leqslant n^2\boldsymbol{\kappa}^2(n)$, $n = 1, 2, \ldots$ Then, $\left(\sum_{j=1}^{n} j\boldsymbol{\kappa}^2(j)\right)^{-1/2} \geqslant n^{-1}\boldsymbol{\kappa}^{-1}(n)$, $n = 1, 2, \ldots$ Then, the proof follows directly from Part **(2)**. This completes the proof. □

Remark 3.2.1. *Part 4 of Lemma 3.2.2 shows that, under Assumption 3.2.2, $\boldsymbol{\kappa}(.)$ satisfies Assumption 3.2.1.*

3.2.2 About the structure of the noise and the regressors

In this subsection, we derive some technical results underlying the structure of the error terms and the regressors. The established results are useful in deriving the joint asymptotic normality between the UE and the RE. In the following, we present some conditions for deriving the joint asymptotic normality of the UE and the RE, which is an important step for the proposed method. Indeed, the optimality of the proposed method is based on the asymptotic results. We first define the concept of $\mathcal{L}^p - mixingales$ in the following definition.

Definition 3.2.1. *Let* $\{\boldsymbol{U}_{nt}\}$ *be a zero-mean stochastic array and* $\{\mathscr{G}^t_{n,s}, -\infty \leq s \leq t \leq \infty,\ n \geq 1\}$ *be an array of sigma subfields of* $\mathscr{F}$. *Then,* $\{U_{nt}, \mathscr{G}^t_{n,s}\}$ *is called an* $\mathcal{L}^r$*– mixingale of size* $-\lambda_0$ *if*

$$||\mathrm{E}(\boldsymbol{U}_{nt}|\mathscr{G}^{t-m}_{n,-\infty})||_r \leqslant a_{nt}\zeta_{r,m}, \quad \text{and} \quad ||U_{nt} - \mathrm{E}(\boldsymbol{U}_{nt}|\mathscr{G}^{t+m}_{n,-\infty})||_r \leqslant a_{nt}\zeta_{r,m+1}, \tag{3.4}$$

where $\{a_{nt}\}$ *is an array of positive constants and* $\zeta_{r,m} = O(m^{-\lambda})$ *for* $\lambda > \lambda_0 > 0$.

Assumption 3.2.3. ($\mathscr{C}_1$) *Let* $L_p = (\tau^0_p - \tau^0_{p-1}), p = 1, \cdots, m$, *then* $(1/L_p)\sum_{t=\tau^0_{p-1}+1}^{\tau^0_{p-1}+[L_p/v]} z_t z_t' \xrightarrow[T\to\infty]{P} Q_p(v)$, *a non-random positive definite matrix uniformly in* $v \in [0, 1]$ *and there exists an* $L_0 > 0$, *such that for all* $L_p > L_0$, *the minimum eigenvalues of* $(1/L_p)\sum_{t=\tau^0_{p-1}+1}^{\tau^0_{p-1}+L_p} z_t z_t'$ *and* $(1/L_p)\sum_{t=\tau^0_{p-1}+1}^{\tau^0_{p-1}} z_t z_t'$ *are bounded away from 0.*

($\mathscr{C}_2$) *The matrix* $\sum_{t=i_1}^{i_2} z_t z_t'$ *is invertible for* $0 \leq i_2 - i_1 \leq \epsilon_0 T$ *for some* $\epsilon_0 > 0$.

($\mathscr{C}_3$) $\tau^0_p = \left[T\lambda^0_p\right]$, *where* $p = 1, \cdots, m_0 + 1$ *and* $0 < \lambda^0_1 < \cdots < \lambda^0_{m_0} < \lambda^0_{m_0+1} = 1$.

($\mathscr{C}_4$) *The minimization problem defined by* (3.3) *is taken over all possible partitions, such that* $\tau_i - \tau_{i-1} > vT, (i = 1, \cdots, m_0 + 1)$ *for some* $v > 0$.

($\mathscr{C}_5$) *For each segment,* $\left(\tau^0_{p-1}, \tau^0_p\right), p = 1, \cdots, m_0 + 1$, *set*

$$\mathbb{X}_{p,i} = T^{-1/2}\mathbb{U}_{\tau^0_{p-1}+i} \circ z_{\tau^0_{p-1}+i} \in \mathbb{R}^{q_1\times\cdots\times q_d\times q_{d+1}},$$

where $\mathscr{F}^{i}_{p,-\infty} = \mathscr{F}^{\tau^0_{p-1}+i}_{T,-\infty}$. *We assume that there exist sequences of non-negative real numbers* $\{c_{pi} : i \geq 1\}$ *and* $\{\psi(j), j \geq 0\}$, *such that* $\psi(j) = O(j^{-1/2}\boldsymbol{\kappa}^{-1}(j))$, *with the function* $\boldsymbol{\kappa}(\cdot)$ *as defined in Assumption 3.2.2, and*

$$\|\mathrm{E}(\mathbb{X}_{pi}|\mathscr{F}^{i-j}_{p,-\infty})\|_2 \leq c_{pi}\psi(j), \quad \|\mathbb{X}_{pi} - \mathrm{E}(\mathbb{X}_{pi}|\mathscr{F}^{i+j}_{p,-\infty})\|_2 \leq c_{pi}\psi(j+1). \tag{3.5}$$

Further, within each break, set $L_p = \tau^0_p - \tau^0_{p-1}$ *and let* l_p, b_p *be such that* $1 \leq l_p \leq l_p + 1 \leq b_p \leq L_p$, *and set* $r_p = \left[L_p/b_p\right]$. *Thus, this separates interval* $\left[\tau^0_{p-1}, \tau^0_p\right]$ *into* $\left[L_p/b_p\right]$ *or* $\left[L_p/b_p\right] + 1$ *cells.*

We assume that $b_p \xrightarrow[L_p\to\infty]{} \infty, l_p \xrightarrow[L_p\to\infty]{} \infty, b_p/L_p \xrightarrow[L_p\to\infty]{} 0$, *and* $l_p/b_p \xrightarrow[L_p\to\infty]{} 0$.

$(\mathscr{C}_6)$ *For* $p = 1, \cdots, m_0 + 1$, *set* $\mathbb{V}_{pi} = \sum_{t=(i-1)b_p+l_p+1}^{ib_p} \mathbb{X}_{p,t}$, *and set*

$$\mathbb{V}_i = \mathbb{V}_{1,i} \boxplus_{(d+1)} \mathbb{V}_{2,i} \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \mathbb{V}_{m_0+1,i} \in \mathbb{R}^{q_1\times\cdots q_d\times(m_0+1)q_{d+1}},$$

1) For $s_1 = 1, \cdots, q_1; s_2 = 1, \cdots, r_2; s_3 = 1, \cdots, r_3; \cdots; s_d = 1, \cdots, q_d; s_{d+1} = 1, \cdots, q_{d+1}$, $\{X^2_{pi,s_1,\cdots,s_{d+1}}/c^2_{pi}, i = 1, 2, \cdots\}$ *is uniformly integrable.*

2) $\max_{1\leq i\leq L_p} c_{pi} = o(T^{-\alpha/2}b_p^{-1/2})$, *for some* $0 < \alpha \leq 1$.

3) $\sum_{i=1}^{r_p}\left(\max_{(i-1)b_p+1\leq t\leq ib_p} c_{pt}\right)^2 = O(T^{-\alpha}b_p^{-1})$.

4) $\sum_{i=1}^{r_p} \mathbb{V}_{p,i(n)}\mathbb{V}'_{p,i(n)} \xrightarrow[L_p\to\infty]{P} \Sigma_{p,n}$, *for* $n = 1, \cdots, d+1$. *Let* $r_{min} = \min_{1\leq i\leq m_0+1} r_i$ *and let* $l_{min} = \min_{1\leq i\leq m_0+1} L_i$, *we have* $\sum_{i=1}^{r_{min}} \mathrm{Vec}(\mathbb{V}_{p,i})\mathrm{Vec}(\mathbb{V}_{p,i})' \xrightarrow[L_p\to\infty]{P} \bigotimes_{j=d+1}^{1} \Sigma_{p,j}$, *and*

(i) $\sum_{i=r_{min}+1}^{r_j}\left(\max_{(i-1)b_j+1\leq t\leq ib_j} c_{j,t}\right)^2 = O(T^{-\alpha}b_j^{-1})$

(ii) $\sum_{i=1}^{r_{min}} \mathbb{V}_{i(j)}\mathbb{V}'_{i(j)} \xrightarrow[L_p\to\infty]{P} \Lambda_j$, *for* $n = 1, \cdots, d+1$ *and* $\sum_{i=1}^{r_{min}} \mathrm{Vec}(\mathbb{V}_i)\mathrm{Vec}(\mathbb{V}_i)' \xrightarrow[L_{min}\to\infty]{P} \Lambda^* = \bigotimes_{j=d+1}^{1} \Lambda_j$, *where* Λ_j *is* $q_j \times q_j$ *positive definite matrix for* $j = 1, \cdots, d$ *and* Λ_{d+1} *is an* $(m_0+1)q_{d+1} \times (m_0+1)q_{d+1}$ *positive definite matrix.*

Remark 3.2.2. *For the special case where $d = 1$, Assumption ($\mathscr{C}_6$) becomes Condition ($\mathscr{C}_6$) of Chen and Nkurunziza (2016). Indeed, if $d = 1$, we have*

$$\sum_{i=1}^{r_{min}} \mathbb{V}_{i(1)}\mathbb{V}'_{i(1)} = \sum_{i=1}^{r_{min}} \mathbb{V}_i\mathbb{V}'_i \xrightarrow[L_p\to\infty]{P} \Lambda_1, \sum_{i=1}^{r_{min}} \mathbb{V}_{i(2)}\mathbb{V}'_{i(2)} = \sum_{i=1}^{r_{min}} \mathbb{V}'_i\mathbb{V}_i = \xrightarrow[L_p\to\infty]{P} \Lambda_2$$

and

$$\sum_{i=1}^{r_{min}} \mathrm{Vec}(\mathbb{V}_i)\mathrm{Vec}(\mathbb{V}_i)' \xrightarrow[L_{min}\to\infty]{P} \Lambda_2 \otimes \Lambda_1$$

where Λ_1 is $q_1 \times q_1$ positive definite matrix for $j = 1, \cdots, d$ and Λ_2 is an $(m_0+1)q_2 \times (m_0+1)q_2$ positive definite matrix.

Remark 3.2.3. *For the special case where $\kappa(j) = j^{\epsilon}$, for some $\epsilon > 0$, the condition in* (3.5) *means that $\{\mathbb{X}_{pi}, \mathscr{F}^{i}_{p,-\infty}\}$ forms an $\mathcal{L}^2$−mixingale array of size $-1/2$.*

The role of the first statement in Condition ($\mathscr{C}_1$) is to overcome the problem of unit root regressors while the role of the second statement, in Condition ($\mathscr{C}_1$), is to avoid local collinearity problem which guarantees the identifiability of the change-points. The role of Condition ($\mathscr{C}_2$) is to guarantee the existence of the tensor estimate. In the model without change-points, this corresponds to the classical requirement of full rank matrix of regressors. Condition ($\mathscr{C}_3$) implies that the length of each regime is proportional to T and the role of this condition is to guarantee that the location of the change-points is asymptotically distinct. Condition ($\mathscr{C}_4$) guarantees that the change-points are distinguishable. Thus, provided that ν is relatively large, the search for the change-points avoids the change-point candidates that are too close. Nevertheless, since ν can be chosen very small, Condition ($\mathscr{C}_4$) does not constitute a limitation for the proposed method. Condition ($\mathscr{C}_5$) defines the dependence structure of the error and the regressors. This condition is so general that it holds for classical (univariate or multivariate) regression models. Finally, Condition ($\mathscr{C}_6$) is useful in deriving tensor type Central limit theorem for tensor $\mathcal{L}^2$−mixingale array of size $-1/2$.

Overall, the conditions in Assumption 3.2.3 are very general and hold for a vast array of models. Specifically, these conditions guarantee that the proposed method can be applied to the classical models where the errors are assumed to be independent and identically distributed as well as to the case where the errors are neither independent nor identically distributed. Further, we do not assume a specific distribution for the tensor noise $\mathbb{U}$ and it is not necessary for $\mathbb{U}$ to be independent of the matrix of covariates $\bar{\boldsymbol{Z}}$. Namely, Condition ($\mathscr{C}_5$) shows that the dependence structure of the noise and the regressors is much weaker than what is seen in the literature. In particular, the above conditions admit a vast array of possible applications, including many auto-correlated and heteroscedastic models.

Using the results of Section 3.2.1, we establish the following series of results that will be used to derive Lemma 3.2.6, the main result of this subsection. Lemma 3.2.6 is useful in establishing some asymptotic results about the UE and RE.

Corollary 3.2.2. *Suppose that ($\mathscr{C}_5$)-($\mathscr{C}_6$) hold, then, for each $i, p = 1, 2, \cdots$*

$$\mathrm{E}[\mathbb{X}_{p,i}|\mathscr{F}^{i+m}_{p,-\infty}] \xrightarrow[m\to\infty]{a.s} \mathbb{X}_{p,i} \quad \text{and} \quad \mathrm{E}[\mathbb{X}_{p,i}|\mathscr{F}^{i-l-1}_{p,-\infty}] \xrightarrow[l\to\infty]{a.s} 0.$$

Proof. The result follows directly from Proposition 3.2.1 by taking the projector $P_i(X) = \mathrm{E}\left[\mathbb{X}_{p,i}|\mathscr{F}^{i}_{p,-\infty}\right]$. □

We now use the above proposition to establish the following corollary. Set $\mathbb{D}_{i,k} = \mathbb{X}_{p,i} - \mathrm{E}[\mathbb{X}_{p,i}|\mathscr{F}^{i+k}_{p,-\infty}]$ and set $D_{i,k,s_1,\cdots,s_{d+1}}$ be the element in s_1^{th} row, s_2^{th} column, s_3^{th} position in the third dimension,..., s_{d+1}^{th} position in the $(d+1)^{th}$ dimension.

Corollary 3.2.3. *Suppose that conditions ($\mathscr{C}_5$)-($\mathscr{C}_6$) of Assumption 3.2.3. Then, for* $s_1 = 1, \cdots q_1, s_2 = 1, \cdots, r_2, \cdots, s_d = 1, \cdots, q_d, s_{d+1} = 1, \cdots, q_{d+1}$, *we have*

1.

$$\mathrm{E}\left(\sum_{i=1}^{L_p}(D_{i,k-1,s_1,s_2,\cdots,s_{d+1}} - D_{i,k,s_1,s_2,\cdots,s_{d+1}})^2\right) = \sum_{i=1}^{L_p}\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}_{p,-\infty}^{i+k})) - \sum_{i=1}^{L_p}\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}_{p,-\infty}^{i+k-1}));$$

2.

$$\sum_{i=1}^{L_p}\sum_{j=1}^{i-1}\mathrm{E}\left[(D_{i,k-1,s_1,s_2,\cdots,s_{d+1}} - D_{i,k,s_1,s_2,\cdots,s_{d+1}})(D_{j,k-1,s_1,s_2,\cdots,s_{d+1}} - D_{j,k,s_1,s_2,\cdots,s_{d+1}})\right] = 0;$$

3.

$$\sum_{i=1}^{L_p}\left[\mathrm{E}(D^2_{i,k-1,s_1,\cdots,s_{d+1}}) - \mathrm{E}(D^2_{i,k,s_1,\cdots,s_{d+1}})\right] = \sum_{i=1}^{L_p}\left[\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}_{p,-\infty}^{i+k})) - \mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}_{p,-\infty}^{i+k-1}))\right].$$

Proof. Proof follows directly from Proposition 3.2.4. □

Using Corollary 3.2.3, we establish the following result.

Corollary 3.2.4. *Suppose that assumptions ($\mathscr{C}_5$) and ($\mathscr{C}_6$) hold and let $\{a_k\}_{-\infty}^{\infty}$ be as in Lemma 3.2.1, Then, the following statements hold for $s_1 = 1, \cdots, q_1, \cdots, s_d = 1, \cdots, q_d$, $s_{d+1} = 1, \cdots, q_{d+1}$,*

1) $\sum_{k=1}^{\infty}\sum_{i=1}^{L_p}(a_k^{-1} - a_{k-1}^{-1})\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}_{p,-\infty}^{i-k})) < \infty.$

2) $\sum_{k=1}^{\infty}\sum_{i=1}^{L_p}(a_{k+1}^{-1} - a_k^{-1})\mathrm{E}(D^2_{i,k,s_1,\cdots,s_{d+1}}) < \infty.$

3) $\sum_{k=1}^{\infty}\sum_{i=1}^{L_p}[a_k^{-1}\mathrm{E}(D^2_{i,k-1,s_1,\cdots,s_{d+1}}) - a_{k+1}^{-1}\mathrm{E}(D^2_{i,k,s_1,\cdots,s_{d+1}})] = \sum_{i=1}^{L_p}a_1^{-1}\mathrm{E}(D^2_{i,0,s_1,\cdots,s_{d+1}}) < \infty.$

4) $\sum_{k=1}^{\infty}\sum_{i=1}^{L_p}[a_{k-1}^{-1}\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}_{p,-\infty}^{i-k})) - a_k^{-1}\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}_{p,-\infty}^{i-k-1}))]$
$= \sum_{i=1}^{L_p}a_0^{-1}\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}_{p,-\infty}^{i-1})) < \infty.$

Proof. This result follows directly from Proposition 3.2.6 with $P_i(X) = \mathrm{E}\left(\mathbb{X}_{p,i}|\mathscr{F}^{i}_{p,-\infty}\right)$. □

By using Corollary 3.2.1, we derive the following result which is useful in establishing a central limit theorem type for some tensor mixingales.

Corollary 3.2.5. *Suppose that the conditions of Proposition C.2.2 hold and let*

$$\mathbb{V}_{j,k}(l) = \sum_{i=l+1}^{l+j}\left[\mathrm{E}(\mathbb{X}_{p,i}|\mathscr{F}^{i+k}_{p,-\infty}) - \mathrm{E}(\mathbb{X}_{p,i}|\mathscr{F}^{i+k-1}_{p,-\infty})\right],$$

$j = 1,\cdots,L_p, k = 1,2,\cdots, l = 0,1,\cdots$ *. Then, for* $s_h = 1,\cdots,q_h, h = 1,2,\cdots,d+1,$

$$\sum_{k=1}^{\infty} a_k^{-1}\mathrm{E}(V^2_{L_p,k,s_1,\cdots,s_{d+1}}(l)) = \sum_{i=l+1}^{l+L_p}\left(\sum_{k=1}^{\infty}(a_{k+1}^{-1} - a_k^{-1})\mathrm{E}(D^2_{i,k,s_1,\cdots,s_{d+1}}) + a_1^{-1}\mathrm{E}(D^2_{i,0,s_1,\cdots,s_{d+1}})\right),$$

and

$$\sum_{k=-\infty}^{-1} a_k^{-1}\mathrm{E}(V^2_{L_p,k,s_1,\cdots,s_{d+1}}(l))$$
$$= \sum_{i=l+1}^{l+L_p}\left(\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i-k}_{p,-\infty})) + a_0^{-1}\mathrm{E}(\mathrm{E}(X^2_{i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i-1}_{p,-\infty}))\right).$$

Proof. The proof follows from Corollary 3.2.1 by taking $P_i(\mathbb{X}_{p,i}) = \mathrm{E}(\mathbb{X}_{p,i}|\mathscr{F}^{i}_{p,-\infty})$. □

By using Corollary 3.2.5, we derive the following proposition which is useful in deriving the main result of this paper.

Proposition 3.2.7. *Suppose* $\{a_k\}_{k=-\infty}^{\infty}$ *are as in Lemma 3.2.1 and* $V_{L_p,k}$ *is as in Corollary 3.2.5. Then,*

$$\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\sum_{k=-\infty}^{\infty} a_k^{-1}\mathrm{E}\left(V^2_{L_p,k,s_1,\cdots,s_{d+1}}(l)\right) \le \left(\sum_{i=l+1}^{l+L_p} c^2_{p,i}\right)\{a_0^{-1}\left(\psi^2(0)+\psi^2(1)\right)$$
$$+ 2\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\psi^2(k)\} < \infty.$$

The proof of Proposition 3.2.7 is given in Appendix B. By using Corollary 3.2.4 and Proposition 3.2.7, we also establish the following lemma which generalizes Lemma 3.2 in McLeish (1977) which becomes a special case with $q_1 = \cdots = q_d = q_{d+1} = 1$.

Lemma 3.2.3. *If Assumption 3.2.3 holds, then for* $L = 1, 2, 3, \ldots; l = 0, 1, 2, \ldots$
$\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left(\max_{j \leqslant L}\left(\sum_{i=l+1}^{l+j} X_{p,i,s_1,\cdots,s_{d+1}}\right)^2\right) \leqslant K \sum_{i=l+1}^{l+L} c_{p,i}^2$, *for some* $K > 0$.

The proof of this lemma is given in Appendix B.

Remark 3.2.4. *From the above corollary, we have*

$$\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left(\left(\sum_{i=1}^{L_p} X_{p,i,s_1,\cdots,s_{d+1}}\right)^2\right) = O\left(\sum_{i=1}^{L_p} c_{p,i}^2\right).$$

For the next proposition, let $\mathbb{X}_{p,i}$ be as in Lemma 3.2.3 and define

$X^a_{p,i,s_1,\cdots,s_{d+1}} = X_{p,i,s_1,\cdots,s_{d+1}} \mathbb{I}[|X_{p,i,s_1,\cdots,s_{d+1}}| \leqslant ac_{p,i}]$, $\mathrm{E}_{i+m} X^a_{p,i,s_1,\cdots,s_{d+1}} = \mathrm{E}(X^a_{p,i,s_1,\cdots,s_{d+1}} | \mathscr{F}^{i+m}_{p,-\infty})$,

$U_{1,i,s_1,\cdots,s_{d+1}} = \mathrm{E}_{i+m} X^a_{p,i,s_1,\cdots,s_{d+1}} - \mathrm{E}_{i-m} X^a_{p,i,s_1,\cdots,s_{d+1}}$,

$U_{2,i,s_1,\cdots,s_{d+1}} = X_{p,i,s_1,\cdots,s_{d+1}} - \mathrm{E}_{i+m} X_{p,i,s_1,\cdots,s_{d+1}} + \mathrm{E}_{i-m} X_{p,i,s_1,\cdots,s_{d+1}}$,

$U_{3,i,s_1,\cdots,s_{d+1}} = \mathrm{E}_{i+m}(X_{p,i,s_1,\cdots,s_{d+1}} - X^a_{p,i,s_1,\cdots,s_{d+1}}) - \mathrm{E}_{i-m}(X_{p,i,s_1,\cdots,s_{d+1}} - X^a_{p,i,s_1,\cdots,s_{d+1}})$. Also, let $v_j^2 = \sum_{i=1}^{j} c_{p,i}^2$, $\tilde{v}_j^2(k) = \sum_{i=k+1}^{k+j} c_{p,i}^2$, $k = 0, 1, \ldots, j$, $j = 1, 2, \ldots$, $\bar{U}_{t,j,s_1,\cdots,s_{d+1}}(l) = \sum_{i=l+1}^{l+j} U_{t,i,s_1,\cdots,s_{d+1}}$,
$t = 1, 2, 3$, $A(a, m) = \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \max_{j \leqslant L} \frac{\bar{U}^2_{1,j,s_1,\cdots,s_{d+1}}}{\tilde{v}_L^2(l)}$,
$B(a, m) = \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \max_{j \leqslant L} \frac{\bar{U}^2_{2,j,s_1,\cdots,s_{d+1}}}{\tilde{v}_L^2(l)}$, $C(a, m) = \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \max_{j \leqslant L} \frac{\bar{U}^2_{3,j,s_1,\cdots,s_{d+1}}}{\tilde{v}_L^2(l)}$.

Proposition 3.2.8. *Suppose that the conditions of Lemma 3.2.3 hold. Then,*

1) *For fixed* (m, a) *and for any* $\epsilon > 0$, *one can choose a, b such that*

   $\mathcal{J}_1(a, b, m) = \mathrm{E}[A(a, m)\mathbb{I}(A(a, m) > b/9)] < \epsilon$.

2) *For any* $\epsilon > 0$, *one can choose m such that* $\mathcal{J}_2(m) = \mathrm{E}(B(m)) < \epsilon$.

3) *For a fixed m, for any* $\epsilon > 0$, *one can choose a such that* $\mathcal{J}_3(a, m) = \mathrm{E}(C(a, m)) < \epsilon$.

The proof of Proposition 3.2.8 is given in Appendix B.

Using this result, we also derive the following proposition which is useful in deriving the asymptotic normality of the UE.

Proposition 3.2.9. *Let b be a constant and let $A_1, A_2, \cdots, A_p$ be non-negative, $q-$integrable random variables for $1 \geq 1$ and let S be a r.v such that $S \leq p \sum_{j=1}^{p} A_j a.s$ with $p \geq 2$. Then,*

1. $\mathrm{E}[S\mathbb{I}(S > b)] \leq p^2\mathrm{E}\left[A_1\mathbb{I}(A_1 > b/p^2)\right] + p^2 \sum_{j=2}^{p} \mathrm{E}(A_j)$.
2. *If* $\mathrm{P}(S \geq 0) = 1$, $\mathrm{E}[S^q\mathbb{I}(S > b)] \leq p^{2q}\mathrm{E}[A_1^q > \frac{b^q}{p^{2q}}] + p^{2q} \sum_{j=2}^{p} \mathrm{E}(A_j^q)$, *for* $q \geq 1$.

The proof of this proposition can be found in Appendix B. By using Propositions 3.2.8 and 3.2.9, we derive below a lemma which is useful in establishing the asymptotic properties of the UE and RE.

Lemma 3.2.4. *Let* $v_j^2 = \sum_{i=1}^{j} c_{p,i}^2$, $\tilde{v}_j^2(k) = \sum_{i=k+1}^{k+j} c_{p,i}^2$, $k = 0, 1, \ldots, j$, $j = 1, 2, \ldots,$ *and* $S_{j,s_1,\cdots,s_{d+1}} = \sum_{i=1}^{j} X_{p,i,s_1,\cdots,s_{d+1}}$, $s_h = 1, \cdots, q_h; h = 1, \ldots, d+1$. *Under Assumption 3.2.3,* $\left\{\sum_{s_1=1}^{q_1} \cdots \sum_{s_d=1}^{q_d} \sum_{s_{d+1}=1}^{q_{d+1}} \max_{j \leqslant L} \frac{\left(S_{k+j,s_1,\cdots,s_{d+1}} - S_{k,s_1,\cdots,s_{d+1}}\right)^2}{\tilde{v}_L^2(k)}; k = 0, 1, \ldots, L; L = 1, 2, \cdots\right\}$ *is a uniformly integrable set. Further,* $\left\{\sum_{s_1=1}^{q_1} \cdots \sum_{s_d=1}^{q_d} \sum_{s_{d+1}=1}^{q_{d+1}} \max_{j \leqslant L} \frac{S^2_{j,s_1,\cdots,s_{d+1}}}{v_L^2}; L = 1, 2, \cdots\right\}$ *is a uniformly integrable set.*

The proof of Lemma 3.2.4 is given in Appendix B. Using Lemma 3.2.4, we derive the following corollary.

Corollary 3.2.6. *Let* $\tilde{v}_i^2 = \sum_{(i-1)b_p+l_p+1}^{ib_p} c_{p,t}^2$. *If Assumption 3.2.3 holds, then the sets*

$\left\{\sum_{s_1=1}^{q_1} \cdots \sum_{s_d}^{q_d} \sum_{s_{d+1}}^{q_{d+1}} \max_{j \leq ib_p} \left(\sum_{t=(i-1)b_p+l_p+1}^{j} X_{p,t,s_1,\cdots,s_{d+1}}\right)^2 / \tilde{v}_i^2, i = 1, 2, \cdots\right\}$,

$\left\{\sum_{s_1=1}^{q_1} \cdots \sum_{s_d}^{q_d} \sum_{s_{d+1}}^{q_{d+1}} \left(\sum_{t=(i-1)b_p+l_p+1}^{ib_p} X_{p,t,s_1,\cdots,s_{d+1}}\right)^2 / \tilde{v}_i^2, i = 1, 2, \cdots\right\}$ *are uniformly integrable.*

Proof. For each i, the uniform integrability of

$$\sum_{s_1=1}^{q_1} \cdots \sum_{s_d}^{q_d} \sum_{s_{d+1}}^{q_{d+1}} \max_{j \leq ib_p} \left(\sum_{t=(i-1)b_p+l_p+1}^{j} X_{p,t,s_1,\cdots,s_{d+1}}\right)^2 / \tilde{v}_i^2$$

follows directly from Lemma 3.2.4 by changing the starting point from 1 to $(i-1)b_p + l_p + 1$ and the end point changes from T to ib_p. This holds uniformly in each i. Moreover, it can be noted that $\sum_{s_1=1}^{q_1} \cdots \sum_{s_d}^{q_d} \sum_{s_{d+1}}^{q_{d+1}} \left(\sum_{t=(i-1)b_p+l_p+1}^{ib_p} X_{p,t,s_1,\cdots,s_{d+1}} \right)^2 / \tilde{v}_i^2$ is a particular case and is thus also uniformly integrable. □

Proposition 3.2.10. *Let $\mathscr{F}_i^*$ be the σ−field generated by $\{U_{ib_p}, U_{ib_p-1}, \cdots\}$, where U_i are random variables defined on $(\Omega, \mathscr{F}, P)$ such that $\mathscr{F}_{i-1}^* \subseteq \mathscr{F}_{p,-\infty}^{i-j}$. Then, if $\{\mathbb{X}_{p,i}, \mathscr{F}_{p,-\infty}^{i}\}$ is an $\mathcal{L}^2$−mixingale of an arbitrary size, $\{\mathrm{E}(\mathbb{X}_{p,i}|\mathscr{F}_{i-1}^*), \mathscr{F}_{p,-\infty}^{i}\}$ is an $\mathcal{L}^2$−mixingale of size $-1/2$.*

Proof. The proof is similar to that given for Proposition A.5 of Chen and Nkurunziza (2016). □

Proposition 3.2.11. *Let $\mathscr{F}_i^*$ be the $\sigma - field$ generated by $\{U_{ib_p}, U_{ib_p-1}, \cdots\}$ with U_i a random variable defined on $(\Omega, \mathscr{F}, P)$ such that $\mathscr{F}_{i-1}^* \subseteq \mathscr{F}_{p,-\infty}^{i-j}$. Then, under Assumption 3.2.3,*

$$\sum_{i=1}^{r_p} \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_{i-1}^*) \xrightarrow[L_p\to\infty]{P} 0 \text{ and } \sum_{i=1}^{r_p} (\mathbb{V}_{p,i} - \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_{i-1}^*)) \xrightarrow[L_p\to\infty]{P} 0.$$

The proof of Proposition 3.2.11 is given in Appendix B.

Proposition 3.2.12. *Suppose that the conditions of Proposition 3.2.11 hold. Then,*

$$\sum_{i=1}^{r_p} \mathbb{W}_{p,i_{(n)}} \mathbb{W}'_{p,i_{(n)}} \xrightarrow[L_p\to\infty]{P} \Sigma_{p,n}, \quad n = 1, \cdots, d+1, \text{ and}$$

$$\sum_{i=1}^{r_p} \mathrm{Vec}(\mathbb{W}_{p,i}) \mathrm{Vec}(\mathbb{W}_{p,i})' \xrightarrow[L_p\to\infty]{P} \Sigma_{p,d+1} \otimes \cdots \otimes \Sigma_{p,1}.$$

Moreover, $\sum_{i=1}^{r_p} \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}[(W_{i,s_1,\cdots,s_{d+1}})^2 \mathbb{I}(\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} W_{i,s_1,\cdots,s_{d+1}}^2 > \epsilon)] \xrightarrow[L_p\to\infty]{P} 0, \forall \epsilon > 0.$

The proof of this proposition can be found in Appendix B

Lemma 3.2.5. *Under Conditions $(\mathscr{C}_5)$ and $(\mathscr{C}_6)$,*

$$\sum_{i=1}^{L_p} \mathbb{X}_{p,i} \xrightarrow[L_p\to\infty]{d} \mathscr{N}_{q_1\times\cdots\times q_{d+1}} \left(0, \bigotimes_{j=d+1}^{1} \Sigma_{p,j}\right).$$

The proof of Lemma 3.2.5 is given in Appendix B. Below, we establish a result which is useful in deriving the asymptotic normality for $(\sum_{i=1}^{L_1} X_{1,i}, \cdots, \sum_{i=1}^{L_{m+1}} X_{m+1,i})$. For the sake of simplicity, let $W_{p,i} = \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_i^*) - \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_{i-1}^*)$, and $\mathbb{W}_i = W_1 \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} W_{m+1,i}$.

Proposition 3.2.13. *Suppose that the conditions of Proposition 3.2.11 hold and let*

$$\zeta_{1,a,b,i,n} = (\mathbb{V}_{a,i_{(n)}} - \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_i^*) + \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_{i-1}^*))(\mathbb{V}_{b,i_{(n)}} + \mathrm{E}(\mathbb{V}_{b,i_{(n)}}|\mathscr{F}_i^*) - \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_{i-1}^*))'$$

and

$$\begin{aligned}\zeta_{2,a,b,i,n} = {} & \mathbb{V}_{a,i_{(n)}}(\mathrm{E}(\mathbb{V}_{b,i_{(n)}}|\mathscr{F}_i^*))' - \mathbb{V}_{a,i_{(n)}}(\mathrm{E}(\mathbb{V}_{b,i_{(n)}}|\mathscr{F}_{i-1}^*))' - (\mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_i^*))\mathbb{V}'_{b,i_{(n)}} \\ & + (\mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_{i-1}^*))\mathbb{V}'_{b,i_{(n)}}.\end{aligned}$$

Then,

$$\sum_{a=1}^{m+1}\sum_{b=1}^{m+1}\sum_{i=1}^{r_{min}} \|\zeta_{1,a,b,i,n}\|_1 = o(1) \text{ and } \sum_{a=1}^{m+1}\sum_{b=1}^{m+1}\sum_{i=1}^{r_{min}} \|\zeta_{2,a,b,i,n}\|_1 = o(1). \tag{3.6}$$

The proof of this propositions can be found in Appendix B.

Proposition 3.2.14. *Let $r_{min} = \min(r_1, \cdots, r_{m+1})$ and $L_{min} = \min(L_1, \cdots, L_{m+1})$. Suppose that the conditions of Proposition 3.2.11 hold. Then,*

$$\sum_{i=1}^{r_{min}} [\mathbb{V}_{i_{(n)}}\mathbb{V}'_{i_{(n)}} - \mathbb{W}_{i_{(n)}}\mathbb{W}'_{i_{(n)}}] \xrightarrow[L_{min}\to\infty]{P} 0, \tag{3.7}$$

for $n = 1, \cdots, d+1$ with $A_{(n)}$ denoting the mode-n matricization of A. Also,

$$\sum_{i=1}^{r_{min}} [\mathrm{Vec}(\mathbb{V}_i)\mathrm{Vec}(\mathbb{V}_i)' - \mathrm{Vec}(\mathbb{W}_i)\mathrm{Vec}(\mathbb{W}_i)'] \xrightarrow[L_{min}\to\infty]{P} 0. \tag{3.8}$$

The proof of Proposition 3.2.14 is given in Appendix B. From Proposition 3.2.14, we derive the following proposition which constitutes a version of the Lindeberg's Condition in the context of random tensors. Thus, the established proposition plays a crucial role in deriving the asymptotic normality of the UE. Set $\|\mathbb{W}_{a,i}\|_{\mathrm{F}}^2 = \sum_{s_1=1}^{q_1} \cdots \sum_{s_d=1}^{q_d} \sum_{s_{d+1}=1}^{q_{d+1}} W^2_{a,i,s_1,\cdots,s_{d+1}}$.

Proposition 3.2.15. *Suppose that the conditions in Proposition 3.2.11 hold. Then,*

$$\sum_{a=1}^{m+1}\sum_{i=1}^{r_a}\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left[(W_{a,i,s_1,\cdots,s_{d+1}})^2 \mathbb{I}\left(\|\mathbb{W}_{a,i}\|_{\mathrm{F}}^2 > \epsilon\right)\right] \xrightarrow[T\to\infty]{P} 0 \tag{3.9}$$

for all $\epsilon > 0$. *In addition,*

$$\sum_{i=1}^{r_{min}} \mathbb{W}_{i_{(n)}}\mathbb{W}'_{i_{(n)}} \xrightarrow[T\to\infty]{P} \Lambda_n, n = 1, \cdots, d+1 \tag{3.10}$$

and

$$\sum_{i=1}^{r_{min}} \mathrm{Vec}(\mathbb{W}_i)\mathrm{Vec}(\mathbb{W}_i)' \xrightarrow[T\to\infty]{P} \bigotimes_{i=d+1}^{1} \Lambda_i. \tag{3.11}$$

The proof of this proposition is given in Appendix B.

Note that

$$T^{-1/2}\mathbb{U} \times_{(d+1)} Z^{0'} = \sum_{i=1}^{L_1} \mathbb{X}_{1,i} \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \sum_{i=1}^{L_{m_0+1}} \mathbb{X}_{m_0+1,i}.$$

To simplify some notations, let $\mathbb{B}_{p,j}(t) = \sum_{i=1}^{\lfloor jt \rfloor} \mathbb{X}_{p,i}$, $\mathbb{B}_j(t) = \mathbb{B}_{1,j}(t) \boxplus_{(d+1)} \mathbb{B}_{2,j}(t) \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \mathbb{B}_{m+1,j}(t)$, let $D^k([0, 1])$ denote the space of all k-column vectors of functions which are right continuous with left limits on $[0, 1]$ and let $\Lambda^* = \bigotimes_{j=d+1}^{1} \Lambda_j$. We present the main result of this subsection in the following lemma.

Lemma 3.2.6. *If Assumption 3.2.3 holds, then* $T^{-1/2}\mathbb{U} \times_{d+1} Z^{0'} \xrightarrow[T\to\infty]{d} \epsilon_1^{*0}$ *and for each* $t \in [0, 1]$, $\mathbb{B}_T(t) \xrightarrow[T\to\infty]{d} \sqrt{t}\,\epsilon_1^{*0}$ *where* $\epsilon_1^{*0} \sim \mathcal{N}_{q_1\times\cdots\times q_d\times(m+1)q_{d+1}}(0, \Lambda^*)$.

The proof of this lemma is given in Appendix B.

Lemma 3.2.6 states that, under the weak dependence structure of the regressors and tensor error, the tensor product of the error term with the regressors converges to a normal tensor variate. This result is essential in establishing the asymptotic normality of the UE given in Section 3.2.3. This result is the second main contribution and it plays an important role in deriving the results of the following subsection. All propositions, lemmas and corollaries given in Section 3.2.1 are the building blocks in the derivation of some key results in obtaining Lemma 3.2.6. It should also be noted that the above results are useful in deriving Lemma 3.2.4 which generalizes some results in McLeish (1977) and Chen and Nkurunziza (2016) which become special cases with $q_1 = \cdots = q_d = q_{d+1} = 1$ and $d = 1$, respectively.

3.2.3 Asymptotic properties of the UE and the RE

In this subsection, we present some asymptotic properties of the UE and RE. The established results are useful in evaluating the relative efficiency of the proposed estimators. To simplify some mathematical expressions, let $L_p = \tau_p^0 - \tau_{p-1}^0$. Note that the condition $\min_{1 \leqslant i \leqslant m_0+1}(L_i) \to \infty$ is equivalent to $T \to \infty$ and thus, under $(\mathscr{C}_1)$- $(\mathscr{C}_4)$, $(L_p)^{-1} \sum_{t=\tau_{p-1}^0+1}^{\tau_{p-1}^0+[L_p]} z_t z_t' \xrightarrow[T\to\infty]{\mathrm{P}} Q_p$. Then, under $(\mathscr{C}_1)$-$(\mathscr{C}_4)$, $T^{-1}\bar{Z}^{0\prime}\bar{Z}^0 \xrightarrow[T\to\infty]{\mathrm{P}} \Gamma$, where Γ is a $(m_0+1)q_{d+1} \times (m_0+1)q_{d+1}$ non-random, positive definite matrix. Also, under $(\mathscr{C}_6)$, $T^{-1}\mathrm{Vec}(\mathbb{U} \times_{d+1} \bar{Z}^{0\prime})\left(\mathrm{Vec}(\mathbb{U} \times_{d+1} \bar{Z}^{0\prime})\right)'$ converges in probability to a non-random matrix $\bigotimes_{j=d+1}^{1} \Lambda_j$. The following proposition gives the asymptotic distribution of the UE. In the sequel, let $\Sigma_{11}^* = \Gamma^{-1}\Lambda_{d+1}\Gamma^{-1} \otimes \Lambda^{**}$ with $\Lambda^{**} = \bigotimes_{j=d}^{1} \Lambda_j$.

Proposition 3.2.16. *Let* $\epsilon_{1,T}^*(\tau) = \sqrt{T}(\hat{\delta}(\tau) - \delta)$. *Under Assumption 3.2.3, we have* $\epsilon_{1,T}^*(\tau) \xrightarrow[T\to\infty]{d} \epsilon_1^*$ *with* $\epsilon_1^* \sim \mathscr{N}_{q_1 \times \cdots \times q_d \times (m_0+1)q_{d+1}}\left(0, \Sigma_{11}^*\right)$.

The proof of Proposition 3.2.16 is outlined in Appendix B. By using Proposition 3.2.16, we derive the asymptotic normality of the RE under the restriction in (3.2). To this end, let $\Omega_i = \mathbb{J}_i,\ R_i'(R_iR_i')^{-1}R_i$, $\mathbb{G}_i = R_i'(R_iR_i')^{-1}$, for $i = 1, \cdots, d$, $\Omega_{d+1} = \Gamma^{-1}R_{d+1}(R_{d+1}\Gamma^{-1}R_{d+1}')^{-1}R_{d+1}$, $\Omega = \bigotimes_{j=d+1}^{1} \Omega_j$, $\mathbb{G}_{d+1}^* = \Gamma^{-1}R_{d+1}'(R_{d+1}\Gamma^{-1}R_{d+1}')^{-1}$,

$$\Sigma_{22}^* = \Sigma_{11}^* - \Omega_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1} \otimes \left(\bigotimes_{j=d}^{1} \Omega_j\Lambda_j\right) - \Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}' \otimes \left(\bigotimes_{j=d}^{1} \Lambda_j\Omega_j'\right)$$
$$+\Omega_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}' \otimes \left(\bigotimes_{j=d}^{1} \Omega_j\Lambda_j\Omega_j'\right).$$

Proposition 3.2.17. *Under Assumption 3.2.3 and restriction (3.2),*

$$\sqrt{T}(\tilde{\boldsymbol{\delta}}(\tau) - \boldsymbol{\delta}) \xrightarrow[T\to\infty]{d} \epsilon_2^* \sim \mathcal{N}_{(m+1)q_1\times\cdots\times q_d\times q_{d+1}}(0, \Sigma_{22}^*).$$

The proof of Proposition 3.2.17 is given in Appendix B.

Next, we derive the joint asymptotic normality of the RE and the UE. In particular, the joint asymptotic normality is established in the context where the restriction in (3.2) may not hold. To this end, as in the matrix parameter case, in order to avoid some degeneracy of the limiting distribution of $\tilde{\boldsymbol{\delta}}$, we consider the following sequence of local alternatives

$$H_{1T} : \boldsymbol{\delta} \times_1 R_1 \times_2 R_2 \times_3 \cdots \times_{d+1} R_{d+1} = \boldsymbol{r} + \frac{\boldsymbol{r}_0}{\sqrt{T}}, T = 1, 2, \cdots \tag{3.12}$$

where $\boldsymbol{r}_0$ is an $l_1 \times \cdots \times l_{d+1}$ tensor with $\|\boldsymbol{r}_0\| < \infty$. To introduce some notation, let $\epsilon_{2,T}^*(\tau) = \sqrt{T}(\tilde{\boldsymbol{\delta}}(\tau) - \boldsymbol{\delta})$, and let $\epsilon_{3,T}^*(\tau) = \sqrt{T}(\hat{\boldsymbol{\delta}}(\tau) - \tilde{\boldsymbol{\delta}}(\tau))$, $\mathbb{G}^* = \mathbb{G}_{d+1}^* \otimes \bigotimes_{i=d}^{1} \mathbb{G}_i$, $\mu^{**} = \left(-\boldsymbol{r}_0 (\bigtimes_{j=1}^{d})_j \mathbb{G}_j\right) \times_{d+1} \mathbb{G}_{d+1}^*$, $\Sigma_{12}^* = \Sigma_{11}^* - \Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}' \otimes \left(\bigotimes_{j=d}^{1} \Lambda_j \Omega_j'\right)$, $\Sigma_{21}^* = \Sigma_{12}^*{}'$, $\Sigma_{31}^{*\prime} = \Sigma_{13}^{*\prime}$,$\Sigma_{23}^* = \Sigma_{32}^{*\prime}$, $\Sigma_{13}^* = \Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}' \otimes \left(\bigotimes_{j=d}^{1} \Lambda_j \Omega_j'\right)$, $\Sigma_{33}^* = \Omega_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}' \otimes \left(\bigotimes_{j=d}^{1} \Omega_j \Lambda_j \Omega_j'\right)$, $\Sigma_{32}^* = \Omega_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1} \otimes \left(\bigotimes_{j=d}^{1} \Omega_j \Lambda_j\right) - \Omega_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}' \otimes \left(\bigotimes_{j=d}^{1} \Lambda_j \Omega_j'\right)$.

Proposition 3.2.18. *Under Assumption 3.2.3, along with (3.12),*

$\epsilon_{1,T}^*(\tau) \boxplus_{(d+1)} \epsilon_{2,T}^*(\tau) \boxplus_{(d+1)} \epsilon_{3,T}^*(\tau) \xrightarrow[T\to\infty]{d} \epsilon_1^* \boxplus_{(d+1)} \epsilon_2^* \boxplus_{(d+1)} \epsilon_3^*$ *where*

$$\epsilon_1^* \boxplus_{(d+1)} \epsilon_2^* \boxplus_{(d+1)} \epsilon_3^* \sim \mathcal{N}_{q_1\times\cdots\times q_d\times 3(m+1)q_{d+1}}\left(0 \boxplus_{(d+1)} \mu^{**} \boxplus_{(d+1)} -\mu^{**}, \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* & \Sigma_{13}^* \\ \Sigma_{21}^* & \Sigma_{22}^* & \Sigma_{23}^* \\ \Sigma_{31}^* & \Sigma_{32}^* & \Sigma_{33}^* \end{pmatrix}\right).$$

Further, $\epsilon_{1,T}^*(\hat{\tau}) \boxplus_{(d+1)} \epsilon_{2,T}^*(\hat{\tau}) \boxplus_{(d+1)} \epsilon_{3,T}^*(\hat{\tau}) \xrightarrow[T\to\infty]{d} \epsilon_1^* \boxplus_{(d+1)} \epsilon_2^* \boxplus_{(d+1)} \epsilon_3^*$.

The proof of Proposition 3.2.18 is outlined in Appendix B

3.3 A class of shrinkage estimators and risk functions

In this section, we propose a class of James-Stein type estimators for the tensor parameter $\boldsymbol{\delta}$. Let $A_{d+1} = R'_{d+1}(R_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}R'_{d+1})^{-1}R_{d+1}$ and for $j = 1, \cdots, d$, let $A_j = R'_j(R_j\Lambda_j R'_j)^{-1}R_j$, $\hat{A}_j = R'_j(R_j\hat{\Lambda}_j R'_j)^{-1}R_j$, $\hat{A}_{d+1} = R'_{d+1}(R_{d+1}\hat{\Gamma}^{-1}\hat{\Lambda}_{d+1}\hat{\Gamma}^{-1}R'_{d+1})^{-1}R_{d+1}$ where $\hat{\Lambda}_j$ is a consistent estimator of Λ_j. Further, let $\hat{\Lambda}_{d+1}$ and $\hat{\Gamma}$ be consistent estimators of Λ_{d+1} and Γ_{d+1}, respectively. Furthermore, let $\boldsymbol{\delta}^* = \left(\hat{\boldsymbol{\delta}}(\hat{\tau}) - \tilde{\boldsymbol{\delta}}(\hat{\tau})\right)(\bigtimes_{j=1}^{d+1})_j \hat{A}_j^{1/2}$, let $\psi = T\text{trace}\left(\boldsymbol{\delta}^{*}_{(d)}{}'\boldsymbol{\delta}^{*}_{(d)}\right)$ and let h be a known Borel measurable and real-valued integrable function. Let $\boldsymbol{\theta}$ be a tensor parameter, $\hat{\boldsymbol{\theta}}$ be an unrestricted tensor estimator for $\boldsymbol{\theta}$ and let $\tilde{\boldsymbol{\theta}}$ be a restricted estimator for $\boldsymbol{\theta}$. We consider the following class of tensor estimators

$$\hat{\boldsymbol{\vartheta}}(h, \hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{\theta}} + h\left(T\|\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\|^2_{\{\Xi_i, i=1,\cdots,d+1\}}\right)\left(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\right). \tag{3.13}$$

In the sequel, we consider that h is continuous. Note that

$$\hat{\boldsymbol{\vartheta}}(1, \hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}) = \hat{\boldsymbol{\theta}} \quad \text{and} \quad \hat{\boldsymbol{\vartheta}}(0, \hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{\theta}}. \tag{3.14}$$

Thus, the UE and the RE both belong to the class of estimators in (3.13) by setting $h = 1$ and $h = 0$, respectively. Another set of estimators that are members of this class are the James-Stein and Positive-rule James-Stein estimators, denoted as $\hat{\boldsymbol{\theta}}^s$ and $\hat{\boldsymbol{\theta}}^{sp}$, respectively. Indeed, by letting $h(x) = 1 - \frac{c}{x}$, $c > 0$, we obtain the James-Stein estimator and by taking $h(x) = \max\{0, 1 - \frac{c}{x}\}$, $c > 0$, we get the Positive-rule James-Stein estimator. As such, this class is also known as shrinkage-type estimators of $\boldsymbol{\theta}$. Overall, the class of estimators combines both the sample information and non-sample information from the uncertain restriction in (3.2). In the context of the tensor change-point model in (3.1), the shrinkage estimators (SEs) are obtained as above by taking $n = T$ and $c = \prod_{j=1}^{d+1} l_i - 2$. Namely, let

$$\hat{\boldsymbol{\delta}}^s = \tilde{\boldsymbol{\delta}}(\hat{\tau}) + \left(1 - \left(\prod_{j=1}^{d+1} l_j - 2\right)\psi^{-1}\right)\left(\hat{\boldsymbol{\delta}}(\hat{\tau}) - \tilde{\boldsymbol{\delta}}(\hat{\tau})\right),$$

$$\hat{\boldsymbol{\delta}}^{s+} = \tilde{\boldsymbol{\delta}}(\hat{\tau}) + \left(1 - \left(\prod_{j=1}^{d+1} l_j - 2\right)\psi^{-1}\right)^{+} \left(\hat{\boldsymbol{\delta}}(\hat{\tau}) - \tilde{\boldsymbol{\delta}}(\hat{\tau})\right).$$

Define $(\hat{\theta} - \theta)^* = (\hat{\theta} - \theta)(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}$, where $\mathbb{W}_j, j = 1, \cdots, d+1$ are non-negative definite matrices. To evaluate the performance of the proposed estimator, we use a criterion known as the asymptotic distributional risk (ADR). This is defined as $\mathrm{ADR}^1(\hat{\theta}, \theta; \mathbb{W}) = \mathrm{E}\left(\mathrm{trace}\left(\rho^{*'}_{(d)}\rho^*_{(d)}\right)\right)$, where the random tensor $\rho^* = \rho(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}$ with $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow[T\to\infty]{d} \rho$. In the following subsection, we establish some preliminary results using identities in Section 2.1 that help to establish the ADR^1.

3.3.1 Preliminary results in shrinkage methods

In this subsection, we present some propositions that follow from results established in Section 2.1 that are useful in deriving the risk functions of the proposed shrinkage estimators. To set up notations, let $\Lambda_{11}^* = \bigotimes_{j=d+1}^{1} \Lambda_{Xj}$, with $\Lambda_{Xj} = \mathbb{G}_j R_j \Lambda_j R_j' \mathbb{G}_j'$ for $j = 1, \cdots, d$, $\Lambda_{Xd+1} = \mathbb{G}_{d+1}^* R_{d+1} \Gamma^{-1} \Lambda_{d+1} \Gamma^{-1} R_{d+1}' \mathbb{G}_{d+1}^{*'}$.

Proposition 3.3.1. *Let* $\mathbb{W}_j = A_j^{1/2} \mathbb{W}_j^* A_j^{1/2}, j = 1, \cdots, d+1$, *where* $\mathbb{W}_j^*$ *are non-negative definite matrices. Let* ϵ_3^* *be a random tensor as defined in Proposition 3.2.18, let* $\epsilon_{31}^* = \epsilon_3^*(\bigtimes_{j=1}^{d+1})_j A_j^{1/2}$, $\mu^*{}_1 = \mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j^{1/2}$, $\Delta = \mathrm{trace}\left(\mu^{*'}_{1_{(d)}} \mu^*_{1_{(d)}}\right)$, *and let h be as in Theorem 2.1.1. Then,* $\mathrm{E}\left[h\left(\mathrm{trace}\left(\epsilon^{*'}_{31_{(d)}} \epsilon^*_{31_{(d)}}\right)\right) \epsilon_3^*(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j\right] = \mathrm{E}\left[h\left(\chi^2_{l+2}(\Delta)\right)\right]\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j\right)$.

The proof of this proposition is given in Appendix B.

Proposition 3.3.2. *Let* $\epsilon_{32}^* = \epsilon_3^*(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}$, $\mu_2^* = \mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}$, $D_1 = \prod_{j=1}^{d+1} \mathrm{trace}(\mathbb{W}_j \mathbb{J}_j)$ *and let* $D_2 = \mathrm{trace}(\mu^{*'}_{2_{(d)}} \mu^*_{2_{(d)}})$ *and let h be as in Theorem 2.1.1. If the conditions of Proposition 3.3.1 hold, then,*

$$\mathrm{E}\left[h\left(\mathrm{trace}\left(\epsilon^{*'}_{31_{(d)}} \epsilon^*_{31_{(d)}}\right)\right) \mathrm{trace}\left(\epsilon^{*'}_{32_{(d)}} \epsilon^*_{32_{(d)}}\right)\right] = \mathrm{E}\left[h(\chi^2_{l_1 \cdots l_{d+1}+2}(\Delta))\right] D_1 + \mathrm{E}\left[h(\chi^2_{l_1 \cdots l_{d+1}+4}(\Delta))\right] D_2.$$

Proof. The result follows directly from Theorem 2.1.2 using the fact that $\epsilon_3^* \sim \mathcal{N}_{q_1\times\cdots\times q_d\times(m+1)q_{d+1}}\left(-\mu^{**}, \Sigma_{33}^*\right)$. □

Next, we derive the following proposition which is a key result in deriving the ADR of the proposed class of shrinkage estimators. For simplicity, let $\Upsilon_{d+1}^* = \Omega_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}'$, $\Upsilon_j^* = \Omega_j\Lambda_j\Omega_j'$, $C_j^* = \Lambda_j$, $B_j^* = \Lambda_j\Omega_j'$, $D_j^* = B_j^{*\prime}$, for $j = 1, \cdots, d$, $B_{d+1}^* = \Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}'$, $C_{d+1}^* = \Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}$, $c_1 = \text{trace}\left(\mu_{(d)}^{**\prime}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right)$, $c_2 = \prod_{j=1}^{d+1} \text{trace}(\mathbb{W}_j B_j^*)$, $c_3 = \prod_{j=1}^{d+1} \text{trace}(\mathbb{W}_j\Upsilon_j^*)$, $c_4 = \text{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}' \left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right)$.

Proposition 3.3.3. *Let ϵ_2^* and ϵ_3^* be as defined in Proposition 3.2.18. Let $h(.)$, $\mathbb{W}_j$, $j = 1, \cdots, d+1$ be as in Proposition 3.3.1. Then,*

$$\mathrm{E}\left[h\left(\text{trace}\left(\epsilon_{31_{(d)}}^{*\prime}\epsilon_{31_{(d)}}^*\right)\right)\text{trace}\left(\left(\epsilon_2^*(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}'\left(\epsilon_3^*(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right)\right]$$

$$= -\mathrm{E}\left[h\left(\chi_{l+2}^2(\Delta)\right)c_1\right] + \mathrm{E}\left[h\left(\chi_{l+2}^2(\Delta)\right)\right][c_2 - c_3] + \mathrm{E}\left[h\left(\chi_{l+4}^2(\Delta)\right)\right]c_1 - \mathrm{E}\left[h\left(\chi_{l+4}^2(\Delta)\right)\right]c_4.$$

Proof. From Proposition 3.2.18 we have $\epsilon_3^* \boxplus_{(d+1)} \epsilon_2^* \sim \mathcal{N}_{q_1\times\cdots\times q_d\times q_{d+1}}\left(-\mu^{**}\boxplus_{(d+1)}\mu^{**}, \begin{pmatrix}\Pi_{11}^* & \Pi_{12}^*\\ \Pi_{21}^* & \Pi_{22}^*\end{pmatrix}\right)$, with $\Pi_{11}^* = \bigotimes_{j=d+1}^{1}\Upsilon_j^*$, $\Pi_{21}^* = \bigotimes_{j=d+1}^{1} B_j^* - \bigotimes_{j=d+1}^{1}\Upsilon_j^*$, $\Pi_{22}^* = \bigotimes_{j=d+1}^{1} C_j^* - \bigotimes_{j=d+1}^{1} D_j^* - \bigotimes_{j=d+1}^{1} B_j^* + \bigotimes_{j=d+1}^{1}\Upsilon_j^*$. Therefore, the result follows by applying Theorem 2.1.3 with the appropriate substitutions. This completes the proof. □

3.3.2 Asymptotic distributional risk (ADR)

In this subsection, we derive the ADR of the class of estimators $\hat{\vartheta}(h, \hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}})$ as defined in (3.13). To this end, we assume that the weight matrices $\mathbb{W}_j = A_j^{1/2}\mathbb{W}_j^* A_j^{1/2}$, $\mathbb{W}_j^*$ non-negative definite matrices for $j = 1, \cdots, d+1$, and let $\mathbb{W} = [\mathbb{W}_1, \mathbb{W}_2, \cdots \mathbb{W}_{d+1}]$.

Lemma 3.3.1. *Let* $\hat{\vartheta}(h, \hat{\theta}, \tilde{\theta})$ *be as in* (3.13) *with h a continuous and square integrable function. If Assumption 3.2.3 holds along with* (3.12)*, then,*

$$\begin{aligned}\mathrm{ADR}^1\left(\hat{\vartheta}(h, \hat{\theta}, \tilde{\theta}), \delta, \mathbb{W}\right) = \mathrm{ADR}^1\left(\tilde{\delta}(\hat{\tau}), \delta, \mathbb{W}\right) - 2\mathrm{E}\left[h\left(\chi^2_{l+2}(\Delta)\right)\right]c_1 + 2\mathrm{E}\left[h\left(\chi^2_{l+2}(\Delta)\right)\right]c_2 \\ -2\mathrm{E}\left[h\left(\chi^2_{l+2}(\Delta)\right)\right]c_3 + 2\mathrm{E}\left[h\left(\chi^2_{l+4}(\Delta)\right)\right]c_1 + \mathrm{E}\left[h^2(\chi^2_{l+2}(\Delta))\right]c_3 \\ -2\mathrm{E}\left[h\left(\chi^2_{l+4}(\Delta)\right)\right]c_4 + \mathrm{E}\left[h^2(\chi^2_{l+4}(\Delta))\right]c_4.\end{aligned}$$

The proof of Lemma 3.3.1 is outlined in Appendix B. From Proposition 3.2.18, we also derive the following lemma which gives the ADR of the UE and the RE.

Lemma 3.3.2. *If Assumption 3.2.3 holds along with* (3.12)*, then,*

$$\mathrm{ADR}^1\left(\hat{\delta}(\hat{\tau}), \delta^0, \mathbb{W}\right) = \prod_{j=1}^{d+1} \mathrm{trace}\left(\mathbb{W}_j C^*_j\right),$$
$$\mathrm{ADR}^1\left(\tilde{\delta}(\hat{\tau}), \delta^0, \mathbb{W}\right) = \mathrm{ADR}^1\left(\hat{\delta}(\hat{\tau}), \delta^0, \mathbb{W}\right) - 2c_2 + c_4 + c_3.$$

The proof of this lemma is given in Appendix B. By using Lemma 3.3.1 and Lemma 3.3.2, we derive the ADR of $\hat{\delta}^s$ and the ADR of $\hat{\delta}^{s+}$. Let $h_1(x) = 1 - ((l-2)/x)$, and let $h_3(x) = [1 - ((l-2)/x)]\mathbb{I}(x < l-2)$, $x > 0$.

Corollary 3.3.1. *Under Assumption 3.2.3 and* (3.12)*,* $\mathrm{ADR}^1\left(\hat{\delta}^s, \delta, \mathbb{W}\right) = \mathrm{ADR}^1\left(\tilde{\delta}(\hat{\tau}), \delta, \mathbb{W}\right)$ $-2\mathrm{E}\left[h_1\left(\chi^2_{l+2}(\Delta)\right)\right]c_1 + 2\mathrm{E}\left[h_1\left(\chi^2_{l+2}(\Delta)\right)\right]c_2 - 2\mathrm{E}\left[h_1\left(\chi^2_{l+2}(\Delta)\right)\right]c_3 + 2\mathrm{E}\left[h_1\left(\chi^2_{l+4}(\Delta)\right)\right]c_1$ $+ \mathrm{E}\left[h_1^2(\chi^2_{l+2}(\Delta))\right]c_3 - 2\mathrm{E}\left[h_1\left(\chi^2_{l+4}(\Delta)\right)\right]c_4 + \mathrm{E}\left[h_1^2(\chi^2_{l+4}(\Delta))\right]c_4$*;*

$$\begin{aligned}\mathrm{ADR}\left(\hat{\delta}^{s+}, \delta^0, \mathbb{W}\right) = \mathrm{ADR}\left(\hat{\delta}^s, \delta^0, \mathbb{W}\right) + 2\mathrm{E}\left[h_3\left(\chi^2_{l+2}(\Delta)\right)\right]c_1 - 2\mathrm{E}\left[h_3\left(\chi^2_{l+2}(\Delta)\right)\right]c_2 \\ +2\mathrm{E}\left[h_3\left(\chi^2_{l+2}(\Delta)\right)\right]c_3 - 2\mathrm{E}\left[h_3\left(\chi^2_{l+4}(\Delta)\right)\right]c_1 - \mathrm{E}\left[h_3^2(\chi^2_{l+2}(\Delta))\right]c_3 \\ +2\mathrm{E}\left[h_3\left(\chi^2_{l+4}(\Delta)\right)\right]c_4 - \mathrm{E}\left[h_3^2(\chi^2_{l+4}(\Delta))\right]c_4.\end{aligned}$$

The proof of this corollary is given in Appendix B. From Corollary 3.3.1, we derive below a result which gives a sufficient condition for the tensor RE to dominate the UE. For simplicity, let $\varpi = \bigotimes_{j=d+1}^{1} A_j^{1/2}\Upsilon_j^*\mathbb{W}_j\Upsilon_j^*A_j^{1/2}$ and let $\mathrm{Ch}_{\max}(B)$ and $\mathrm{Ch}_{\min}(B)$ denote the maximum and minimum eigenvalues of a matrix B, respectively.

Corollary 3.3.2. *Suppose that Assumption 3.2.3 holds along with* (3.12)*. Then, if* $\Delta \leq \frac{2c_2 - c_3}{\mathrm{Ch}_{\max}(\varpi)}$, $\mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right) \leqslant \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right)$*. Moreover, if* $\Delta \geqslant \frac{2c_2 - c_3}{\mathrm{Ch}_{\min}(\varpi)}$*, then,* $\mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right) \geqslant \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right)$.

The proof of this corollary is outlined in Appendix B.

From Corollary 3.3.1, we also derive a result which gives a sufficient condition for the tensor SEs to dominate the UE. Namely, in the following corollary, we show that for certain weighting matrices, $\mathbb{W}_j$, $j = 1, \cdots, d+1$, the SEs always dominate the UE. Towards that end, let $\Pi^* = \bigotimes_{j=d+1}^{1} A_j^{1/2}\left(4 \bigotimes_{j=d+1}^{1} B_j^* + (l-2) \bigotimes_{j=d+1}^{1} \Upsilon_j^*\right) \bigotimes_{j=d+1}^{1} \mathbb{W}_j \Upsilon_j^* A_j^{1/2}$, $\Pi^{**} = \frac{\Pi^* + \Pi^{*\prime}}{2}$.

Corollary 3.3.3. *Let* $c_2 \geqslant \max\{c_3/2, \mathrm{Ch}_{\max}(\Pi^{**})/4\}$ *and suppose that Assumption 3.2.3 holds along with* (3.12)*. Then, for all* $\Delta \geq 0$*,*

$$\mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^{s+}, \boldsymbol{\delta}^0, \mathbb{W}\right) \leqslant \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^{s}, \boldsymbol{\delta}^0, \mathbb{W}\right) \leqslant \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right).$$

The proof of Corollary 3.3.3 is given in Appendix B. We note that the established sufficient condition for the risk dominance of SEs in Corollary 3.3.3 is more general than the one given in existing literature such as, for example, in Saleh (2006), Hossain et al. (2016) among others. Thus, the cases studied in the quoted papers are special cases with $d = 1$ and where the shrinking factor and the RE are uncorrelated. Namely, other than the fact that this is a vector case, the scenario studied in the quoted papers correspond to the case where $\boldsymbol{\Sigma}_{32} = \boldsymbol{\Sigma}_{23}' = \mathbf{0}$. To illustrate the importance of the condition above, note that if $d = 1$ or $d = 2$, $\boldsymbol{\Sigma}_{32} = \boldsymbol{\Sigma}_{23}' = \mathbf{0}$, the classical sufficient condition for the risk dominance is $\prod_{i=1}^{d} \mathrm{trace}(\mathbb{W}_i \Xi_i^*) \geqslant \frac{p+2}{2} \mathrm{Ch}_{\max}\left(\bigotimes_{i=d}^{1} \mathbb{W}_i \Xi_i^*\right)$, which is a special case of the condition given in Corollary 3.3.3. Indeed, in this case, the condition in Corollary 3.3.3 yields $\prod_{i=1}^{d} \mathrm{trace}(\mathbb{W}_i \Xi_i^*) \geqslant \frac{p+2}{4} \mathrm{Ch}_{\max}\left(\bigotimes_{i=d}^{1} \mathbb{W}_i \Xi_i^*\right)$ and noting that, for $b \geqslant 0$, $\left\{x : x \geqslant \frac{b}{2}\right\} \subset \left\{x : x \geqslant \frac{b}{4}\right\}$, the condition of Corollary 3.3.3 is more general than the one given in the quoted literature.

3.4 The case of unknown number of change-points

In the previous sections, we assumed that the number of change points, m, was known. However, this is not always the case and we often face the challenge of analyzing a data set in which very little information is given about the number of the change-points. In this section, we describe a method that can be used to estimate m and $\boldsymbol{\tau}$ and how to determine the UE and RE when both parameters are unknown. We also present an asymptotic result which is used to overcome the challenge due to the randomness of the dimensions of the proposed tensor estimators.

3.4.1 Estimating the number of change points

In this subsection, we describe a method to estimate m and $\boldsymbol{\tau}$. We estimate m by choosing the value that gives the best fitting model. Thus, we consider the following penalty function to choose the best fitting model

$$\mathrm{IC}(m) = -2\,\mathrm{SSR}_T^U(\hat{\boldsymbol{\tau}}(m)) + (m+1)\nu(q_{d+1})\gamma(T), \tag{3.15}$$

where $\hat{\boldsymbol{\tau}}(m)$ is established in (2.3) corresponding to each m; $\nu(x) = x + 1$; $\gamma(T)$ is a non-decreasing function of T; and m is the potential number of change points. Note that, we can also include the restriction in the penalty function to obtain the restricted estimator for m. However, since the goal is to obtain a consistent estimator for m, we choose to ignore such a penalty function for the sake of simplicity. The function in (3.15) is known as the least squares-based information criterion and in the case where $\gamma(T) = \log(T)$, the function yields the Schwarz information criterion (SIC) as in Schwarz (1978). We also prove that, as T is large, the $\mathrm{IC}(m)$ reaches its minimum value when $m = m_0$ where m_0 is the true number of change-points. As such, by minimizing the $\mathrm{IC}(m)$, one can detect m_0. The outline of the algorithm used to estimate m and $\boldsymbol{\tau}$ is derived from the dynamic algorithm described in Qu and Perron (2007) and can be found in Nkurunziza et al. (2019). Let $\hat{m}$ be an estimator of m_0, obtained from (3.15).

Theorem 3.4.1. *Under Assumption 3,* **(1).** $\lim_{T\to\infty} \mathrm{P}\left(\mathrm{IC}(m_0) < \mathrm{IC}(m)\right) = 1,\ \forall\ m < m_0$; **(2).** $\lim_{T\to\infty} \mathrm{P}\left(\mathrm{IC}(m_0) < \mathrm{IC}(m)\right) = 1,\ \forall\ m > m_0$; **(3).** $\lim_{T\to\infty} \mathrm{P}\left(\mathrm{IC}(m_0) < \mathrm{IC}(m)\right) = 1,\ \forall\ m \neq m_0$. **(4).** *$\hat{m}$ is a consistent estimator for m_0.*

The proof of this theorem is given in Appendix B. Note that Parts **(1)-(3)** of Theorem 3.4.1 show that the proposed penalty function, IC(m) reaches its minimum value when $m = m_0$ and thus, this guarantees that our algorithm allows us to detect the exact value of the number change-points m_0. Importantly, Part (4) of Theorem 3.4.1 shows that our algorithm produces an estimator $\hat{m}$ which converges in probability to the exact value of the number of the change-points m_0.

3.4.2 Asymptotic results of estimators with random dimensions

In this subsection, we present a probabilistic result which allows us to overcome the problem related to the fact that when m is replaced by an estimator, the dimensions of the tensor estimators become random variables. Indeed, the dimensions of $\hat{\boldsymbol{\delta}}(\boldsymbol{\tau})$, $\tilde{\boldsymbol{\delta}}(\boldsymbol{\tau})$, $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}})$ and $\tilde{\boldsymbol{\delta}}(\tilde{\boldsymbol{\tau}})$, are functions of m and because of that, let $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}, m)$ denote the $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}})$ and let $\tilde{\boldsymbol{\delta}}(\tilde{\boldsymbol{\tau}}, m)$ denote $\tilde{\boldsymbol{\delta}}(\tilde{\boldsymbol{\tau}})$. Further, let $\hat{m}$ be a consistent estimator for m_0 and let $\hat{\boldsymbol{\tau}}(\hat{m})$ be the estimator of $\boldsymbol{\tau}(m)$. For the sake of simplicity, we denote $\hat{\boldsymbol{\tau}}$ and $\tilde{\boldsymbol{\tau}}$ to stand for $\hat{\boldsymbol{\tau}}(\hat{m})$ and $\tilde{\boldsymbol{\tau}}(\hat{m})$, respectively. The UE and RE are obtained as in Section 2 by replacing m with $\hat{m}$ for $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}, m)$ and $\tilde{\boldsymbol{\delta}}(\tilde{\boldsymbol{\tau}}, m)$. As such, the UE and RE become $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}, \hat{m})$ and $\tilde{\boldsymbol{\delta}}(\tilde{\boldsymbol{\tau}}, \hat{m})$, respectively. It is important to note that as the dimensions of $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}, \hat{m})$ and $\tilde{\boldsymbol{\delta}}(\tilde{\boldsymbol{\tau}}, \hat{m})$ are functions of $\hat{m}$, it is not possible to derive the limiting distribution of $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}, \hat{m})$ and $\tilde{\boldsymbol{\delta}}(\tilde{\boldsymbol{\tau}}, \hat{m})$. Due to that fact, neither the relative risk dominance of the UE and the RE nor the construction of shrinkage estimators follow from the results in literature as, for example, in Saleh (2006), and Chen and Nkurunziza (2016). To overcome the random dimension problem, we use the following result. Let $\rho_T(\hat{\tau}, \hat{m}) = \sqrt{T}(\hat{\boldsymbol{\delta}}(\hat{\tau}, \hat{m}) - \boldsymbol{\delta})\mathbb{I}_{\{\hat{m}=m\}}$, $\zeta_T(\tilde{\tau}, \hat{m}) = \sqrt{T}(\tilde{\boldsymbol{\delta}}(\tilde{\tau}, \hat{m}) - \boldsymbol{\delta})\mathbb{I}_{\{\hat{m}=m\}}$ and $\xi_T(\hat{\tau}, \hat{m}) = \sqrt{T}(\hat{\boldsymbol{\delta}}(\hat{\tau}, \hat{m}) - \tilde{\boldsymbol{\delta}}(\tilde{\tau}, \hat{m}))$.

Theorem 3.4.2. *Let* $g : \mathbb{R}^{q_1\times\cdots\times q_d\times(m+1)q_{d+1}} \times \mathbb{R}^{q_1\times\cdots\times q_d\times(m+1)q_{d+1}} \times \mathbb{R}^{q_1\times\cdots\times q_d\times(m+1)q_{d+1}} \longrightarrow \mathbb{R}^{a_1\times\cdots\times a_{d+1}}$ *be a continuous function with* a_i*,* $i = 1, \cdots, d+1$ *independent of m, and suppose that Assumption 3.2.3 holds along with* (3.12)*. Then,*

1. *if* $\boldsymbol{r}_0 \neq 0$*,* $g(\epsilon^*_{1,T}(\hat{\tau}, \hat{m}), \epsilon^*_{2,T}(\tilde{\tau}, \hat{m}), \epsilon^*_{3,T}(\hat{\tau}, \hat{m}))) \xrightarrow[T\to\infty]{d} g(\epsilon^*_1, \epsilon^*_2, \epsilon^*_3)$;
2. *if* $\boldsymbol{r}_0 = 0$*,* $g(\epsilon^*_{1,T}(\hat{\tau}, \hat{m}), \epsilon^*_{2,T}(\tilde{\tau}, \hat{m}), \epsilon^*_{3,T}(\hat{\tau}, \hat{m}))) \xrightarrow[T\to\infty]{d} g(\epsilon^*_{10}, \epsilon^*_{20}, \epsilon^*_{30})$*, where*

$$\epsilon^*_{10} \boxplus_{(d+1)} \epsilon^*_{20} \boxplus_{(d+1)} \epsilon^*_{30} \sim \mathcal{N}_{q_1\times\cdots\times q_d\times 3(m+1)q_{d+1}}\left(0, \begin{pmatrix} \Sigma^*_{11} & \Sigma^*_{12} & \Sigma^*_{13} \\ \Sigma^*_{21} & \Sigma^*_{22} & \Sigma^*_{23} \\ \Sigma^*_{31} & \Sigma^*_{32} & \Sigma^*_{33} \end{pmatrix}\right).$$

Proof. The proof follows by using vec operator along with Lemma 5.1 of Nkurunziza et al. (2019) and Proposition 3.2.18. □

Remark 3.4.1. *Recall that the optimality of the proposed estimators, which is established in Corollary 3.3.3, heavily relies on Proposition 3.2.18. However, in this context, the dimension of the random tensors* $\rho_T(\hat{\tau}, \hat{m})$*,* $\zeta_T(\tilde{\tau}, \hat{m})$ *and* $\xi_T(\hat{\tau}, \hat{m})$ *are random variables. The notion of asymptotic distribution of* $\rho_T(\hat{\tau}, \hat{m}) \boxplus_{(d+1)} \zeta_T(\tilde{\tau}, \hat{m}) \boxplus_{(d+1)} \xi_T(\hat{\tau}, \hat{m})$ *does not make any sense here. Nevertheless, Theorem 3.4.2 tells us that, since* $g(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \mathrm{trace}\left(\boldsymbol{x}^{*'}_{(d)}\boldsymbol{x}^*_{(d)}\right)$ *is a real-valued function, we can still use the distribution of* $\epsilon^*_1 \boxplus_{(d+1)} \epsilon^*_2 \boxplus_{(d+1)} \epsilon^*_3$ *given in Proposition 3.2.18 in order to get the ADR of the proposed estimators.*

3.5 Simulation studies and illustrative examples

3.5.1 Simulation studies

In this subsection, we present some simulation results that illustrate the performance of the proposed method. We carry out the simulations for the case where there is no change-point (i.e. $m_0 = 0$) as well as for the case where $m_0 = 1$, $m_0 = 2$ and $m_0 = 3$. Nevertheless, to save the space of this paper, we only report the results for the cases where $m_0 = 0$ and $m_0 = 1$. Subsequently, we also present results in the case where $m_0 = 2$ is unknown and must be estimated.

First case: First, consider the case where there is no change-point and set $m_0 = 0$. Consider the estimation problem of two dimensional 64×64 signals image of a square, a circle, a triangle and a $\top$ similar to that in Figure 3.1 and set the responses to also be two dimensional matrices.

Figure 3.1: Signal images used for parameter estimation.

Set $d = 2$, $q_1 = q_2 = 64$, $q_3 = 1$ and let $\delta = \mathbb{B}_1$ be $64 \times 64 \times 1$ (which is equivalent to 64×64 matrix) and set the number of observations/responses to be $T = 20$. We let covariates z_i to be scalars randomly drawn from a uniform distribution on the interval (0,1.5) and the resulting matrix of covariates becomes $\bar{Z} = (z_1, \cdots, z_{20})'$. The error terms $\mathbb{U}_i$ are 64×64 randomly drawn from a normal distribution with mean 0 and variance 1 and the resulting stacked error term along the 3rd dimension, $\mathbb{U}$, is a $64 \times 64 \times 20$ dimensional tensor in which the i^{th} face corresponds to the i^{th} error term of the i^{th} 64×64 response matrix. Stacking along the 3^{rd} dimension gives the response tensor $\mathbb{Y}$ which is $64 \times 64 \times 20$ with $\mathbb{Y} = \delta \times_3 \bar{Z} + \mathbb{U}$. To set up the restriction, let $l_1 = 20$, $l_2 = 64$, $R_1 = \left[I_{l_1}, 0_{l_1 \times (q_1 - l_1)} \right]$, $R_2 = I_{l_2}$, and let r_0 be an $l_1 \times l_2 \times 1$ tensor of zeros. Then, we compute the UE, the RE and the SEs as in this chapter.

We have an idea that the signal is mainly concentrated in the middle of the image and the surrounding voxels are empty signals. Thus, we know that along the first dimension (i.e. down the rows of the parameter matrix), the first, say, 20 rows are 0. Hence, we set the mode-1 restriction, R_1, to be as previously defined. Moreover, such an empty signal would be spread throughout the columns from 1 to 64, hence, we set the mode-2 restriction to be the 64×64 identity matrix. r_0 is set to be 20×64 zero matrix to support the hypothesis that the first 20 rows and 64 columns of our matrix parameter display an empty signal.

We repeated this simulation for several different image signals and the results are displayed in Figure 3.2. This shows that the UE for each signal parameter mostly displayed the true signal in the centre, however, it failed to display any information about the area surrounding the image. However, the RE has offered more information as the top portion of the signal is uniform and gives an idea that there may not be any signal for the top portion of this parameter. The SEs, although not as clearly uniform as the RE, are far less grainy and display more consistency in colouring than the image of the UE. As such, it can be seen that with some prior knowledge about an image, the RE outperforms the UE and the SEs are not too far behind.

In addition to the findings illustrated by Figure 3.2 we study the efficiency of the proposed estimators by comparing their relative mean square efficiencies (RMSE) with respect to $\hat{\boldsymbol{\delta}}$. The RMSE is defined as $\text{RMSE}(\hat{\boldsymbol{\delta}}^*) = \text{ADR}(\hat{\boldsymbol{\delta}})/\text{ADR}(\hat{\boldsymbol{\delta}}^*)$, where $\hat{\boldsymbol{\delta}}^*$ is a proposed estimator of $\boldsymbol{\delta}$. All parameters, dimensions and initial restrictions are as discussed in the above simulation. We run the simulation with r deviating from 0 by units of $1/\sqrt{T}$ at each iteration, i.e. $\boldsymbol{r} = \boldsymbol{r}_0 + \left(\frac{\Delta}{\sqrt{T}}\right)\mathbb{E}$, where $\Delta = 0, 1, \cdots, 6$ and $\mathbb{E}$ is $20 \times 64 \times 1$ tensor of ones.

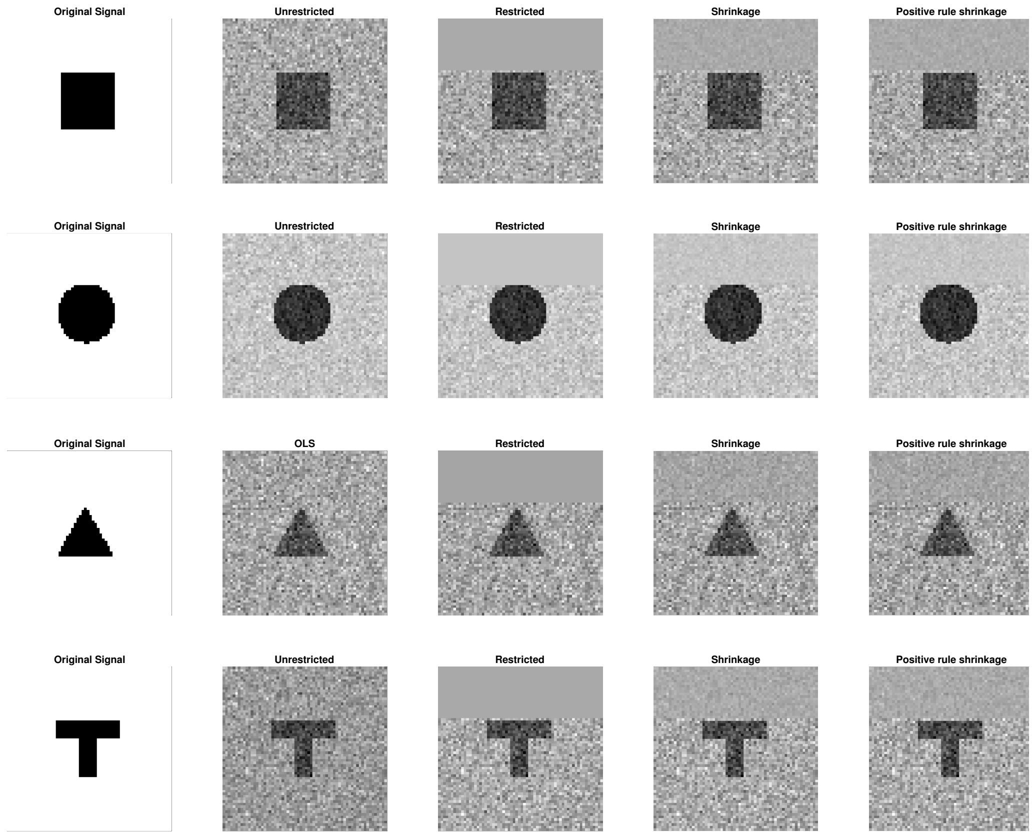


Figure 3.2: Comparison of several signal images with their respective estimators.

For each restriction, we compute the UE, the RE, the SEs and RMSE of each estimator with respect to the UE. For each deviation, Δ, we replicate the simulation 1000 times and obtain the average RMSE for each estimator. The results of this simulation are displayed in Figure 3.3. This shows that, in the neighbourhood of the restriction, the RE dominates all estimators while if fails from around $\Delta = 1.5$. However, the SEs continue to be more efficient than the UE as we deviate away from a true constraint. This corroborates with the theoretical results given in Corollary 3.3.2 and Corollary 3.3.3.

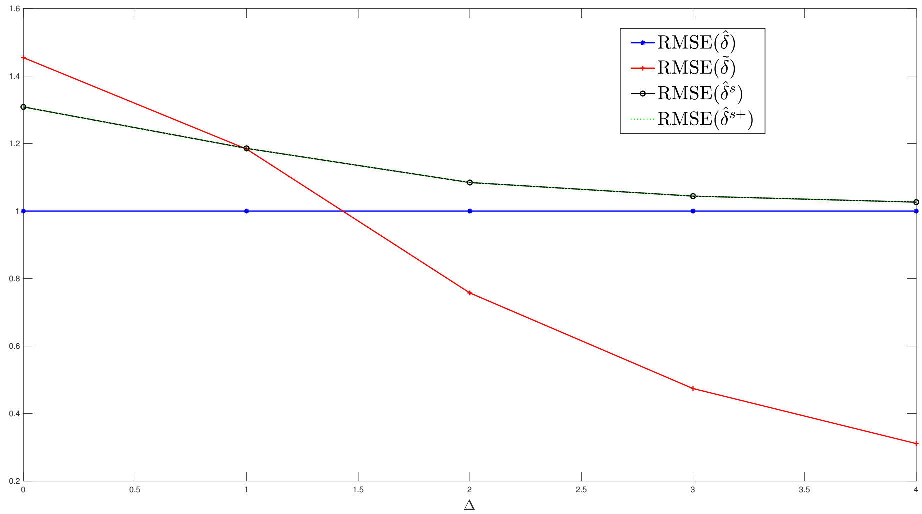
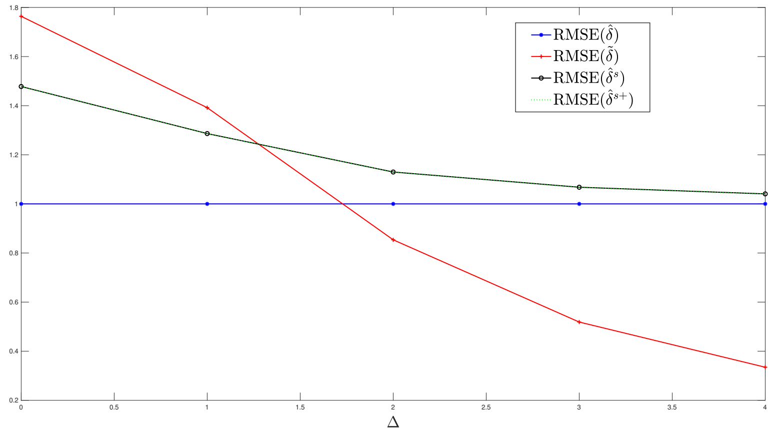


Figure 3.3: The RMSE versus Δ plot of the four estimators of the square signal parameter.

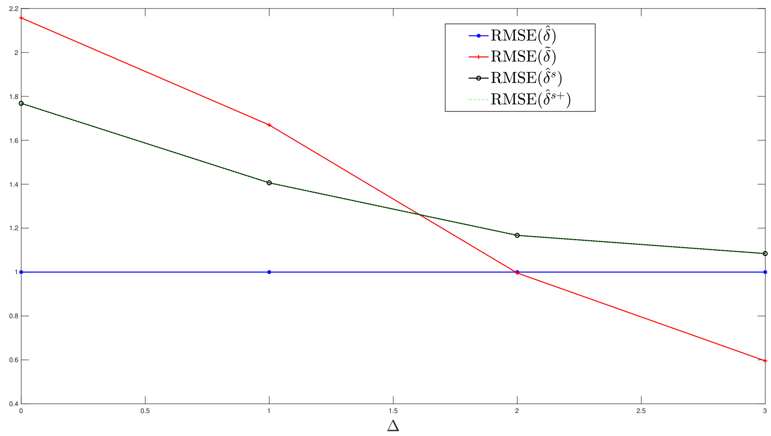
Second case: We also perform the simulations for the cases where $m_0 = 1$. For $T = 80$ and $T = 150$, we consider that the first 30 observations depend on the 2-dimensional square signal parameter and the remaining $T - 30$ depend on the 2-dimensional triangle signal parameter. Further, as in the previous simulations, we set $d = 2$, $q_1 = q_2 = 64$, $q_3 = 1$. The parameter $\mathbb{B}_1$ is set as the 2-dimensional square signal and $\mathbb{B}_2$ is set as the 2-dimensional triangle signal and $\delta = \mathbb{B}_1 \boxplus_3 \mathbb{B}_2$ is the $64 \times 64 \times 2$ model parameter. Note that the two images compose the two faces of δ, which is three dimensional. Figure 3.4 illustrates the two signal image parameters. We generate the covariates z_i as in the first case and the resulting matrix of covariates becomes $\bar{Z} = \left((z_1, \cdots, z_{30}, \bar{0}'_{T-30})\ (\bar{0}'_{30}, z_{31}, \cdots, z_T)\right)'$, which is a $T \times 2$ block diagonal matrix, where $\bar{0}_n$ is the vector of n zeros. The error, $\mathbb{U}$, is a $64 \times 64 \times T$-tensor randomly drawn as above. We set R_1 and R_2 as in the first case and we set $R_3 = I_2$, to represent that the mode-1 and mode-2 restrictions apply for both faces of the 3-dimensional image parameter, δ. We set r_0 to be an $20 \times 64 \times 2$ tensor of zeros. After the initializations, we assume that the location of the change-points is unknown and we run the dynamic programming algorithm to obtain $\hat{\tau}$ and $\tilde{\tau}$. Using $\hat{\tau}$ and $\tilde{\tau}$, we then build the corresponding covariate matrix, $\bar{Z}$, and obtain the UE and RE as in Proposition 3.1.1. We build the shrinkage estimators from the UE and RE as defined in Section 3.3. As in the first case, we compute the RMSE for the UE, the RE and the SEs and the obtained results are displayed in Figure 3.5. Once again, this shows that in the neighbourhood of the hypothesized restriction, the RE dominates all estimators while it performs poorly as one moves far away from the restriction. Further, the SEs continue to dominate the UE even when the restriction fails. This visual protrait further corroborates the theoretical results of Corollary 3.3.2 and Corollary 3.3.3.

$\mathbb{B}_1$ $\mathbb{B}_2$

Figure 3.4: The two signal image parameters for the case where $m_0 = 1$.



(a) The case where $m_0 = 1$ and $T = 80$



(b) The case where $m_0 = 1$ and $T = 150$

Figure 3.5: The RMSE versus Δ plot of the UE, the RE and SEs

Third case: To illustrate the model and estimation in the case of unknown number of change-points, we take the case where $m_0 = 2$ and we set $T = 80$ observations where the first 32 depend on the 2-dimensional square signal parameter, the next 30 depend on the 2-dimensional triangle signal parameter, and the remaining depend on an irregular centred image. Then, as in the previous simulations, we set $d = 2$, $q_1 = q_2 = 64$, $q_3 = 1$. $\mathbb{B}_1$ is set as the 2-dimensional square signal, $\mathbb{B}_2$ is set as the 2-dimensional triangle signal, $\mathbb{B}_3$ is set as a 2-dimensional irregular shape signal and $\delta = \mathbb{B}_1 \boxplus_{(3)} \mathbb{B}_2 \boxplus_{(3)} \mathbb{B}_3$ is the $64 \times 64 \times 3$ model parameter. Note, that in this simulation the three images compose the three faces of δ, which is three dimensional. Once again, we let covariates z_i be scalars randomly drawn from a uniform distribution on the interval (0,1.5) and the resulting matrix of covariates becomes $\bar{Z} = \left((z_1, \cdots, z_{30}, \bar{0}'_{50})\ (\bar{0}'_{30}, z_{31}, \cdots, z_{62}, \bar{0}'_{18})\ (\bar{0}'_{62}, z_{63}, \cdots, z_{80})\right)'$, which is an 80×3 block diagonal matrix. The error tensor, $\mathbb{U}$, is randomly drawn from random normal with mean 1 and variance 1.5 and has dimensions $64 \times 64 \times 80$. We set the maximum number of change-points to be 5. For each number of change-points, we run the dynamic programming algorithm and store the unrestricted SSR and the estimated change-point locations. The program then selects the number of change-points that gives the minimum value of the penalty function in (3.15) . We set that number to be $\hat{m}$, the estimated number of change-points. After we find $\hat{m}$, we set R_1 and R_2 as in the previous simulations and we set $R_3 = I_3$, to represent that the mode-1 and mode-2 restrictions apply for all three faces of our 3-dimensional image parameter, $\boldsymbol{\delta}$. We set $\boldsymbol{r}_0$ to be an $20 \times 64 \times 3$ tensor of zeros. We once again deviate away from the initial restriction by units of $1/\sqrt{80}$ at each iteration for $\Delta = 0, 1, 2, 3$. We replicate the simulation 1000 times and obtain the average RMSE for the UE, RE and SEs. We repeated the simulation for the case when $T = 200$. The simulation results are displayed in Figure 3.6 and Figure 3.7. As in the first two cases, it can be seen by the plots in Figure 3.6 and Figure 3.7 that in the neighbourhood of the restriction, the RE once again outperforms the other estimators but performs poorly the farther away we move from the restriction. In addition, the shrinkage estimators continue to dominate the UE even when the restriction fails. This further corroborates our theory.

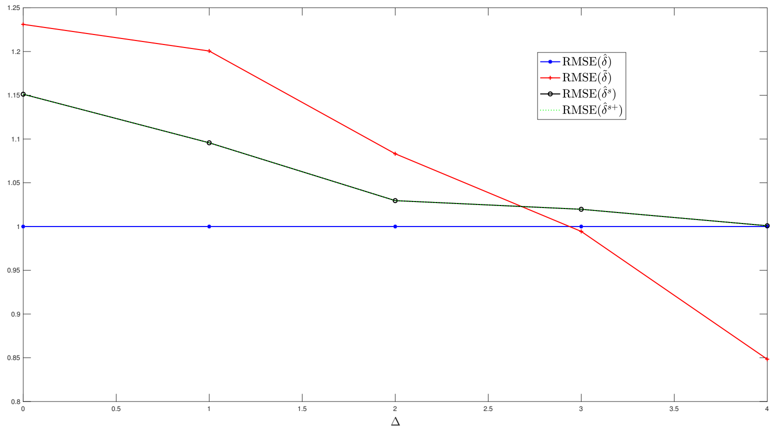


Figure 3.6: The RMSE versus Δ plot of the four estimators in the case where $m = 2$ is unknown and $T = 80$.

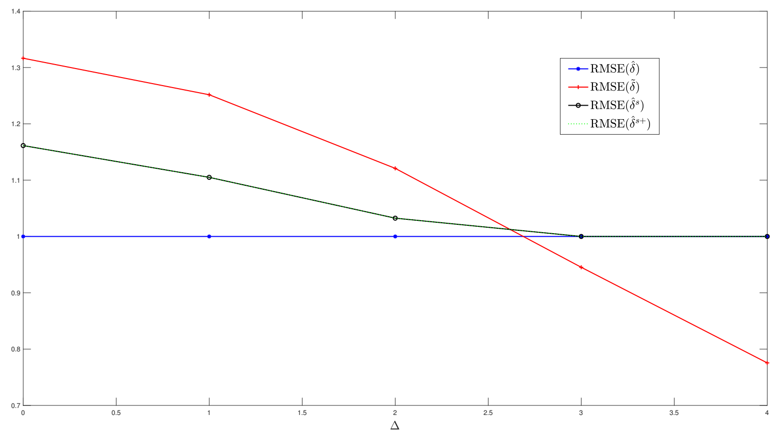


Figure 3.7: The RMSE versus Δ plot of the four estimators in the case where $m = 2$ is unknown and $T = 200$.

3.5.2 Real data analysis

In this subsection, we summarize some real neuro-imaging data analysis results. In particular, we illustrate the application of the proposed method to MRI imaging data and we show that the SEs as well as the RE perform slightly better than the UE. We also conduct a change-point detection analysis on real neuro-imaging time series.

MRI dataset

In this subsection, we show results of an analysis on a real neuro-imaging dataset in order to illustrate the application of our methods. For the analysed dataset, the attention deficit hyperactivity disorder (ADHD) data was acquired from the ADHD-200 Sample Initiative which consists of 775 subjects consisting of 491 normal controls and 285 combined ADHD subjects. We obtained the preprocessed anatomical (MRI) data provided by the Burner pipeline (Bellec et al. (2017)). We removed 7 images due to poor quality or missing data. For each subject, the MRI data is a $197 \times 233 \times 189$ tensor, which will be taken as our observed three-dimensional responses. To facilitate the analysis, we down-sized the data and the resulting dimensions of each observation $\mathbb{Y}_i$ are $30 \times 36 \times 30$.

From previous literature such as in Solanto et al. (2009), Wolf et al. (2009) and Yu-Feng et al. (2007), it was suggested that ADHD is associated with an abnormality of the brain region known as the fusiform gyrus. Specifically, it was found that for those diagnosed with ADHD, the fusiform gyrus was much darker when compared with the control indicating that those regions are inactive. As such, using this prior knowledge and to illustrate how restrictions can be used for more efficient estimation, we set R_1 to be a 10×30 matrix where the first 5 rows and columns 8 to 12 are set to be the identity matrix. In addition, the part of R_1 corresponding to rows 6 to 10 and columns 20 to 24 were also set to the identity matrix. R_2 is set to be 5×36 matrix where the sub-matrix consisting of rows 1 to 5 and columns 16 to 20 is set to also be the identity matrix of size 5. R_3 is 5×30 matrix with sub-matrix consisting of rows 1 to 5 and columns 6 to 10 are also set to be the identity matrix of size 5. All other elements of R_1, R_2 and R_3 were set as 0. We let $R_4 = [0, 0, 1]$ and $\boldsymbol{r}_0$ is a $10 \times 5 \times 5$

zero tensor. To provide some practical interpretation of these matrices, note that R_1, R_2 and R_3 select the region of interest in the brain and R_4 selects those diagnosed with ADHD. A visual representation of this restriction can be found below in Figure 4.4.

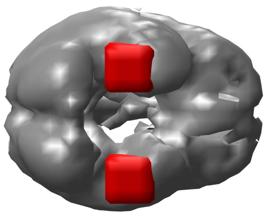
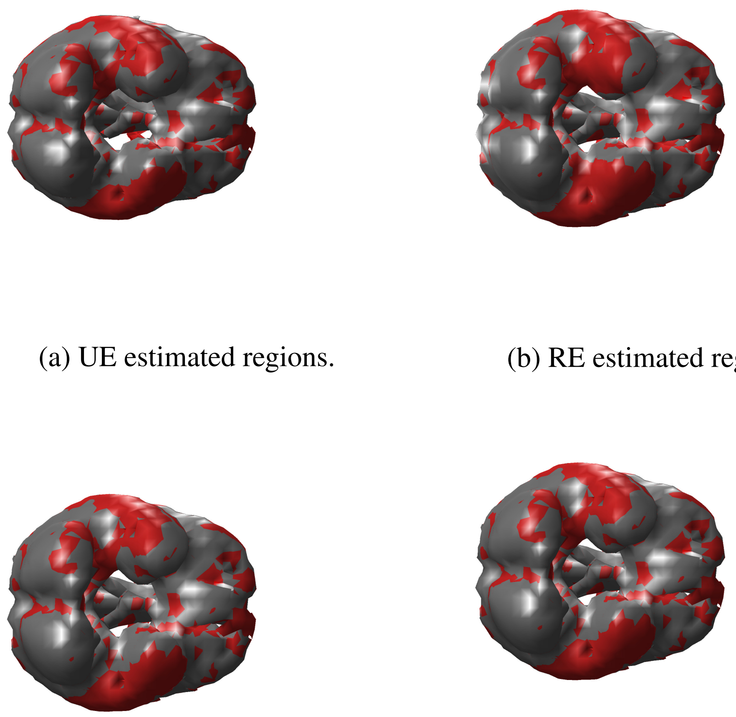


Figure 3.8: A visual representation of the restriction (red). This is an approximate location of part of the fusiform gyrus.

Further, we included the age, gender and ADHD diagnosis as covariates and obtained the estimators for $\mathbb{B}$ and the RMSE by performing a bootstrap with 1000 replications. Figure 4.5 displays the estimated regions overlaid on a randomly selected subject. As can be seen in Figure 4.5, there are some subtle differences in the estimated regions associated with each estimator. We bootstrapped residuals to produce replicates. The RMSE of the restricted estimator was found to be 1.0036, the RMSE of the James-Stein shrinkage estimator was 1.0003 and the RMSE of Positive James-Stein shrinkage estimator was 1.00035. Thus, as the RMSEs of the SEs are larger than 1, we conclude that the SEs are more efficient than the UE. This confirms that, even when working with real neuro-imaging data, the SEs perform better than the UE. In addition, as the RMSE of the RE is also slightly larger than 1, we conclude that the RE is more efficient than the UE. In other words, an abnormality/inactivation of the fusiform gyrus may exist for those with ADHD.



(a) UE estimated regions.

(b) RE estimated regions.

(c) James-Stein estimated regions.

(d) Positive James-Stein estimated regions.

Figure 3.9: Estimated regions (red) that may be associated with ADHD overlaid on a randomly-drawn subject (grey).

fMRI dataset

In this subsection, we use real neuro-imaging data to detect the existence of change-points. In particular, we analyze the non-stationary properties of resting state functional magnetic resonance imaging or fMRI. This resting state scan requires an individual to remain entirely still with eyes closed and to keep any thought process blank for a period of time. These scans are useful in pinpointing the brain regions involved in underlying brain activity and are sometimes known as the default network. During this period of time, three-dimensional images of the brain are taken at each time point, resulting in a four-dimensional time-series of brain scans of each individual. The stationarity of this time-series is often a crucial assumption that must be checked to ensure that the brain results are not due to an external stimulus such as an individual's thought process or a loud, unexpected sound occurring during the scanning session. As in Aston and Kirch (2012), we use the 1000 connectome resting state functional magnetic resonance imaging or fMRI Biswal et al. (2010). We use the Beijing scan site data which consist of 198 resting state fMRI scans, where three-dimensional images of size $64 \times 64 \times 33$ voxels are taken over 225 time points and every pair of consecutive time points are 2 seconds apart. We also included the age and gender as covariates. For each subject, we checked if a change-point could be detected. By using the proposed method, the results reported in Table 3.1 show that, for each subject, a change-point was detected at a certain time-point. As can be seen in Table 3.1, a change-point has been detected for the subjects and the estimated location of that change-point was also found.

Subject Number	$\hat{m}$	$\hat{\tau}$
00440	1	71
10973	1	46
17315	1	44
48501	1	76
49782	1	50
01018	1	40
69518	1	123

Table 3.1: Results of fMRI data analysis for several subjects. Each subject was found to have one non-stationarity ($\hat{m}$) in their resting state scan and the time-point at which the non-stationarity occurred was recorded as $\hat{\tau}$.

3.6 Conclusion

In this chapter, we studied an estimation problem about the tensor coefficient in a tensor regression model with multiple and unknown change-points in the context where the tensor parameter is suspected to satisfy some restrictions. We introduced the proposed tensor coefficient estimators including the UE, RE and the James-Stein and Positive-rule Stein estimators. Under the $\mathcal{L}^2-$mixingale assumptions, we derived the joint asymptotic distribution of the UE and RE which is the tensor generalization of Lemma 3.4 of Chen and Nkurunziza (2016). In addition, we defined the asymptotic distributional risk under the quadratic loss function and derived the ADR[1] for the class of shrinkage estimators as well as the proposed estimators. Using the results, we established some sufficient conditions for the SEs and RE to outperform the UE. We also proposed methods to consistently estimate the number of change-points in the case where m_0 is unknown. To demonstrate the applications of the restrictions and the overall proposed methods, we corroborated the results with some simulation studies and we also analyzed MRI and fMRI datasets.

Chapter 4

Generalized Tensor Regression

In this chapter, we study estimation methods for a generalized tensor regression model. Unlike Chapter 3 where an identity link function on the tensor linear predictors was used, the model in this chapter extends results of multivariate linear regression to a tensor generalized model with a known link function. As we are interested in developing estimation methods in the context of a generalized model, in this chapter we do not consider multiple change-points but we provide assumptions that are weaker than the conditions of Assumption 3.2.3 of Chapter 3. Moreover, we study the asymptotic distributional risks under a general constraint and a general loss function that includes the quadratic loss and restriction of Chapter 3 as a special case.

In this chapter, our methods differ from recent works in several ways. First, while some references such as Zhou et al. (2013), Li and Zhang (2017), Raskutti et al. (2015) and Hoff (2015) assume independent and/or identically distributed errors, we consider a generalized tensor model with a link function that includes the Gaussian assumption within its framework. In addition, the model takes into account matrix regressors on multiple-modes allowing for more complex interactions and connections among covariates. Incorporating multi-mode covariates would include many other different types of models such as spatiotemporal growth models, network population models and dyadic data with node attributes. See Xu et al. (2019) for more details on this topic. Second, under some general assumptions, we weaken the dependence structure of the error terms of the model to be as weak as that of an $\mathcal{L}^2-$ mixingale. Third, we introduce a general restriction upon the tensor parameter and derive the unrestricted and restricted estimators by minimizing a quasi-score function. While some of the quoted papers, such as Zhou et al. (2013) and Hoff (2015), have presented a penalty function as a form of model regularization and dimension reduction, our work differs in that the restriction can implement prior knowledge for parameter reduction or can even be used for statistical inference. For example, when working with three-dimensional neuro-imaging data, a previous study may have suggested that a certain region of the brain may/may not have an effect on the diagnosis of disease of interest. The restrictions can be chosen such that the first three restriction matrices would select the region and the fourth would select the covariate of interest or even all covariates. Fourth, we propose a class of shrinkage estimators which, to the best of our knowledge, has not been investigated in the context of generalized tensor regression models with tensor observations and matrix regressors.

The remainder of this chapter is organized as follows. In Section 4.1, we present the general tensor regression model and constraint on the tensor parameter and estimation method. In Section 4.2, we establish some asymptotic results of the estimators that are derived from assumptions intended to weaken the dependence structure of the tensor error terms. In Section 4.3, we present a class of tensor shrinkage estimators and the asymptotic distributional risk. We present some conditions for which some shrinkage estimators dominate the unrestricted estimator. Section 4.4 summarizes some simulation studies and application of two real datasets.

4.1 The generalized tensor model and estimation

In this section, we present the generalized tensor regression model and a general constraint to be applied to the tensor parameter. We then derive an estimating score function for the tensor parameter. The obtained estimating score function is useful in deriving some asymptotic results.

4.1.1 The generalized tensor regression model and constraints

We consider the following model with n observations and link function $f(\cdot)$. We assume that $f(\cdot)$ is a $d-$dimensional, twice continuously differentiable function that maps elements component-wise. Let

$$\mathbb{Y}_i = f(\Theta_i) + \mathbb{U}_i, \quad i = 1, \cdots, n, \tag{4.1}$$

where $\mathbb{Y}_i \in \mathbb{R}^{q_1 \times q_2 \times \cdots \times q_d}$, $\Theta_i = \mathbb{B}(\bigtimes_{j=1}^{d})_j \boldsymbol{X}_{ij}$, $\mathbb{B} \in \mathbb{R}^{p_1 \times p_2 \times \cdots \times p_d}$, $\boldsymbol{X}_{ij} \in \mathbb{R}^{q_j \times p_j}$, $j = 1, 2, \cdots, d$. Define $\mathbb{Y} = \mathbb{Y}_1 \boxplus_{(d+1)} \mathbb{Y}_2 \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \mathbb{Y}_n$, $\mathbb{U} = \mathbb{U}_1 \boxplus_{(d+1)} \mathbb{U}_2 \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \mathbb{U}_n$ and $\mu_i = \mathrm{E}\left(\mathbb{Y}_i \middle| \boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots, \boldsymbol{X}_{id}\right) = f(\Theta_i)$, $i = 1, 2, \ldots, n$ and let $\boldsymbol{\mu} = f(\Theta_1) \boxplus_{(d+1)} f(\Theta_2) \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} f(\Theta_n)$. We also define $\boldsymbol{X} = \left[\bigotimes_{j=d}^{1} \boldsymbol{X}'_{1j}, \cdots, \bigotimes_{j=d}^{1} \boldsymbol{X}'_{nj}\right]'$. Thus, the tensor regression model in (4.1) becomes

$$\mathbb{Y} = \boldsymbol{\mu} + \mathbb{U} \tag{4.2}$$

where the random tensor $\mathbb{Y}$ represents the stacked response and $\boldsymbol{X}$ represents the regressors while $\mathbb{U}$ represents the tensor of stacked error terms. We suppose that $\mathrm{E}(\mathbb{U}|\boldsymbol{X}) = \mathbf{0}$ with $\mathrm{E}\left(\|\mathbb{U}\|^2\right) < \infty$ and $\mathrm{E}\left(\|\boldsymbol{X}\|^2\right) < \infty$. Here, the error term and the regressors do not need to be independent and we do not assume any specific distribution for the error term or the regressors. Further, as given in the next section, the components of the error term may be correlated.

Remark 4.1.1. *The model in Equation* (4.1) *is the general form of the model in Equation* (3.1) *for the case where* $m_0 = 0$*. Indeed, by setting the link function* f *as* $f(x) = x$ *and taking* $p_j = q_j,\ j = 1, \cdots, d$; $p_{d+1} = q_{d+1} = 1$; $n = T$; *and* $\boldsymbol{X}_{ij} = I_{q_j},\ j = 1, \cdots, d$ *and* $\boldsymbol{X}_{id+1} = z_i \in \mathbb{R},\ i = 1, \cdots, n$, *we get the model in Equation* (3.1)*.*

Additionally, we consider the scenario where uncertain prior information about the target parameter exists. Thus, we establish a statistical method which combines the imprecise prior knowledge and the sample information from the model in (4.2). In particular, we consider the case where the prior knowledge is in the form of some constraints on the tensor parameter $\mathbb{B}$. Namely, we consider the following general constraint on the tensor parameter

$$\hbar^*(\mathbb{B}) = 0, \tag{4.3}$$

where $\hbar^*(\mathbb{B})$ is an $l_1 \times l_2 \times \cdots \times l_d$ function. To guarantee the existence and the consistency of the restricted estimator (RE) of the tensor parameter $\mathbb{B}$, we assume that the following assumption holds.

Assumption 4.1.1. 1. *In a neighbourhood of* $\mathbb{B}^*$*, the function* $\hbar^*(\mathbb{B})$ *is a twice continuously differentiable function. Let* $\hbar(\mathbb{B}) = \mathrm{Vec}(\hbar^*(\mathbb{B}))$*;*
2. *Let* $H(\mathbb{B}) = \frac{\partial \hbar(\mathbb{B})}{\partial \mathrm{Vec}(\mathbb{B})}$*.* $H(\mathbb{B})$ *is* $l_1 l_2 \cdots l_d \times p_1 p_2 \cdots p_d$ *matrix with full column rank. Further, in a neighbourhood of* $\mathbb{B}^*$*,* $\frac{\partial H(\mathbb{B})}{\partial (\mathrm{Vec}(\mathbb{B}))'}$ *exists and is bounded by a constant and integrable function.*

Remark 4.1.2. *Note that the constraint in* (4.3) *can be set as the restriction considered in* (3.2)*. Specifically, we can set* $\hbar^*$ *as*

$$\hbar^*(\mathbb{B}^*) = \mathbb{B}^* \times_1 \boldsymbol{R}_1 \times_2 \boldsymbol{R}_2 \times \cdots \times_d \boldsymbol{R}_d - \boldsymbol{r}, \tag{4.4}$$

where $\boldsymbol{R}_i$ *are* $l_i \times p_i$*, non-random matrices with rank* $l_i \leqslant p_i$*,* $i = 1, 2, \cdots, d$ *and* $\boldsymbol{r}$ *is* $l_1 \times l_2 \times \cdots \times l_d$ *non-random tensors.*

4.1.2 Estimating score function in a generalized tensor regression model

In this subsection, we present an estimating function for the model in (4.1). This estimating function is useful in the derivation of the unrestricted estimator (UE) and, under a general constraint in (4.3), it is useful in establishing the restricted estimator (RE). Based on the least square optimization function, the estimating function is defined as

$$\mathcal{G}_n(\mathbb{B}) = \left[\frac{\partial \mathrm{Vec}(\boldsymbol{\mu})}{\partial \mathrm{Vec}(\mathbb{B})'}\right]' \mathrm{Vec}(\mathbb{Y} - \boldsymbol{\mu}) = \left[\frac{\partial \mathrm{Vec}(\boldsymbol{\mu})'}{\partial \mathrm{Vec}(\mathbb{B})}\right] \mathrm{Vec}(\mathbb{Y} - \boldsymbol{\mu}) \tag{4.5}$$

where $\boldsymbol{\mu}$ and $\mathbb{Y}$ are given by (4.2). We break down each of these components: $\mathrm{Vec}(\boldsymbol{\mu}) = (\mathrm{Vec}(f(\Theta_1))',\ \mathrm{Vec}(f(\Theta_2))',\ \cdots,\ \mathrm{Vec}(f(\Theta_n))')'$ and the first derivative with respect to $\mathrm{Vec}(\mathbb{B})'$ is

$$\frac{\partial \mathrm{Vec}(\boldsymbol{\mu})'}{\partial \mathrm{Vec}(\mathbb{B})} = \left(\left(\frac{\partial \mathrm{Vec}(f(\Theta_1))'}{\partial \mathrm{Vec}(\mathbb{B})}\right), \left(\frac{\partial \mathrm{Vec}(f(\Theta_2))'}{\partial \mathrm{Vec}(\mathbb{B})}\right), \cdots, \left(\frac{\partial \mathrm{Vec}(f(\Theta_n))'}{\partial \mathrm{Vec}(\mathbb{B})}\right)\right). \tag{4.6}$$

Using the chain rule, we get

$$\frac{\partial \mathrm{Vec}(\boldsymbol{\mu})'}{\partial \mathrm{Vec}(\mathbb{B})} = \left(\left(\frac{\partial \mathrm{Vec}(f(\Theta_1))}{\partial \mathrm{Vec}(\Theta_1)} \bigotimes_{j=d}^{1} X_{1j}\right)', \left(\frac{\partial \mathrm{Vec}(f(\Theta_2))}{\partial \mathrm{Vec}(\Theta_2)} \bigotimes_{j=d}^{1} X_{2j}\right)', \cdots, \left(\frac{\partial \mathrm{Vec}(f(\Theta_n))}{\partial \mathrm{Vec}(\Theta_n)} \bigotimes_{j=d}^{1} X_{nj}\right)'\right)'.$$

Let $\mathcal{D}_n = \frac{\partial \mathrm{Vec}(\boldsymbol{\mu})}{\partial \mathrm{Vec}(\Theta)'}$, where $\Theta = (\Theta_1, \cdots, \Theta_n)$. As a result, the estimating function in (4.5) becomes

$$\mathcal{G}_n(\mathbb{B}) = \sum_{i=1}^{n} \left[\left(\bigotimes_{j=d}^{1} \boldsymbol{X}'_{ij}\right)\left(\frac{\partial \mathrm{Vec}(f(\Theta_i))}{\partial \mathrm{Vec}(\Theta_i)}\right)' \mathrm{Vec}(\mathbb{U}_i)\right] = \boldsymbol{X}'\mathcal{D}'_n \mathrm{Vec}(\mathbb{Y} - \boldsymbol{\mu}). \tag{4.7}$$

The (unrestricted) score estimating equation for estimating the tensor parameter $\mathbb{B}$ is given by

$$\mathcal{G}_n(\mathbb{B}) = 0. \tag{4.8}$$

Let $\hat{\mathbb{B}}$ be the unique root of the estimating equation (4.8). This is known as the unrestricted estimator (UE). By combining the estimating equation in (4.8) with the restriction in (4.3), we propose the system of restricted estimating equations

$$\mathcal{G}_n(\mathbb{B}) + \boldsymbol{H}'(\mathbb{B})\boldsymbol{\lambda}_n = 0, \quad \hbar^*(\mathbb{B}) = 0, \tag{4.9}$$

where $\boldsymbol{\lambda}_n$ is an $l_1 l_2 \cdots l_d$-column vector of Lagrange multipliers.

Remark 4.1.3. *In the context of the model in* (3.1) *and the restriction* (3.2)*,* (4.8) *and* (4.9) *are equivalent to the equations* (B.1) *and* (B.2)*, respectively.*

4.2 Asymptotic results

In this section, we prove the existence of the solutions of the estimating score equations (4.8) and (4.9). We also show that the resulting estimators are consistent and asymptotically normal. To that end, we first introduce some important definitions and assumptions that will enable us to derive the asymptotic normality of the estimating score function in (4.5). From these assumptions, we will then establish some preliminary results. Subsequently, we use these preliminary results to establish the asymptotic normality of the unrestricted score function. This result is then used to derive the asymptotic normality of the UE and the RE. To introduce some notations, suppose that $\mathcal{G}_n(\mathbb{B})$ is differentiable with respect to $\mathrm{Vec}(\mathbb{B})$ and let $\boldsymbol{C}_n(\mathbb{B}) = -\frac{\partial \mathcal{G}_n(\mathbb{B})}{\partial(\mathrm{Vec}(\mathbb{B}))'}$. In particular, by taking the derivative of $\mathcal{G}_n(\mathbb{B})$ with respect to $-\mathrm{Vec}(\mathbb{B})$, we have

$$\begin{aligned} \boldsymbol{C}_n(\mathbb{B}) &= -\sum_{i=1}^{n} \frac{\partial^2 \mathrm{Vec}(f(\Theta_i))'}{\partial \mathrm{Vec}(\Theta_i)\partial \mathrm{Vec}(\Theta_i)'} \left(\bigotimes_{j=d}^{1} \boldsymbol{X}'_{ij} \right) \mathrm{Vec}(\mathbb{U}_i) \\ &\quad + \sum_{i=1}^{n} \left(\bigotimes_{j=d}^{1} \boldsymbol{X}'_{ij} \right) \left(\frac{\partial \mathrm{Vec}(f(\Theta_i))}{\partial \mathrm{Vec}(\Theta_i)} \right)' \left(\frac{\partial \mathrm{Vec}(f(\Theta_i))}{\partial \mathrm{Vec}(\Theta_i)} \right) \left(\bigotimes_{j=d}^{1} \boldsymbol{X}_{ij} \right). \end{aligned} \tag{4.10}$$

4.2.1 Some definitions and assumptions

We present the following assumption which gives some conditions for the existence of a solution of the estimating equation (4.8) as well as the consistency of the UE.

Assumption 4.2.1. *1. There is a true $\mathbb{B}^* \in \mathbb{R}^{p_1 \times \cdots \times p_d}$ and for all $\mathbb{B}$ within an α neighbourhood $\mathcal{N}^{\alpha}_{\mathbb{B}^*} = \{\mathbb{B} : \|\mathbb{B} - \mathbb{B}^*\| \leq \alpha\}$ such that $\mathcal{G}_n(\mathbb{B})$, $\boldsymbol{C}_n(\mathbb{B})$ and $\frac{\partial \boldsymbol{C}_n(\mathbb{B})}{\partial(\mathrm{Vec}(\mathbb{B}))}$ exist and*

$\mathcal{G}_n(\mathbb{B})$ and $\boldsymbol{C}_n(\mathbb{B})$ are continuous and bounded in absolute value by finitely integrable functions and $\frac{\partial \boldsymbol{C}_n(\mathbb{B})}{\partial(\mathrm{Vec}(\mathbb{B}))}$ is bounded by a function with a finite expectation,

2. *$\frac{1}{n}\boldsymbol{C}_n(\mathbb{B}) \xrightarrow[n\to\infty]{\mathrm{P}} \boldsymbol{\Phi}$, where $\boldsymbol{\Phi}$ is non-random and positive definite matrix. Let $\mu_0 > 0$ be the minimum eigenvalue of $\boldsymbol{\Phi}$.*

To derive some asymptotic results, let $(\Omega, \mathscr{F}, P)$ be a probability space and let

$$\boldsymbol{Z}_{n,t} = n^{-1/2}\left(\bigotimes_{j=d}^{1} \boldsymbol{X}'_{tj}\right)\left(\frac{\partial \mathrm{Vec}(f(\Theta_t))}{\partial \mathrm{Vec}(\Theta_t)}\right)' \mathrm{Vec}(\mathbb{U}_t). \tag{4.11}$$

Note that, for each $t = 1, 2, \ldots, n$ for each $n = 1, 2, \ldots$ $\boldsymbol{Z}_{n,t}$ is $\prod_{i=1}^{d} p_i$-column vector. Let $Z_{n,t,s}$ be the s^{th} component of the vector $\boldsymbol{Z}_{n,t}$, $s = 1, 2, \ldots, \prod_{i=1}^{d} p_i$. Note that $\mathcal{G}_n(\mathbb{B})$ is proportional to the sum of the terms in (4.11) with respect to t. In the following assumption, we present several conditions which will subsequently be used to relax the dependence structure of the model in (4.1). These conditions will be used to show that the array in (4.11) forms an $\mathcal{L}^p-$mixingale.

Assumption 4.2.2. *1. There exists a positive constant array $\{c_{nt}, t = 1, \ldots, n; n = 1, \ldots\}$ such that $\{Z_{n,t,s}/c_{nt}\}$, $s = 1, \cdots, p_1 p_2 \cdots p_d$ is $\mathcal{L}^r$-bounded for $r > 2$ uniformly in t and n;*

2. *$\boldsymbol{Z}_{n,t}$ is near-epoch dependent in $\mathcal{L}^2$ of size -1 on an $\alpha-$mixing array such that $\alpha_m = O\left(m^{-(1+2\varsigma)r/(r-2)}\boldsymbol{\kappa}^{-1}(m)\right)$, $0 \leqslant \varsigma < 1/2$, where $\boldsymbol{\kappa}(\cdot)$ is a positive and increasing function such that for some $M > 0$, $\boldsymbol{\kappa}(x) \geqslant 1$ for all $x \geqslant M$ and $\sum_{n=1}^{\infty}\left(\sum_{j=0}^{n}\boldsymbol{\kappa}(j)\right)^{-1/2} < \infty$.*

3. *For some $\alpha \in (0, 1]$, let $b_n = [n^{-\alpha}]$ and $r_n = [n/b_n]$ and define $M_{ni} = \max_{(i-1)b_n < t \leqslant ib_n}\{c_{nt}\}$ for $i = 1, \cdots, r_n$ and $M_{n,r_n+1} = \max_{r_n b_n < t \leqslant n}\{c_{nt}\}$, the following conditions hold:*

$$\max_{1\leqslant i \leqslant r_n+1} M_{ni} = o(b_n^{-1/2}); \quad \sum_{i=1}^{r_n} M_{ni} = O(b_n^{\varsigma-1/2}); \quad \sum_{i=1}^{r_n} M_{ni}^2 = O(n^{-\alpha}b_n^{-1}),\ 0 \leqslant \varsigma < 1/2. \tag{4.12}$$

4. $\sum_{i=1}^{r_n} \boldsymbol{V}_{n,i}\boldsymbol{V}_{n,i}' \xrightarrow[n\to\infty]{\mathrm{P}} \boldsymbol{\Phi}^*$, *where* $\boldsymbol{V}_{n,i} = \sum_{t=(i-1)b_n+l_n+1}^{ib_n} \boldsymbol{Z}_{n,t},\ i = 1, \cdots, r_n,\ 1 \leqslant l_n < n$ *and* $\boldsymbol{\Phi}^*$ *is a non-random symmetric and positive definite matrix.*

Condition (2) ensures that the for the special case where $\boldsymbol{\kappa}(j) = j^{\epsilon_0},\ \epsilon_0 > 0$, the dependence structure of the error and regressors form an $\mathcal{L}^2-$ mixingale array. Conditions (1), (3) and (4) are useful in deriving a functional central limit theorem for $\mathcal{L}^2$-mixingale arrays of size $-1/2-\varsigma$. The derived result is useful in establishing the joint asymptotic normality of the UE and the RE. Note that some parts of Assumption 4.2.2 are weaker than those of Assumption 3.2.3. Specifically, part 1 of Assumption 4.2.2 gives part 1 of Condition ($\mathscr{C}_6$) of Assumption 3.2.3 and part 2 of Assumption 4.2.2 gives Condition ($\mathscr{C}_5$) of Assumption 3.2.3 as will be shown in Lemma 4.2.1. Parts 3 and 4 are equivalent to parts 2-4 of Assumption 3.2.3.

4.2.2 On the asymptotic distribution of the estimating score function

In this subsection, we present some notations and some preliminary results which are useful in deriving the main results of this paper. We consider that we have some filtration $\{\mathscr{F}_{n,s}^t, -\infty \leq s \leq t \leq \infty,\ n \geq 1\}$ and we prove that, under Assumption 4.2.2, $\{\boldsymbol{Z}_{n,t}, \mathscr{F}_{n,-\infty}^t\}$ is an $\mathcal{L}^2-$mixingale of size $-1/2-\varsigma$. Thanks to this result, we derive some preliminary results that help us establish the joint asymptotic normality of the UE and the RE. To simplify some notations, we take $\mathrm{E}^m\boldsymbol{U} = \mathrm{E}[\boldsymbol{U}|\mathscr{F}_{n,-\infty}^m]$ and $\mathrm{E}_n^m\boldsymbol{U} = \mathrm{E}[\boldsymbol{U}|\mathscr{F}_{n,k}^m]$, for some random array $\boldsymbol{U}$ and integers $-\infty < k < m < \infty$.

Lemma 4.2.1. *Suppose that* $\{Z_{n,t}\}$ *satisfies Parts 1-3 of Assumption 4.2.2, then* $\|\mathrm{E}^{t-m}\boldsymbol{Z}_{n,t}\|_p \leqslant \zeta_{p,m}c_{nt}$ *and* $\|\boldsymbol{Z}_{n,t} - \mathrm{E}^{t+m}\boldsymbol{Z}_{n,t}\|_p \leqslant \zeta_{p,m+1}c_{nt}$, *where* $\zeta_{p,m} = O_p(m^{-\lambda^*})$, $\lambda^* = \min\{\frac{(r-p)(1+2\varsigma)}{p(r-2)} + \log_m \boldsymbol{\kappa}(m), 1+\delta_0\}$, $\delta_0 > 0$.

Proof. For a fixed m, $\mathrm{E}_{t-[m/2]}^{t+[m/2]}\boldsymbol{Z}_{nt}$ is an $\mathscr{F}_{n,-\infty}^{t+i}-$measurable function for $i = -[m/2], -[m/2]+$

$1, \cdots, [m/2]$ and is hence strong mixing of the same size.Then, we have

$$\begin{aligned} \|\mathrm{E}^{t-m}\boldsymbol{Z}_{nt}\|_p &= \|\mathrm{E}^{t-m}(\boldsymbol{Z}_{nt} - \mathrm{E}_{t-[m/2]}^{t+[m/2]}\boldsymbol{Z}_{nt} + \mathrm{E}_{t-[m/2]}^{t+[m/2]}\boldsymbol{Z}_{nt})\|_p \\ &\leq \|\mathrm{E}^{t-m}(\boldsymbol{Z}_{nt} - \mathrm{E}_{t-[m/2]}^{t+[m/2]}\boldsymbol{Z}_{nt})\|_p + \|\mathrm{E}^{t-m}(\mathrm{E}_{t-[m/2]}^{t+[m/2]}\boldsymbol{Z}_{nt})\|_p. \end{aligned}$$

Using Lemma 4.1 of Davidson (1992) and Jensen's inequality, we have

$$\begin{aligned} \|\mathrm{E}^{t-m}(\mathrm{E}_{t-[m/2]}^{t+[m/2]}\boldsymbol{Z}_{nt})\|_p &\leq 2(2^{1/p}+1)\alpha_{[m/2]}^{1/p-1/r}\|\mathrm{E}_{t-[m/2]}^{t+[m/2]}\boldsymbol{Z}_{nt}\|_r \\ \leq \quad & 2(2^{1/p}+1)\alpha_{[m/2]}^{1/p-1/r}\|\boldsymbol{Z}_{nt}\|_r \leq 2(3)\alpha_{[m/2]}^{1/p-1/r}\|\boldsymbol{Z}_{nt}\|_r. \end{aligned}$$

Further, by the definition of near epoch, we have $\|\mathrm{E}^{t-m}(\boldsymbol{Z}_{nt} - \mathrm{E}_{t-[m/2]}^{t+[m/2]}\boldsymbol{Z}_{nt})\|_p \leq \nu_{[m/2]}d_{nt}$. Thus, we have

$$\|\mathrm{E}^{t-m}\boldsymbol{Z}_{nt}\|_p \leq 6\alpha_{[m/2]}^{1/p-1/r}\|\boldsymbol{Z}_{nt}\|_r + \nu_{[m/2]}d_{nt} \leq (6\alpha_{[m/2]}^{1/p-1/r} + \nu_{[m/2]})\max\{\|\boldsymbol{Z}_{nt}\|_r, d_{nt}\}.$$

Then,

$$\|\mathrm{E}^{t-m}\boldsymbol{Z}_{nt}\|_p \leq 6\alpha_{[m/2]}^{1/p-1/r} + \nu_{[m/2]})2Kc_{nt} = \zeta_{p,m}c_{nt}$$

with $\zeta_{p,m}c_{nt} = B(6\alpha_{[m/2]}^{1/p-1/r} + \nu_{[m/2]})$, for $B \geq 4$.
We also have $\|\boldsymbol{Z}_{nt} - \mathrm{E}^{t+m}\boldsymbol{Z}_{nt}\|_p \leq 2\|\boldsymbol{Z}_{nt} - \mathrm{E}_{t-m}^{t+m}\boldsymbol{Z}_{nt}\|_p \leq 2\nu_m d_{nt} \leq \zeta_{m+1}^{(p)}c_{nt}$ by Lemma 4.2 of Davidson (1992). By part 3 of Assumption 4.2.2, $\alpha_m^{1/p-1/r} = O\left(m^{-\frac{(r-p)(1+2\varsigma)}{p(r-2)}}\boldsymbol{\kappa}^{-(1/p-1/r)}(m)\right)$ and $\nu_m = O(m^{-\lambda}) = O(m^{-1-\delta_0})$, for $\lambda > 1$, for some $\delta_0 > 0$. Therefore, we have $\zeta_{p,m} = O(m^{-\lambda^*})$, where $\lambda^* = \min\{\frac{(r-p)(1+2\varsigma)}{p(r-2)}\boldsymbol{\kappa}^{(1/p-1/r)}(m), 1+\delta_0\}$. The result follows. □

Remark 4.2.1. *For the special case where* $\boldsymbol{\kappa}(m) = m^{\epsilon_0}$ *for some* $\epsilon_0 > 0$, *we have that* $\{\boldsymbol{Z}_{n,t}\}$ *is an* $\mathcal{L}^p-$ *mixingale of size* $\max\{-\frac{(r-p)(1+2\varsigma)}{p(r-2)}, -1\}$ *with respect to the constant array* $\{c_{nt}\}$. *In particular, if* $p = 2$, $\{\boldsymbol{Z}_{n,t}, \mathscr{F}_{n,-\infty}^t\}$ *forms an* $\mathcal{L}^2-$ *mixingale array of size* $\max\{-1-\varsigma/2, -1\}$, *for* $0 \leqslant \varsigma \leqslant 1$.

From Lemma 4.2.1, we establish the following lemma. In particular, the established lemma will be used to derive the asymptotic properties of the estimating score function in (4.5).

Lemma 4.2.2. *Let* $v_j^2 = \sum_{i=1}^{j} c_{n,i}^2$, $\tilde{v}_j^2(k) = \sum_{i=k+1}^{k+j} c_{n,i}^2$, $k = 0, 1, \ldots, j$, $j = 1, \ldots$, $S_{j,s} = \sum_{i=1}^{j} Z_{n,i,s}$, $s = 1, \cdots, \prod_{i=1}^{d} p_i$, *and* $\tilde{v}_i^2 = \sum_{(i-1)b_n+l_n+1}^{ib_n} c_{n,t}^2$. *Under Assumption 4.2.2,*

1. The sets $\left\{ \sum_{s=1}^{\prod_{i=1}^{d} p_i} \max_{j \leqslant L} \frac{\left(S_{k+j,s} - S_{k,s}\right)^2}{\tilde{v}_L^2(k)}; k = 0, \ldots, L; L = 1, \cdots \right\}$, $\left\{ \sum_{s=1}^{\prod_{i=1}^{d} p_i} \max_{j \leqslant L} \frac{S_{j,s}^2}{v_L^2}; L = 1, 2, \cdots \right\}$ *are uniformly integrable;*

2. The sets $\left\{ \sum_{s=1}^{\prod_{i=1}^{d} p_i} \max_{j \leqslant ib_n} \left(S_{j,s} - S_{(i-1)b_n+l_n,s}\right)^2 \Big/ \tilde{v}_i^2, i = 1, 2, \cdots \right\}$, $\left\{ \sum_{s=1}^{\prod_{i=1}^{d} p_i} \frac{\left(S_{ib_n,s} - S_{(i-1)b_n+l_n,s}\right)^2}{\tilde{v}_i^2}, i = 1, 2, \cdots \right\}$ *are uniformly integrable.*

The proof of this lemma follows directly from Lemma 3.2.4 by taking $q_2 = \cdots = q_{d+1} = 1$ and $q_1 = \prod_{i=1}^{d} p_i$, $m_0 = 0$. For the convenience of the reader, we also provide alternate proofs of Lemma 4.2.2 and all related results in Appendix C.

From the above Lemma 3.2.4, we establish the following theorem which is the main contribution of this section. The established theorem leads immediately to the asymptotic distribution of the estimating function, $\mathcal{G}_n(\mathbb{B}^*)$. For the sake of simplicity, let $\mathcal{P}_{n,j}(t) = \sum_{i=1}^{\lfloor jt \rfloor} \boldsymbol{Z}_{n,i}$, $t \in [0, 1]$ and let $D^k([0,\ 1])$ denote the space of all k-column vectors of functions which are right continuous with left limits on $[0,\ 1]$.

Theorem 4.2.1. *Under Assumption 4.2.2,* $\mathcal{P}_{n,n}(1) = n^{-1/2}\mathcal{G}_n(\mathbb{B}^*) \xrightarrow[n\to\infty]{D} \boldsymbol{U} \sim \mathcal{N}_{p_1 \cdots p_d}(0,\ \boldsymbol{\Phi}^*)$. *Further, for each* $t \in [0,\ 1]$, $\mathcal{P}_{n,n}(t) \xrightarrow[n\to\infty]{d} \sqrt{t}\,\boldsymbol{U}$.

Proof. Let $\gamma_n = \sum_{i=1}^{r_n} \left(\boldsymbol{V}_{n,i} - \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_i^*) + \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_{i-1}^*)\right) + \sum_{i=1}^{r_n} \sum_{t=(i-1)b_n+1}^{(i-1)b_n+l_n} \boldsymbol{Z}_{n,t} + \sum_{t=r_nb_n+1}^{n} \boldsymbol{Z}_{n,t}$. We have

$$\boldsymbol{S}_n = \sum_{t=1}^{n} \boldsymbol{Z}_{n,t} = \gamma_n + \sum_{i=1}^{r_n} \boldsymbol{W}_{n,i} + \left(\sum_{i=r_n+1}^{n} \left(\sum_{t=(i-1)b_n+1}^{ib_n} \boldsymbol{Z}_{n,t}\right)\right). \tag{4.13}$$

Then, from Proposition C.2.5, we have

$$\gamma_n \xrightarrow[n\to\infty]{\mathrm{P}} 0. \tag{4.14}$$

Moreover, letting $S_j^* = \{t : t \in \bigcup_{i=r_n+1}^{n} [(i-1)b_n + 1, ib_n]\}$, we have

$$\sum_{s=1}^{\prod_{i=1}^{d} p_i} \mathrm{E}\left[\left(\sum_{i=r_n}^{n} \sum_{t=(i-1)b_n+1}^{ib_n} Z_{n,t,s}\right)^2\right] = \sum_{s=1}^{\prod_{i=1}^{d} p_i} \mathrm{E}\left[\left(\sum_{t\in S_j^*} Z_{n,t,s}\right)^2\right],$$

and then, together with Lemma 4.2.2, we get

$$\sum_{s=1}^{\prod_{i=1}^{d} p_i} \mathrm{E}\left[\left(\sum_{i=r_n}^{n} \sum_{t=(i-1)b_n+1}^{ib_n} Z_{n,t,s}\right)^2\right] = O\left(\sum_{i=r_n+1}^{n} \sum_{t=(i-1)b_n+1}^{ib_n} c_{n,t}^2\right) = O\left(\sum_{i=r_n+1}^{n} \left(\max_{(i-1)b_n+1\leqslant t\leqslant ib_n} c_{n,t}\right)^2 b_n\right) = o(1).$$

Hence

$$\sum_{i=r_n+1}^{n} \left(\sum_{t=(i-1)b_n+1}^{ib_n} \boldsymbol{Z}_{n,t}\right) \xrightarrow[n\to\infty]{\mathrm{P}} 0. \tag{4.15}$$

In addition, by Proposition C.2.5, we have $\sum_{i=1}^{r_n} \boldsymbol{W}_{n,i}\boldsymbol{W}'_{n,i} \xrightarrow[n\to\infty]{\mathrm{P}} \boldsymbol{\Phi}^*$ and $\sum_{i=1}^{r_n} \sum_{s=1}^{\prod_{i=1}^{d} p_i} \mathrm{E}\left[(W_{n,i,s})^2 \mathbb{I}\left(\sum_{s=1}^{\prod_{i=1}^{d} p_i} W_{n,i,s}^2 > \epsilon\right)\right] \xrightarrow[n\to\infty]{} 0$, for all $\epsilon > 0$. Then, by the martingale difference sequence central limit theorem, $\sum_{i=1}^{r_n} \boldsymbol{W}_{n,i} \xrightarrow[T\to\infty]{d} \mathcal{N}_{p_1 p_2 \cdots p_d}(0, \boldsymbol{\Phi}^*)$. Hence, together with (B.11), (B.12), (B.13) along with Slutsky's theorem, we prove the first claim. The second claim follows from the first statement along with the fact that $\mathcal{P}_{n,n}(t) = \left(\lfloor tn \rfloor n^{-1}\right)^{1/2} \mathcal{P}_{n,\lfloor tn \rfloor}(1)$, this completes the proof. □

Below, we prove that, under additional conditions, $\{\mathcal{P}_{n,n}(t)\}_1^\infty$ converges weakly to a Gaussian process. As an intermediate result, we first establish the following lemma.

Lemma 4.2.3. *Under Assumption 4.2.2, if*

$$\sup_{0<\alpha<1-a,0\leqslant a<1} \limsup_{n\to\infty} \alpha^{-1} \sum_{i=\lfloor na \rfloor}^{\lfloor n(a+\alpha) \rfloor} c_{n,i}^2 < \infty, \tag{4.16}$$

then

1. $\left\{\max_{u\leqslant t\leqslant u+\alpha} \left\|\mathcal{P}_{n,n}(t) - \mathcal{P}_{n,n}(u)\right\|^2 \big/ \alpha; n \geq N(u,\alpha), 0 \leqslant s \leqslant 1, \alpha \in \mathrm{T}\right\}$ *is a uniformly integrable set for some sequence* T *of* α *approaching 0 and nonrandom finite valued function* $N(u, \alpha)$.

2. $\{\mathcal{P}_{n,n}(.)\}_{n=1}^{\infty}$ *is tight in Stone's topology on* $D^{\prod_{i=1}^{d} p_i}([0,1])$;

3. *for each* $t \in [0,\ 1]$, *the set* $\left\{\|\mathcal{P}_{n,n}(t)\|^2 : n = 1, 2, \dots\right\}$ *is uniformly integrable;*

4. *the weak limit process of any convergent subsequence of* $\{\boldsymbol{S}(.)\}$ *is almost surely continuous.*

Proof. One can verify that,

$$
\begin{aligned}
&\max_{s\leqslant t\leqslant s+\alpha} \frac{\left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2}{\alpha} \mathbb{I}\left(\max_{s\leqslant t\leqslant s+\alpha} \frac{\left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2}{\alpha} > b\right) \\
&= \max_{s\leqslant t\leqslant s+\alpha} \frac{\bar{v}^2\left(\lfloor sj\rfloor, \lfloor jt\rfloor\right)}{\alpha} \frac{\left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2}{\bar{v}^2\left(\lfloor sj\rfloor, \lfloor jt\rfloor\right)} \mathbb{I}\left(\max_{s\leqslant t\leqslant s+\alpha} \frac{\left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2 \bar{v}^2\left(\lfloor sj\rfloor, \lfloor jt\rfloor\right)}{\bar{v}^2\left(\lfloor sj\rfloor, \lfloor jt\rfloor\right)} > b\alpha\right) \\
&\leqslant \alpha^{-1} \max_{s\leqslant t\leqslant s+\alpha} \bar{v}^2\left(\lfloor sj\rfloor, \lfloor jt\rfloor\right) Y_j(\alpha, s) \mathbb{I}\left(Y_j(\alpha, s) > \frac{b\alpha}{\max_{s\leqslant t\leqslant s+\alpha} \bar{v}^2\left(\lfloor sj\rfloor, \lfloor jt\rfloor\right)}\right),
\end{aligned}
$$

with $\bar{v}^2\left(\lfloor sj\rfloor, \lfloor jt\rfloor\right) = \sum_{i=\lfloor sj\rfloor}^{\lfloor jt\rfloor} c_{n,i}^2$ and $Y_j(\alpha, s) = \max_{s\leqslant t\leqslant s+\alpha} \left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2 / \bar{v}^2\left(\lfloor sj\rfloor, \lfloor jt\rfloor\right)$. From Lemma 4.2.2, $\left\{Y_j(\alpha, s) : j \geq N(s, \alpha)\right\}$ is uniformly integrable. Then,

$$
\begin{aligned}
&\sup_{j,s} \mathrm{E}\left[\max_{s\leqslant t\leqslant s+\alpha} \frac{\left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2}{\alpha} \mathbb{I}\left(\max_{s\leqslant t\leqslant s+\alpha} \frac{\left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2}{\alpha} > b\right)\right] \\
&\leqslant \alpha^{-1} \sum_{i=\lfloor sj\rfloor}^{\lfloor j(s+\alpha)\rfloor} c_{n,i}^2 \sup_{j,s} \mathrm{E}\left[Y_j(\alpha, s) \mathbb{I}\left(Y_j(\alpha, s) > b\left(\alpha^{-1} \sum_{i=\lfloor sj\rfloor}^{\lfloor j(s+\alpha)\rfloor} c_{n,i}^2\right)^{-1}\right)\right],
\end{aligned}
$$

and then, using Condition (A.9), we get

$$
\begin{aligned}
&\sup_{j,s} \mathrm{E}\left[\max_{s\leqslant t\leqslant s+\alpha} \frac{\left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2}{\alpha} \mathbb{I}\left(\max_{s\leqslant t\leqslant s+\alpha} \frac{\left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2}{\alpha} > b\right)\right] \\
&\leqslant C_0 \sup_{j,s} \mathrm{E}\left[Y_j(\alpha, s) \mathbb{I}\left(Y_j(\alpha, s) > b\left(\alpha^{-1} \sum_{i=\lfloor sj\rfloor}^{\lfloor j(s+\alpha)\rfloor} c_{n,i}^2\right)^{-1}\right)\right],
\end{aligned}
\tag{4.17}
$$

for some $C_0 > 0$. Further, from (A.9), $\lim_{\alpha\to 0} \limsup_{j\to\infty} \left(\alpha^{-1} \sum_{i=\lfloor sj\rfloor}^{\lfloor j(s+\alpha)\rfloor} c_{n,i}^2\right)^{-1} = 0$, for arbitrary $0 \leqslant s \leqslant 1$. Therefore, together with (4.17), we get

$$
\sup_{0\leqslant s\leqslant 1} \sup_{j} \mathrm{E}\left[\max_{s\leqslant t\leqslant s+\alpha} \frac{\left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2}{\alpha} \mathbb{I}\left(\max_{s\leqslant t\leqslant s+\alpha} \frac{\left\|\mathcal{P}_{j,j}(t) - \mathcal{P}_{j,j}(s)\right\|^2}{\alpha} > b\right)\right] = 0,
$$

for arbitrary α, this completes the proof of Part **1**. The proof of Part **2** follows by combining Part **1** and Theorem 8.4 of Billingsley (1968). Part **3** follows from Part **1**. Finally, Part **1** follows from Part **1** along with Proposition VI.3.26 in Jacod and Shiryaev (1987). This completes the proof. $\square$

Note that the condition in (A.9) permits the inclusion of cases where $Z_{n,i}$ have growing moments. From Theorem 4.2.1 and Lemma 4.2.3, we derive the following result.

Theorem 4.2.2. *Suppose that* $\lim_{n\to\infty} \mathrm{E}\left[\mathcal{P}_{n,n}(t)\mathcal{P}'_{n,n}(t)\right] = \boldsymbol{\Lambda}^*(t)$ *for all* $t \in [0, 1]$ *along with the conditions of Lemma 4.2.3. Then,* $\mathcal{P}_{n,n}(t) \xrightarrow[n\to\infty]{d} \boldsymbol{S}(t)$ *where* $\{\boldsymbol{S}(t) : 0 \leqslant t \leqslant 1\}$ *is a vector-valued Gaussian process with almost surely continuous paths and independent increments.*

Proof. From Part **(3)** of Lemma 4.2.3, we have $\left(\mathcal{P}'_{n,n}(t_1), \mathcal{P}'_{n,n}(t_2), \ldots, \mathcal{P}'_{n,n}(t_k)\right)' \xrightarrow[n\to\infty]{d} (\boldsymbol{S}'(t_1), \boldsymbol{S}'(t_2) \ldots, \boldsymbol{S}'(t_k))'$, for all $(t_1, t_2, \ldots, t_k)' \in ([0\,1])^k$. Then, by combining Theorem 4.2.1 and Lemma 4.2.3, we conclude that $\mathcal{P}_{n,n}(t) \xrightarrow[n\to\infty]{d} \boldsymbol{S}(t)$ where $\{\boldsymbol{S}(t) : 0 \leqslant t \leqslant 1\}$ is a Gaussian process with almost surely continuous paths. Then, the proof is completed if we prove that, for any set $\{t_1, \cdots, t_k : 0 < t_1 < \cdots < t_k < 1\}$ and for all $i < j$, $\mathcal{P}_{n,n}(t_i) - \mathcal{P}_{n,n}(t_{i-1})$ and $\mathcal{P}_{n,n}(t_j) - \mathcal{P}_{n,n}(t_{j-1})$ are asymptotically uncorrelated. To this end, one can verify that, $\left\|\mathrm{E}\left[(\mathcal{P}_{n,n}(t_i) - \mathcal{P}_{n,n}(t_{i-1}))\left(\left(\mathcal{P}_{n,n}(t_j) - \mathcal{P}_{n,n}(t_{j-1})\right)\right)'\right]\right\| = \left\|\mathrm{E}\left[\left(\sum_{k=\lfloor nt_{i-1}\rfloor+1}^{\lfloor nt_i\rfloor} Z_{n,k}\right)\left(\sum_{k=\lfloor nt_{j-1}\rfloor+1}^{\lfloor nt_j\rfloor} Z_{n,k}\right)'\right]\right\|$ and then, letting $p = \prod_i^d p_i$, for fixed $\alpha > 0$, we get

$$\left\|\mathrm{E}\left[(\mathcal{P}_{n,n}(t_i) - \mathcal{P}_{n,n}(t_{i-1}))\left(\left(\mathcal{P}_{n,n}(t_j) - \mathcal{P}_{n,n}(t_{j-1})\right)\right)'\right]\right\| \leqslant \sum_{l=1}^{p}\left\|\sum_{k=\lfloor nt_{i-1}\rfloor+1}^{\lfloor nt_i\rfloor} Z_{n,k,l}\right\|_2 \left\|\sum_{k=\lfloor nt_{j-1}\rfloor+1}^{\lfloor n(t_{j-1}+\alpha)\rfloor} Z_{n,k,l}\right\|_2 + \sum_{l=1}^{p}\sum_{g=1}^{n}\sum_{h=1}^{n}\left|\mathrm{E}(Z_{n,g,l}Z_{n,h,l})\right|\mathbb{I}\left(|s-t| \geqslant n(t_{j-1}+\alpha) - nt_i\right).$$

Further, since $\lfloor nt\rfloor - \lfloor ns\rfloor \to \infty$ for all $t > s$, by some algebraic computations, we get $\lim_{n\to\infty}\sum_{l=1}^{n}\sum_{k=1}^{n}\left|\mathrm{E}(Z_{n,g,l}Z_{n,h,l})\right|\mathbb{I}\left(|s-t| \geqslant n(t_{j-1}+\alpha) - nt_i\right) = 0$. Hence, together with (A.9), we get $\lim_{n\to\infty}\mathrm{E}\left[(\mathcal{P}_{n,n}(t_i) - \mathcal{P}_{n,n}(t_{i-1}))\left(\mathcal{P}_{n,n}(t_j) - \mathcal{P}_{n,n}(t_{j-1})\right)'\right] = 0$, this completes the proof. $\square$

4.2.3 On existence and consistency of the UE and the RE

In this subsection, we derive two lemmas which show that the estimating equations (4.8) and (4.9) have solutions. The established results also show that the UE and the RE are consistent estimators for the tensor parameter $\mathbb{B}$ if the solutions are unique. To prove Lemma 4.2.5 and Lemma 4.2.6, we use the following lemma from Aitchison and Silvey (1958).

Lemma 4.2.4. *Let g be a continuous function mapping from $\mathbb{R}^r$ onto itself. If for every ψ such that $||\psi|| = 1$ we have $\psi' g(\psi) < 0$ then there exists a point $\hat{\psi}$ such that $||\hat{\psi}|| < 1$ and $g(\hat{\psi}) = 0$.*

The proof of Lemma 4.2.4 can be found in Aitchison and Silvey (1958).

Lemma 4.2.5. *Suppose that Assumption 4.2.2 and Assumption 4.2.1 hold. Then, for an arbitrarily small $\delta > 0$ and some $0 < \epsilon < 1$, there exists $n_{\epsilon,\delta}$ such that for all $n > n_{\epsilon,\delta}$, $\mathcal{G}_n(\mathbb{B}) = 0$ has a solution $\hat{\mathbb{B}}$ with probability greater than $1 - \epsilon$ and with $||\hat{\mathbb{B}} - \mathbb{B}^*|| \leq \delta$. Moreover, if there exists an n_0 such that a solution to $\mathcal{G}_n(\mathbb{B}) = 0$ is unique for all $n > n_0$, then $\hat{\mathbb{B}} \xrightarrow[n\to\infty]{\mathrm{P}} \mathbb{B}^*$.*

Proof. We have the equation

$$\frac{1}{n}\mathcal{G}_n(\mathbb{B}) = 0. \tag{4.18}$$

Let an α neighbourhood of $\mathbb{B}^*$ be denoted as $\mathcal{N}^{\alpha}_{\mathbb{B}^*} = \{\mathbb{B} : ||\mathbb{B} - \mathbb{B}^*|| \leq \alpha\}$. Let $\delta < \min\{\alpha, 1\}$ and suppose $\mathbb{B} \in \mathcal{N}^{\delta}_{\mathbb{B}^*}$. Then, expanding (4.18) about $\mathbb{B}^*$, we get the equation

$$\frac{1}{n}\mathcal{G}_n(\mathbb{B}^*) - \frac{1}{n}\boldsymbol{C}_n(\mathbb{B}^*)(\mathrm{Vec}(\mathbb{B}) - \mathrm{Vec}(\mathbb{B}^*)) + \frac{1}{n}\tilde{r}_1(\mathbb{B}) = 0, \tag{4.19}$$

where $\tilde{r}_1(\mathbb{B})$ is the remainder term of higher order derivatives. Then, we have by Theorem 4.2.1, $\frac{1}{n}\mathcal{G}_n(\mathbb{B}^*) = o_p(1)$ and by Condition (1) of Assumption 4.2.1, $\frac{1}{n}\tilde{r}_1(\mathbb{B}) = o_p(||\mathbb{B} -$

$\mathbb{B}^*\|^2)$. Further, $\frac{1}{n}\boldsymbol{C}_n(\mathbb{B}) - \boldsymbol{\Phi} = o_p(1)$. Hence, (4.19) becomes the equation

$$o_p(1) - \boldsymbol{\Phi}(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) - o_p(1)(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + o_p(\|\mathbb{B} - \mathbb{B}^*\|^2) = 0, \quad (4.20)$$

Since $\delta < 1$, then we get the equation

$$-\boldsymbol{\Phi}(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + o_p(1) = 0, \quad (4.21)$$

Define $\psi = (\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*))/\delta$ and $g(\psi) = -\delta\boldsymbol{\Phi}\psi + o_p(1)$ and fix ϵ, $0 < \epsilon < 1$. Choose a sufficiently small δ such that for some natural number $N_{\epsilon,\delta} > 0$, we have $\text{P}(\|o_p(1)\| < \delta^2) > 1 - \epsilon$, for all $n > N_{\epsilon,\delta}$. Then, for sufficiently large n we have with probability greater than $1 - \epsilon$ that

$\psi' g(\psi) = -\frac{1}{\delta}(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*))'\boldsymbol{\Phi}(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + \frac{1}{\delta}(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*))' o_p(1)$, then

$$\psi' g(\psi) \leqslant -\frac{1}{\delta}\mu_0\|\mathbb{B} - \mathbb{B}^*\|^2 + \frac{\delta^2}{\delta}\|\mathbb{B} - \mathbb{B}^*\|, \quad (4.22)$$

where the inequality follows by Condition (2) of Assumption 4.2.1 for the first term and the Cauchy-Schwarz inequality for the second term. Thus, by choosing $\mathbb{B}$ such that $\|\mathbb{B} - \mathbb{B}^*\| = \delta$, we have

$$\psi' g(\psi) \leqslant -\delta\mu_0 + \delta^2 < 0 \quad (4.23)$$

with probability $1 - \epsilon$ for a sufficiently small δ and a sufficiently large n. Therefore, by Lemma 4.2.4, we have for a sufficiently small δ and $0 < \epsilon < 1$ there exists an $N_{\epsilon,\delta}$ such that for all $n > N_{\epsilon,\delta}$, we have a $\hat{\psi}$ with $g(\hat{\psi}) = 0$ with probability greater than $1 - \epsilon$. That is, (4.23) has a solution for $\|\mathbb{B} - \mathbb{B}^*\| < \delta$, say $\hat{\mathbb{B}}$, and assuming the solution is unique then it will also be consistent for $\mathbb{B}^*$. □

Lemma 4.2.6. *Suppose that Assumption 4.1.1 holds along with the conditions of Lemma 4.2.5. Then,* $\mathcal{G}_n(\mathbb{B}) + H(\mathbb{B})'\lambda_n = 0$ *has a consistent solution in* $\{(\mathbb{B}) : \|\mathbb{B} - \mathbb{B}^*\| \leq \delta\}$*, denoted as* $\tilde{\mathbb{B}}$.

Proof. We have the system of equations

$$\frac{1}{n}\mathcal{G}_n(\mathbb{B}) + H(\mathbb{B})'\boldsymbol{\lambda}_n = 0, \quad \text{and} \quad \hbar^*(\mathbb{B}) = 0. \tag{4.24}$$

Let an α neighbourhood of $\mathbb{B}^*$ be denoted as $\mathcal{N}^{\alpha}_{\mathbb{B}^*} = \{\mathbb{B} : \|\mathbb{B} - \mathbb{B}^*\| \leq \alpha\}$. Let $\delta < \min\{\alpha, 1\}$ and suppose $\mathbb{B} \in \mathcal{N}^{\delta}_{\mathbb{B}^*}$. Then, expanding (4.24) about $\mathbb{B}^*$, we get the system of equations

$$\begin{aligned} \frac{1}{n}\mathcal{G}_n(\mathbb{B}^*) - \frac{1}{n}\boldsymbol{C}_n(\mathbb{B}^*)(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + \frac{1}{n}\tilde{r}_1(\mathbb{B}) + H(\mathbb{B})'\boldsymbol{\lambda}_n &= 0, \\ \hbar^*(\mathbb{B}^*) + H(\mathbb{B})(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + \tilde{r}_2(\mathbb{B}) &= 0, \end{aligned} \tag{4.25}$$

where $\tilde{r}_1(\mathbb{B})$ and $\tilde{r}_2(\mathbb{B})$ are the remainder terms of higher order derivatives. Then, we have by Theorem 4.2.1, $\frac{1}{n}\mathcal{G}_n(\mathbb{B}^*) = o_p(1)$ and by (1) of Assumption 4.2.1, $\frac{1}{n}\tilde{r}_1(\mathbb{B}) = o_p(\|\mathbb{B}-\mathbb{B}^*\|^2)$. In addition, $\frac{1}{n}\boldsymbol{C}_n(\mathbb{B}) - \boldsymbol{\Phi} = o_p(1)$, $\tilde{r}_2(\mathbb{B}) = O(\|\mathbb{B} - \mathbb{B}^*\|^2)$, and $\hbar^*(\mathbb{B}^*) = 0$. Hence, (4.25) becomes

$$\begin{aligned} &o_p(1) - \boldsymbol{\Phi}(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) - o_p(1)(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + o_p(\|\mathbb{B} - \mathbb{B}^*\|^2) + H(\mathbb{B})'\boldsymbol{\lambda}_n = 0, \\ &H(\mathbb{B})(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + O(\|\mathbb{B} - \mathbb{B}^*\|^2) = 0. \end{aligned} \tag{4.26}$$

Since $\delta < 1$, then we get the system of equations

$$-\boldsymbol{\Phi}(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + H(\mathbb{B})'\boldsymbol{\lambda}_n + o_p(1) = 0, \tag{4.27}$$

$$H(\mathbb{B})(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + O(\|\mathbb{B} - \mathbb{B}^*\|^2) = 0.$$

Then, we multiply both sides of (4.27) by $H(\mathbb{B}^*)\boldsymbol{\Phi}^{-1}$, and we get the equation

$$- H(\mathbb{B}^*)(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + H(\mathbb{B}^*)\boldsymbol{\Phi}^{-1}H(\mathbb{B})'\boldsymbol{\lambda}_n + o_p(1) = 0. \tag{4.28}$$

Furthermore, since $H(\mathbb{B})$ is full column rank, we have that $H(\mathbb{B}^*)\boldsymbol{\Phi}^{-1}H(\mathbb{B}^*)$, is invertible and for a sufficiently small $\boldsymbol{\delta}$, $H(\mathbb{B}^*)\boldsymbol{\Phi}^{-1}H(\mathbb{B})$, would also be invertible. Hence using (4.28), we can solve for $\boldsymbol{\lambda}_n$ to obtain $\boldsymbol{\lambda}_n = [H(\mathbb{B}^*)\boldsymbol{\Phi}^{-1}H(\mathbb{B})']^{-1}(H(\mathbb{B}^*)(\text{Vec}(\mathbb{B}) - \text{Vec}(\mathbb{B}^*)) + o_p(1)$, this gives

$$\boldsymbol{\lambda}_n[H(\mathbb{B}^*)\boldsymbol{\Phi}^{-1}H(\mathbb{B})']^{-1}(H(\mathbb{B}^*)O(\|\mathbb{B} - \mathbb{B}^*\|^2) + o_p(1) = O(\|\mathbb{B} - \mathbb{B}^*\|^2) + o_p(1). \tag{4.29}$$

Substituting $\boldsymbol{\lambda}_n$ into (4.27), we get

$$-\boldsymbol{\Phi}(\mathrm{Vec}(\mathbb{B}) - \mathrm{Vec}(\mathbb{B}^*)) + H(\mathbb{B})' O(\|\mathbb{B} - \mathbb{B}^*\|^2) + o_p(1) = 0. \tag{4.30}$$

Since $H(\mathbb{B})$ is bounded, then the first equation of (4.27) can be rewritten as

$$-\boldsymbol{\Phi}(\mathrm{Vec}(\mathbb{B}) - \mathrm{Vec}(\mathbb{B}^*)) + \delta^2 \nu(\mathbb{B}) + o_p(1) = 0, \tag{4.31}$$

for some bounded continuous function $\nu(\mathbb{B})$ with $\|\nu(\mathbb{B})\| < K,\ K > 0$. Define $\psi = (\mathrm{Vec}(\mathbb{B}) - \mathrm{Vec}(\mathbb{B}^*))/\delta$, $g(\psi) = -\delta\boldsymbol{\Phi}\psi + \delta^2\nu(\mathbb{B}) + o_p(1)$ and fix $\epsilon,\ 0 < \epsilon < 1$. Choose a sufficiently small δ such that for some natural number $N_{\epsilon,\delta} > 0$, we have $\mathrm{P}(\|o_p(1)\| < \delta^2) > 1 - \epsilon$, for all $n > N_{\epsilon,\delta}$. Then, for sufficiently large n we have with probability greater than $1 - \epsilon$ that

$$\psi' g(\psi) = -\frac{1}{\delta}(\mathrm{Vec}(\mathbb{B}) - \mathrm{Vec}(\mathbb{B}^*))'\boldsymbol{\Phi}(\mathrm{Vec}(\mathbb{B}) - \mathrm{Vec}(\mathbb{B}^*)) + \delta(\mathrm{Vec}(\mathbb{B}) - \mathrm{Vec}(\mathbb{B}^*))'\nu(\mathbb{B}) + \frac{1}{\delta}(\mathrm{Vec}(\mathbb{B}) - \mathrm{Vec}(\mathbb{B}^*))' o_p(1).$$

Then,

$$\psi' g(\psi) \leqslant -\frac{1}{\delta}\mu_0\|\mathbb{B} - \mathbb{B}^*\|^2 + \delta K\|\mathbb{B} - \mathbb{B}^*\| + \frac{\delta^2}{\delta}\|\mathbb{B} - \mathbb{B}^*\|, \tag{4.32}$$

where the inequality follows from Condition 2) of Assumption 4.2.1 for the first term and the Cauchy-Schwarz inequality for the other terms. Thus, by choosing $\mathbb{B}$ such that $\|\mathbb{B} - \mathbb{B}^*\| = \delta$, we have

$$\psi' g(\psi) \leq -\delta\mu_0 + \delta^2 K + \delta^2 < 0 \tag{4.33}$$

with probability $1 - \epsilon$ for a sufficiently small δ and a sufficiently large n. Therefore, by Lemma 4.2.4, we have for a sufficiently small δ and $0 < \epsilon < 1$ there exists an $N_{\epsilon,\delta}$ such that for all $n > N_{\epsilon,\delta}$, we have a $\hat{\psi}$ with $g(\hat{\psi}) = 0$ with probability greater than $1 - \epsilon$. That is, (4.33) has a solution for $\|\mathbb{B} - \mathbb{B}^*\| < \delta$, say $\tilde{\mathbb{B}}$, and assuming the solution is unique then it will also be consistent for $\mathbb{B}^*$. □

4.2.4 Asymptotic properties of UE and RE

In this subsection, we present some asymptotic properties of the UE and RE using Theorem 4.2.1. We first establish the following theorem which gives the asymptotic distribution of the unrestricted estimator, $\hat{\mathbb{B}}$. In the sequel, we set $\boldsymbol{\epsilon}_{1,n} = \sqrt{n}(\hat{\mathbb{B}} - \mathbb{B}^*)$.

Theorem 4.2.3. *Suppose that Assumption 4.2.1 and Assumption 4.2.2 hold. Then,* $\boldsymbol{\epsilon}_{1,n} = \sqrt{n}(\hat{\mathbb{B}} - \mathbb{B}^*) \xrightarrow[n\to\infty]{D} \boldsymbol{\epsilon}_1 \sim \mathscr{N}_{p_1\times p_2\times\cdots\times p_d}\left(0,\ \boldsymbol{\Phi}^{-1}\boldsymbol{\Phi}^*\boldsymbol{\Phi}^{-1}\right)$.

Proof. It suffices to show that $\sqrt{n}(\mathrm{Vec}(\hat{\mathbb{B}}) - \mathrm{Vec}(\mathbb{B}^*)) \xrightarrow[n\to\infty]{D} \mathrm{Vec}(\boldsymbol{\epsilon}_1)$ with $\mathrm{Vec}(\boldsymbol{\epsilon}_1) \sim \mathscr{N}_{p_1p_2\cdots p_d}\left(0, \boldsymbol{\Phi}^{-1}\boldsymbol{\Phi}^*\boldsymbol{\Phi}^{-1}\right)$. Since $\hat{\mathbb{B}}$ is an estimator, it satisfies (4.8). We expand $\mathcal{G}_n(\mathbb{B})$ in a Taylor series around $\mathrm{Vec}(\mathbb{B}^*)$ and then, together with the consistency of $\hat{\mathbb{B}}$, we get

$$\mathcal{G}_n(\mathbb{B}) = \mathcal{G}_n(\mathbb{B}^*) - \boldsymbol{C}_n(\mathbb{B}^*)\left(\mathrm{Vec}(\mathbb{B}) - \mathrm{Vec}(\mathbb{B}^*)\right) + o_p(\mathscr{E}_n) = 0, \tag{4.34}$$

where $\mathscr{E}_n = \|\hat{\mathbb{B}} - \mathbb{B}^*\|$. By rearranging the expansion, we have $\boldsymbol{C}_n(\mathbb{B}^*)\left(\mathrm{Vec}(\hat{\mathbb{B}}) - \mathrm{Vec}(\mathbb{B}^*)\right) = \mathcal{G}_n(\mathbb{B}^*) - o_p(\mathscr{E}_n)$. Then,

$$\mathrm{Vec}(\hat{\mathbb{B}}) - \mathrm{Vec}(\mathbb{B}^*) = \boldsymbol{C}_n^{-1}(\mathbb{B}^*)\mathcal{G}_n(\mathbb{B}^*) - \boldsymbol{C}_n^{-1}o_p(\mathscr{E}_n).$$

Then, we have $\sqrt{n}(\mathrm{Vec}(\hat{\mathbb{B}}) - \mathrm{Vec}(\mathbb{B}^*)) = n\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\frac{1}{\sqrt{n}}\mathcal{G}_n(\mathbb{B}^*) + o_p(1)$ and by Condition (2) of Assumption 4.2.2, Theorem 4.2.1 and Slutsky's theorem, we get $n\boldsymbol{C}_n^{-1}\frac{1}{\sqrt{n}}\mathcal{G}_n(\mathbb{B}^*) \xrightarrow[n\to\infty]{D} \boldsymbol{\Phi}^{-1}\boldsymbol{U}_0$, where $\boldsymbol{U}_0 \sim \mathscr{N}_{p_1p_2\cdots p_d}(0, \boldsymbol{\Phi}^*)$. Therefore, we get that $\sqrt{n}(\mathrm{Vec}(\hat{\mathbb{B}}) - \mathrm{Vec}(\mathbb{B}^*)) \xrightarrow[n\to\infty]{D} \mathrm{Vec}(\boldsymbol{\epsilon}_1)$, where $\mathrm{Vec}(\boldsymbol{\epsilon}_1) \sim \mathscr{N}_{p_1p_2\cdots p_d}(0, \boldsymbol{\Phi}^{-1}\boldsymbol{\Phi}^*\boldsymbol{\Phi}^{-1})$. This completes the proof. □

Next, we derive the asymptotic distribution for the restricted estimator in the case where the general constraint in (4.3) may not hold. Specifically, we consider the following sequence of local alternative constraints

$$\hbar(\mathbb{B}^*) = \frac{\mathrm{Vec}(\boldsymbol{r}_0)}{\sqrt{n}}, \quad n = 1, 2, \ldots \tag{4.35}$$

where $\boldsymbol{r}_0$ is an $l_1 \times l_2 \times \cdots \times l_d$ non-random tensor. For simplicity, let

$$\boldsymbol{\epsilon}_{2,n} = \sqrt{n}(\tilde{\mathbb{B}} - \mathbb{B}^*), \boldsymbol{\epsilon}_{3,n} = \sqrt{n}(\hat{\mathbb{B}} - \tilde{\mathbb{B}}), \boldsymbol{H}(\mathbb{B}^*) = \bigotimes_{i=d}^{1} \boldsymbol{H}_i(\mathbb{B}^*),$$

$$\boldsymbol{J}_{1,n} = \boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)' \left[\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'\right]^{-1} \boldsymbol{H}(\mathbb{B}^*),$$

$$\boldsymbol{J}_{2,n} = \boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)' \left[\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'\right]^{-1},$$

$$\boldsymbol{J} = \boldsymbol{\Phi}^{-1}\boldsymbol{H}(\mathbb{B}^*)' \left[\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{\Phi}^{-1}\boldsymbol{H}(\mathbb{B}^*)'\right]^{-1} \boldsymbol{H}(\mathbb{B}^*).$$

To derive the asymptotic distribution of $\boldsymbol{\epsilon}_{2,n}$ under the sequence of local alternative constraints in (4.35), we first establish the following proposition.

Proposition 4.2.1. *Suppose that the conditions of Theorem 4.2.3 and Lemma 4.2.6 hold along with the sequence of alternative constraints in* (4.35). *Then,*

$$\text{Vec}\,(\boldsymbol{\epsilon}_{2,n}) \;=\; \left[\bigotimes_{i=d}^{1} \boldsymbol{I}_{p_i} - \boldsymbol{J}_{1,n}\right]\text{Vec}\,(\boldsymbol{\epsilon}_{1,n}) - \boldsymbol{J}_{2,n}\text{Vec}(\boldsymbol{r}_0) + o_p(1). \tag{4.36}$$

Proof. From Theorem 4.2.1, we have $\frac{1}{n}\mathcal{G}_n(\mathbb{B}) \xrightarrow[n\to\infty]{\text{P}} 0$ and by Lemma 4.2.6, $\tilde{\mathbb{B}}$ is a consistent estimator of the true tensor parameter, $\mathbb{B}^*$. Next, by using Lagrangian multipliers, the RE $\tilde{\mathbb{B}}$ satisfies

$$\mathcal{G}_n(\mathbb{B}) + H(\mathbb{B})'\boldsymbol{\lambda}_n = 0, \tag{4.37}$$

for some non-random vector $\boldsymbol{\lambda}_n$ of size $l_1 l_2 \cdots l_d$. The Taylor expansion of $\hbar(\tilde{\mathbb{B}})$ gives

$$\hbar(\tilde{\mathbb{B}}) = \hbar(\mathbb{B}^*) + \boldsymbol{H}(\mathbb{B}^*)(\text{Vec}(\tilde{\mathbb{B}}) - \text{Vec}(\mathbb{B}^*)) + o_p(\tilde{\mathscr{E}}_n) = 0, \tag{4.38}$$

with $\tilde{\mathscr{E}}_n = \|\tilde{\mathbb{B}} - \mathbb{B}^*\|$. Similarly, by expanding (4.37) and using (4.38), we have

$$\mathcal{G}_n(\mathbb{B}^*) - \boldsymbol{C}_n(\mathbb{B}^*)(\text{Vec}(\tilde{\mathbb{B}}) - \text{Vec}(\mathbb{B}^*)) + \boldsymbol{H}(\mathbb{B}^*)'\boldsymbol{\lambda}_n + o_p(\tilde{\mathscr{E}}_n^*) = 0, \tag{4.39}$$

with $\tilde{\mathscr{E}}_n^* = \|\tilde{\mathbb{B}} - \mathbb{B}^*\| + \|\boldsymbol{\lambda}_n\|$. Further, from (4.38), we get

$$\boldsymbol{H}(\mathbb{B}^*)(\text{Vec}(\tilde{\mathbb{B}}) - \text{Vec}(\mathbb{B}^*)) = -\hbar(\mathbb{B}^*) + o_p(\tilde{\mathscr{E}}_n). \tag{4.40}$$

Multiplying (4.39) by $\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)$ gives

$$\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\mathcal{G}_n(\mathbb{B}^*) - \boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{C}_n(\mathbb{B}^*)(\mathrm{Vec}(\tilde{\mathbb{B}}) - \mathrm{Vec}(\mathbb{B}^*))$$
$$+\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'\boldsymbol{\lambda}_n + \boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)o_p(\tilde{\mathscr{E}}_n^*) = 0.$$

Using (4.40), we get

$$\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\mathcal{G}_n(\mathbb{B}^*) + \boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'\boldsymbol{\lambda}_n + \boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)o_p(\tilde{\mathscr{E}}_n^*)$$
$$+\hbar(\mathbb{B}^*) + o_p(\tilde{\mathscr{E}}_n) = 0.$$

Setting $\hbar(\mathbb{B}^*) = \frac{\mathrm{Vec}(\boldsymbol{r}_0)}{\sqrt{n}}$, we get

$$\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\mathcal{G}_n(\mathbb{B}^*) + \boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'\boldsymbol{\lambda}_n + \boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)o_p(\tilde{\mathscr{E}}_n^*)$$
$$+\frac{\mathrm{Vec}(\boldsymbol{r}_0)}{\sqrt{n}} + o_p(\tilde{\mathscr{E}}_n) = 0.$$

Since $\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'$ is non-singular a.s, then we get

$$\begin{aligned}\boldsymbol{\lambda}_n &= -\boldsymbol{J}_{2,n}'\mathcal{G}_n(\mathbb{B}^*) - \left[\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'\right]^{-1}\frac{\mathrm{Vec}(\boldsymbol{r}_0)}{\sqrt{n}} - \boldsymbol{J}_{2,n}'o_p(\tilde{\mathscr{E}}_n^*)\\ &\quad - \left[\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'\right]^{-1} o_p(\tilde{\mathscr{E}}_n), \end{aligned} \tag{4.41}$$

with $\boldsymbol{J}_{2,n} = \boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'\left[\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'\right]^{-1}$. Substituting (4.41) into (4.39), we have

$$\mathcal{G}_n(\mathbb{B}^*) - \boldsymbol{C}_n((\tilde{\mathbb{B}}^*)\mathrm{Vec}\,(\boldsymbol{\epsilon}_{2,n})\,/\,\sqrt{n} - \boldsymbol{J}_{1,n}\mathcal{G}_n(\mathbb{B}^*) - \boldsymbol{C}_n(\mathbb{B}^*)\boldsymbol{J}_{2,n}\frac{\mathrm{Vec}(\boldsymbol{r}_0)}{\sqrt{n}} - \boldsymbol{C}_n(\mathbb{B}^*)\boldsymbol{J}_{2,n}o_p(\tilde{\mathscr{E}}_n)$$
$$-\boldsymbol{J}_{1,n}o_p(\tilde{\mathscr{E}}_n^*) + o_p(\tilde{\mathscr{E}}_n) = 0$$

where $\boldsymbol{J}_{1,n} = \boldsymbol{H}(\mathbb{B}^*)'\left[\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{H}(\mathbb{B}^*)'\right]^{-1}\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{C}_n^{-1}(\mathbb{B}^*)$. Solving for $\mathrm{Vec}\,(\boldsymbol{\epsilon}_{2,n})$, we get

$$\begin{aligned}\mathrm{Vec}\,(\boldsymbol{\epsilon}_{2,n})\,/\,\sqrt{n} &= \boldsymbol{C}_n^{-1}(\mathbb{B}^*)\left[\bigotimes_{i=d}^{1}\boldsymbol{I}_{p_i} - \boldsymbol{J}_{1,n}\right]\mathcal{G}_n(\mathbb{B}^*) - \boldsymbol{J}_{2,n}\frac{\mathrm{Vec}(\boldsymbol{r}_0)}{\sqrt{n}} - \boldsymbol{J}_{2,n}o_p(\tilde{\mathscr{E}}_n)\\ &\quad - \boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{J}_{1,n}o_p(\tilde{\mathscr{E}}_n^*) + o_p(\tilde{\mathscr{E}}_n), \end{aligned} \tag{4.42}$$

and the result follows. □

Remark 4.2.2. *Note that if the constraint is*

$$\hbar^*(\mathbb{B}^*) = \mathbb{B}^* \times_1 \boldsymbol{R}_1 \times_2 \boldsymbol{R}_2 \times \cdots \times_d \boldsymbol{R}_d - \boldsymbol{r} = \frac{\boldsymbol{r}_0}{\sqrt{n}}, \tag{4.43}$$

where $\boldsymbol{R}_i$ are $l_i \times p_i$, non-random matrices with rank $l_i \leqslant p_i$, $i = 1, 2, \cdots, d$ and $\boldsymbol{r}$ and $\boldsymbol{r}_0$ are $l_1 \times l_2 \times \cdots \times l_d$ non-random tensors. Then, $\hbar(\mathbb{B}^*) = \bigotimes_{i=d}^{1} \boldsymbol{R}_i \mathrm{Vec}(\mathbb{B}^*) - \mathrm{Vec}(\boldsymbol{r}_0)$ and $\boldsymbol{H}(\mathbb{B}^*) = \bigotimes_{i=d}^{1} \boldsymbol{R}_i = \boldsymbol{R}$. Let $\boldsymbol{J}^*_{1,n} = \boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{R}' \left[\boldsymbol{R}\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{R}'\right]^{-1} \boldsymbol{R}$, $\boldsymbol{J}^*_{2,n} = \boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{R}' \left[\boldsymbol{R}\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{R}'\right]^{-1}$, $\boldsymbol{J}^* = \boldsymbol{\Phi}^{-1} R' \left[R\boldsymbol{\Phi}^{-1}R'\right]^{-1} R$. In this case, we get

$$\mathrm{Vec}\left(\boldsymbol{\epsilon}_{2,n}\right) \;=\; \left[\bigotimes_{i=d}^{1} I_{p_i} - \boldsymbol{J}^*_{1,n}\right] \mathrm{Vec}\left(\boldsymbol{\epsilon}_{1,n}\right) - \boldsymbol{J}^*_{2,n}\mathrm{Vec}(\boldsymbol{r}_0) + o_p(1). \tag{4.44}$$

We use Proposition 4.2.1 to derive the joint asymptotic distribution of the UE and the RE in the following theorem.

Theorem 4.2.4. *Suppose that the conditions of Proposition 4.2.1 hold, then*

$$\boldsymbol{\epsilon}_{1,n} \boxplus_{(d+1)} \boldsymbol{\epsilon}_{2,n} \boxplus_{(d+1)} \boldsymbol{\epsilon}_{3,n} \xrightarrow[n\to\infty]{d} \boldsymbol{\epsilon}_1 \boxplus_{(d+1)} \boldsymbol{\epsilon}_2 \boxplus_{(d+1)} \boldsymbol{\epsilon}_3 \sim \mathscr{N}_{p_1 \times p_2 \times \cdots \times 3p_d}\left(0 \boxplus_{(d+1)} \boldsymbol{\delta} \boxplus_{(d+1)} -\boldsymbol{\delta}, \boldsymbol{\Sigma}\right),$$

$$\text{where } \boldsymbol{\Sigma} = \begin{pmatrix} \tilde{\boldsymbol{\Phi}} & \tilde{\boldsymbol{\Phi}}(\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J}') & \tilde{\boldsymbol{\Phi}}\boldsymbol{J}' \\ (\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J})\tilde{\boldsymbol{\Phi}} & (\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J})\tilde{\boldsymbol{\Phi}}(\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J}') & (\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J})\tilde{\boldsymbol{\Phi}}\boldsymbol{J}' \\ \boldsymbol{J}\tilde{\boldsymbol{\Phi}} & \boldsymbol{J}\tilde{\boldsymbol{\Phi}}(\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J}') & \boldsymbol{J}\tilde{\boldsymbol{\Phi}}\boldsymbol{J}' \end{pmatrix}.$$

Proof. Let $\boldsymbol{J}^*_2 = \boldsymbol{\Phi}^{-1}\boldsymbol{H}(\mathbb{B}^*)' \left[\boldsymbol{H}(\mathbb{B}^*)\boldsymbol{\Phi}^{-1}\boldsymbol{H}(\mathbb{B}^*)'\right]^{-1}$. From Proposition 4.2.1, we have $\mathrm{Vec}(\boldsymbol{\epsilon}_{2,n}) = \sqrt{n}\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\left[\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J}_{1,n}\right]\mathcal{G}_n(\mathbb{B}^*) - \boldsymbol{J}_{2,n}\mathrm{Vec}(\boldsymbol{r}_0) + o_p(1)$, and $\mathrm{Vec}(\boldsymbol{\epsilon}_{3,n}) = \sqrt{n}\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{J}_{1,n}\mathcal{G}_n(\mathbb{B}^*) + o_p(1)$. Moreover, note that $\mathrm{Vec}(\boldsymbol{\epsilon}_{1,n} \boxplus_{(d+1)} \boldsymbol{\epsilon}_{2,n} \boxplus_{(d+1)} \boldsymbol{\epsilon}_{3,n}) = ((\mathrm{Vec}(\boldsymbol{\epsilon}_{1,n}))', (\mathrm{Vec}(\boldsymbol{\epsilon}_{2,n}))', (\mathrm{Vec}(\boldsymbol{\epsilon}_{3,n}))')'$, then

$$\begin{aligned}
&\mathrm{Vec}(\boldsymbol{\epsilon}_{1,n} \boxplus_{(d+1)} \boldsymbol{\epsilon}_{2,n} \boxplus_{(d+1)} \boldsymbol{\epsilon}_{3,n}) \\
&= \left((n\boldsymbol{C}_n^{-1}(\mathbb{B}^*), (n\boldsymbol{C}_n^{-1}(\mathbb{B}^*)[\boldsymbol{I} - \boldsymbol{J}_{1,n}])', (n\boldsymbol{C}_n^{-1}(\mathbb{B}^*)\boldsymbol{J}_{1,n})'\right)' \frac{1}{\sqrt{n}}\mathcal{G}_n(\mathbb{B}^*) \\
&+ (0, -(\boldsymbol{J}_2\mathrm{Vec}(\boldsymbol{r}_0))', (\boldsymbol{J}_2\mathrm{Vec}(\boldsymbol{r}_0))')' + o_p(1),
\end{aligned}$$

where, for the sake of simplicity, we set $\boldsymbol{I}_{p_1p_2\cdots p_d} = \boldsymbol{I}$. Hence, by Slutsky's theorem, we have $\mathrm{Vec}(\boldsymbol{\epsilon}_{1,n} \boxplus_{(d+1)} \boldsymbol{\epsilon}_{2,n} \boxplus_{(d+1)} \boldsymbol{\epsilon}_{3,n})' \xrightarrow[n\to\infty]{d} \left(\boldsymbol{I}, \boldsymbol{I} - \boldsymbol{J}', \boldsymbol{J}'\right)' \boldsymbol{\Phi}^{-1}\boldsymbol{U}_0 + \left(0, -(\boldsymbol{J}_2^*\mathrm{Vec}(\boldsymbol{r}_0))', (\boldsymbol{J}_2^*\mathrm{Vec}(\boldsymbol{r}_0))'\right)'$ with $\boldsymbol{U}_0 \sim \mathcal{N}_{p_1\cdots p_d}(0, \boldsymbol{\Phi}^*)$. The result follows. □

From Theorem 4.2.4, we derive the following corollary which gives the joint asymptotic distribution of $\boldsymbol{\epsilon}_{1,n}$, $\boldsymbol{\epsilon}_{2,n}$ and $\boldsymbol{\epsilon}_{3,n}$ in the special case where the sequence of local alternatives restriction is as in (4.43).

Corollary 4.2.1. *Under Assumption 4.2.2 and the sequence of local alternatives in* (4.43)*, we have* $\boldsymbol{\epsilon}_{1,n} \boxplus_{(d+1)} \boldsymbol{\epsilon}_{2,n} \boxplus_{(d+1)} \boldsymbol{\epsilon}_{3,n} \xrightarrow[n\to\infty]{d} \boldsymbol{\epsilon}_1 \boxplus_{(d+1)} \boldsymbol{\epsilon}_2 \boxplus_{(d+1)} \boldsymbol{\epsilon}_3$,
where $\boldsymbol{\epsilon}_1 \boxplus_{(d+1)} \boldsymbol{\epsilon}_2 \boxplus_{(d+1)} \boldsymbol{\epsilon}_3 \sim \mathcal{N}_{p_1\times p_2\times\cdots\times 3p_d}(0 \boxplus_{(d+1)} \boldsymbol{\delta}^* \boxplus_{(d+1)} -\boldsymbol{\delta}^*, \boldsymbol{\Sigma}^*)$ *and*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \tilde{\boldsymbol{\Phi}} & \tilde{\boldsymbol{\Phi}}(\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J}^{*'}) & \tilde{\boldsymbol{\Phi}}\boldsymbol{J}^{*'} \\ (\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J}^*)\tilde{\boldsymbol{\Phi}} & (\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J}^*)\tilde{\boldsymbol{\Phi}}(\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J}^{*'}) & (\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J}^*)\tilde{\boldsymbol{\Phi}}\boldsymbol{J}^{*'} \\ \boldsymbol{J}^*\tilde{\boldsymbol{\Phi}} & \boldsymbol{J}^*\tilde{\boldsymbol{\Phi}}(\boldsymbol{I}_{p_1p_2\cdots p_d} - \boldsymbol{J}^{*'}) & \boldsymbol{J}^*\tilde{\boldsymbol{\Phi}}\boldsymbol{J}^{*'} \end{pmatrix}.$$

Proof. The result follows from Theorem 4.2.4 by taking $\boldsymbol{H}_i(\mathbb{B}^*) = \boldsymbol{R}_i$, $i = 1, \cdots, d$. □

The results of Theorem 4.2.4 will be used in Section 4.3 to compare the relative efficiency of the proposed estimators.

4.3 A class of shrinkage tensor estimators and relative efficiency

In this section we propose shrinkage estimators in the context of the generalized tensor regressor model in (4.2). As such, the shrinkage estimators (SEs) for the model in (4.2) are obtained from the class of shrinkage estimators (3.13) of Section 3.3. Namely, following the notations of Section 2.2, the James-Stein and Positive rule James-Stein estimators are taken to be

$$\hat{\mathbb{B}}^{s} = \tilde{\mathbb{B}} + \left[1 - \left(\prod_{j=1}^{d} l_i - 2\right) \Big/ \left(n\|\hat{\mathbb{B}} - \tilde{\mathbb{B}}\|^2_{\{\hat{\Xi}_i, i=1,\cdots,d\}}\right)\right]\left(\hat{\mathbb{B}} - \tilde{\mathbb{B}}\right), \tag{4.45}$$

$$\hat{\mathbb{B}}^{sp} = \tilde{\mathbb{B}} + \max\left\{\left[1 - \left(\prod_{j=1}^{d} l_i - 2\right) \Big/ \left(n\|\hat{\mathbb{B}} - \tilde{\mathbb{B}}\|^2_{\{\hat{\Xi}_i, i=1,\cdots,d\}}\right)\right], 0\right\}\left(\hat{\mathbb{B}} - \tilde{\mathbb{B}}\right),$$

respectively, where $\hat{\Xi}_i = \boldsymbol{H}_i(\hat{\mathbb{B}})' \left[\boldsymbol{H}_i(\hat{\mathbb{B}}) n \boldsymbol{C}_n^{-1}(\hat{\mathbb{B}}) \boldsymbol{H}_i(\hat{\mathbb{B}})'\right]^{-1} \boldsymbol{H}_i(\hat{\mathbb{B}}), \quad i = 1, 2, \ldots, d$. In order to evaluate the performance of the proposed class of estimators, we compute their asymptotic distributional risk with respect to the loss function ℓ, denoted by ADR^{ℓ}. Let $\boldsymbol{W}_i,\ i = 1, \cdots, d$ be nonnegative definite matrices and let $\boldsymbol{W} = [\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_d]$. We consider the loss function $\mathrm{L}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}) = \ell\left(\mathrm{trace}\left(\rho^{*\prime}_{(d)}\rho^{*}_{(d)}\right)\right)$ for $\rho^{*} = \rho(\bigtimes_{i=1}^{d})_i \boldsymbol{W}_i^{1/2}$ where $\ell(t)$ is a non-negative, non-decreasing concave function on $(0, +\infty)$ such that $\ell'(t)$ exists and $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow[n\to\infty]{d} \rho$. Then, the ADR^{ℓ} of an estimator $\hat{\boldsymbol{\theta}}$ of a parameter θ is defined as

$$\mathrm{ADR}^{\ell}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}, \boldsymbol{W}) = \mathrm{E}\left[\ell\left(\mathrm{trace}\left(\rho^{*\prime}_{(d)}\rho^{*}_{(d)}\right)\right)\right]. \tag{4.46}$$

Remark 4.3.1. *For the special case where $\ell(t) \equiv t$, $t \geqslant 0$,* (4.46) *yields the usual asymptotic distributional risk with respect to the quadratic loss function as defined in Section 3.3, i. e.*

$$\mathrm{ADR}^{1}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}, \boldsymbol{W}) = \mathrm{E}\left[\mathrm{trace}\left(\rho^{*\prime}_{(d)}\rho^{*}_{(d)}\right)\right]. \tag{4.47}$$

To analyze the relative efficiency of the tensor estimators in (3.13) under loss function ℓ, we consider the case where $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$ such that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \boxplus_{(d+1)} \sqrt{n}\left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \boxplus_{(d+1)} \sqrt{n}\left(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\right) \xrightarrow[n\to\infty]{d} \boldsymbol{\vartheta}_1 \boxplus_{(d+1)} \boldsymbol{\vartheta}_2 \boxplus_{(d+1)} \boldsymbol{\vartheta}_3 \tag{4.48}$$

$\mathrm{Vec}(\boldsymbol{\vartheta}_2) = (\boldsymbol{I}_{p_1 p_2 \cdots p_d} - \boldsymbol{J}_0)\mathrm{Vec}(\boldsymbol{\vartheta}_1) + \mathrm{Vec}(\delta)$ and $\mathrm{Vec}(\boldsymbol{\vartheta}_3) = \boldsymbol{J}_0 \mathrm{Vec}(\boldsymbol{\vartheta}_1) - \mathrm{Vec}(\delta)$, where $\boldsymbol{J}_0$ is a matrix with the same structure as the matrix $\boldsymbol{J}$ given in Theorem 4.2.4. In the following lemma, we show that if the SEs dominate the UE under quadratic loss function, then they also dominate the UE under $l(\cdot)$. This is useful in deriving some conditions under which the

SEs dominate the UE under $l(\cdot)$. For some simplification, let
$\boldsymbol{\vartheta}_i^* = \boldsymbol{\vartheta}_i(\bigotimes_{i=1}^{d})_i \boldsymbol{W}_i^{1/2}$, $i = 1, 2, 3, 4$, let $\boldsymbol{I} = \boldsymbol{I}_{p_1 p_2 \cdots p_d}$ and let

$$
\begin{aligned}
g(\boldsymbol{\vartheta}_1^*) &= \left(\mathrm{Vec}(\boldsymbol{\vartheta}_1^*)\right)' (\boldsymbol{J}_0'\boldsymbol{J}_0 - \boldsymbol{J}_0' - \boldsymbol{J}_0)\left(\mathrm{Vec}(\boldsymbol{\vartheta}_1^*)\right) \\
&+ 2\left((\boldsymbol{I} - \boldsymbol{J}_0)\mathrm{Vec}(\boldsymbol{\vartheta}_1^*)\right)' \left(\boldsymbol{J}_0\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_1^*)\right) \\
&+(\boldsymbol{J}_0\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_1^*))' \left(\boldsymbol{J}_0\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_1^*)\right) + 2\left((\boldsymbol{I} - \boldsymbol{J}_0)\mathrm{Vec}(\boldsymbol{\vartheta}_1^*)\right)' \mathrm{Vec}(\delta) \\
&+\|\delta\|^2 - 2\left((\boldsymbol{I} - \boldsymbol{J}_0)\mathrm{Vec}(\boldsymbol{\vartheta}_1^*)\right)' \left(\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\delta)\right) \\
&-2(\boldsymbol{J}_0\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_1^*))' \left(\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\delta)\right) \\
&+ \left(\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\delta)\right)' \left(\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\delta)\right) \\
&+2\left(\mathrm{Vec}(\delta)\right)' \left[\left(\boldsymbol{J}_0\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_1^*)\right) - \left(\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\delta)\right)\right].
\end{aligned}
$$

Lemma 4.3.1. *Under* (4.48)*, we have*

$$
\mathrm{ADR}^{\ell}(\hat{\boldsymbol{\vartheta}}(h, \hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}), \boldsymbol{\theta}; \boldsymbol{W}) \leqslant \mathrm{ADR}^{\ell}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}) + \mathrm{E}\left[l'\left(\|\boldsymbol{\vartheta}_1^*\|^2\right) g(\boldsymbol{\vartheta}_1^*)\right].
$$

Proof. From Corollary 4.2.1, we get $\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}(h, \hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}) - \boldsymbol{\theta}\right) \xrightarrow[n\to\infty]{d} \boldsymbol{\vartheta}_4 = \boldsymbol{\vartheta}_2 + h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_3$. Let $\boldsymbol{\vartheta}_i^* = \boldsymbol{\vartheta}_i(\bigotimes_{i=1}^{d})_i \boldsymbol{W}_{\boldsymbol{i}}^{1/2}$, $i = 1, 2, 3, 4$. Then, $\mathrm{trace}\left(\boldsymbol{\vartheta}_{4(d)}^{*'}\boldsymbol{\vartheta}_{4(d)}^*\right) = \|\boldsymbol{\vartheta}_4^*\|^2$ and then

$$
\begin{aligned}
\mathrm{trace}\left(\boldsymbol{\vartheta}_{4(d)}^{*'}\boldsymbol{\vartheta}_{4(d)}^*\right) &= (\mathrm{Vec}(\boldsymbol{\vartheta}_2^* + h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_3^*))'(\mathrm{Vec}(\boldsymbol{\vartheta}_2^* + h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_3^*)) \\
&= \|\boldsymbol{\vartheta}_2^*\|^2 + 2\mathrm{Vec}(\boldsymbol{\vartheta}_2^*)'\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_3^*) \\
&+ (\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_3^*))'\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_3^*).
\end{aligned}
$$

Note that $\mathrm{Vec}(\boldsymbol{\vartheta}_2) = (\boldsymbol{I}_{p_1 p_2 \cdots p_d} - \boldsymbol{J}_0)\mathrm{Vec}(\boldsymbol{\vartheta}_1) + \mathrm{Vec}(\delta)$ and $\mathrm{Vec}(\boldsymbol{\vartheta}_3) = \boldsymbol{J}_0\mathrm{Vec}(\boldsymbol{\vartheta}_1) - \mathrm{Vec}(\delta)$, where $\boldsymbol{J}_0$ is a matrix of the same nature as $\boldsymbol{J}$. Then,

$$
\begin{aligned}
\mathrm{trace}\left(\boldsymbol{\vartheta}_{4(d)}^{*'}\boldsymbol{\vartheta}_{4(d)}^*\right) &= \left((\boldsymbol{I}_{p_1 p_2 \cdots p_d} - \boldsymbol{J}_0)\mathrm{Vec}(\boldsymbol{\vartheta}_1^*) + \mathrm{Vec}(\delta)\right)' \left((\boldsymbol{I}_{p_1 p_2 \cdots p_d} - \boldsymbol{J}_0)\mathrm{Vec}(\boldsymbol{\vartheta}_1^*) + \mathrm{Vec}(\delta)\right) \\
&+ 2\left((\boldsymbol{I}_{p_1 p_2 \cdots p_d} - \boldsymbol{J}_0)\mathrm{Vec}(\boldsymbol{\vartheta}_1^*) + \mathrm{Vec}(\delta)\right)' \left(\boldsymbol{J}_0\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_1^*) - \mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\delta)\right) \\
&- \mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\delta))' \left(\boldsymbol{J}_0\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_1^*) - \mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\delta)\right) \\
&+ (\boldsymbol{J}_0\mathrm{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi, i=1,\cdots,d\}})\boldsymbol{\vartheta}_1^*).
\end{aligned}
$$

Hence, setting $\boldsymbol{I}_{p_1p_2\cdots p_d} = \boldsymbol{I}$, we have

$$\begin{aligned}
&\text{trace}\left(\boldsymbol{\vartheta}^{**\prime}_{4(d)}\boldsymbol{\vartheta}^{*}_{4(d)}\right) = \|\boldsymbol{\vartheta}^*_1\|^2 + (\text{Vec}(\boldsymbol{\vartheta}^*_1))'\,(\boldsymbol{J}_0'\boldsymbol{J}_0 - \boldsymbol{J}_0' - \boldsymbol{J}_0)\,(\text{Vec}(\boldsymbol{\vartheta}^*_1)) \\
&+2\,((\boldsymbol{I}-\boldsymbol{J}_0)\text{Vec}(\boldsymbol{\vartheta}^*_1))'\,\text{Vec}(\delta) \\
&+(\text{Vec}(\delta))'(\text{Vec}(\delta)) + 2\,((\boldsymbol{I}-\boldsymbol{J}_0)\text{Vec}(\boldsymbol{\vartheta}^*_1))'\left(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1)\right) \\
&-2\,((\boldsymbol{I}-\boldsymbol{J}_0)\text{Vec}(\boldsymbol{\vartheta}^*_1))'\left(\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\delta)\right) \\
&+(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1))'\left(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1)\right) \\
&-2(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1))'\left(\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\delta)\right) \\
&+\left(\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\delta)\right)'\left(\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\delta)\right) \\
&+2\,(\text{Vec}(\delta))'\left[\left(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1)\right) - \left(\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\delta)\right)\right],
\end{aligned}$$

and then, $\text{trace}\left(\boldsymbol{\vartheta}^{**\prime}_{4(d)}\boldsymbol{\vartheta}^{*}_{4(d)}\right) = \|\boldsymbol{\vartheta}^*_1\|^2 + \left(\text{Vec}(\boldsymbol{\vartheta}^*_1)\right)'(\boldsymbol{J}_0'\boldsymbol{J}_0 - \boldsymbol{J}_0' - \boldsymbol{J}_0)\left(\text{Vec}(\boldsymbol{\vartheta}^*_1)\right)$

$$\begin{aligned}
&+ 2\left((\boldsymbol{I}-\boldsymbol{J}_0)\text{Vec}(\boldsymbol{\vartheta}^*_1)\right)'\left(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1)\right) \\
&+ (\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1))'\left(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1)\right) = \|\boldsymbol{\vartheta}^*_1\|^2 + g(\boldsymbol{\vartheta}^*_1),
\end{aligned}$$

where

$$\begin{aligned}
g(\boldsymbol{\vartheta}^*_1) \;=\;& (\text{Vec}(\boldsymbol{\vartheta}^*_1))'\,(\boldsymbol{J}_0'\boldsymbol{J}_0 - \boldsymbol{J}_0' - \boldsymbol{J}_0)\,(\text{Vec}(\boldsymbol{\vartheta}^*_1)) \\
&+2\,((\boldsymbol{I}-\boldsymbol{J}_0)\text{Vec}(\boldsymbol{\vartheta}^*_1))'\left(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1)\right) \\
&+(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1))'\left(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1)\right) \\
&+2\,((\boldsymbol{I}-\boldsymbol{J}_0)\text{Vec}(\boldsymbol{\vartheta}^*_1))'\,\text{Vec}(\delta) + (\text{Vec}(\delta))'(\text{Vec}(\delta)) \\
&-2\,((\boldsymbol{I}-\boldsymbol{J}_0)\text{Vec}(\boldsymbol{\vartheta}^*_1))'\left(\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\delta)\right) \\
&-2(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1))'\left(\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\delta)\right) \\
&+\left(\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\delta)\right)'\left(\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\delta)\right) \\
&+2\,(\text{Vec}(\delta))'\left[\left(\boldsymbol{J}_0\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\boldsymbol{\vartheta}^*_1)\right) - \left(\text{Vec}(h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi,i=1,\cdots,d\}})\delta)\right)\right].
\end{aligned}$$

Hence, $L(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) = l\left(\|\boldsymbol{\vartheta}_1^*\|^2 + g(\boldsymbol{\vartheta}_1^*)\right)$. Then, using the fact that $l(.)$ is a concave function, we have $l(t+y) \leqslant l(t) + yl'(t)$. Then $L(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) \leqslant l\left(\|\boldsymbol{\vartheta}_1^*\|^2\right) + l'\left(\|\boldsymbol{\vartheta}_1^*\|^2\right) g(\boldsymbol{\vartheta}_1^*)$. Hence, by taking expectations of both sides of the inequality, we have

$$\mathrm{ADR}^{\ell}(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) \leqslant \mathrm{ADR}^{\ell}(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) + \mathrm{E}\left[l'\left(\|\boldsymbol{\vartheta}_1^*\|^2\right) g(\boldsymbol{\vartheta}_1^*)\right].$$

This completes the proof. □

From Lemma 4.3.1, we derive the following result which shows that if $\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}})$ dominates asymptotically $\hat{\boldsymbol{\theta}}$ under quadratic loss function, it also dominates it asymptotically under the loss function $l(.)$.

Lemma 4.3.2. *Let $f(\|x\|^2)$ be the pdf of a $p_1 \times p_2 \times \cdots \times p_d$ random tensor $\boldsymbol{\vartheta}_1$. Suppose that* $\mathrm{ADR}^1(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) \leqslant \mathrm{ADR}^1(\hat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W})$ *where the expectation is taken with respect to a probability measure whose density is proportional to $f(\|\boldsymbol{x}\|^2)l'\left(\|\boldsymbol{x}\|^2\right)$. Then,* $\mathrm{ADR}^{\ell}(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) \leqslant \mathrm{ADR}^{\ell}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W})$.

Proof. From Lemma 4.3.1,

$$\mathrm{ADR}^{\ell}(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) \leqslant \mathrm{ADR}^{\ell}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W}) + \mathrm{E}\left[l'\left(\|\boldsymbol{\vartheta}_1^*\|^2\right)(g(\boldsymbol{\vartheta}_1^*))\right].$$

Adding and subtracting $\|\boldsymbol{\vartheta}_1^*\|^2$ to $g(\boldsymbol{\vartheta}_1^*)$, we get

$$\mathrm{ADR}^{\ell}(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) \leqslant \mathrm{ADR}^{\ell}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W}) + \mathrm{E}\left[l'\left(\|\boldsymbol{\vartheta}_1^*\|^2\right)\left(g(\boldsymbol{\vartheta}_1^*) + \|\boldsymbol{\vartheta}_1^*\|^2 - \|\boldsymbol{\vartheta}_1^*\|^2\right)\right].$$

Thus,
$\mathrm{ADR}^{\ell}(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) \leqslant \mathrm{ADR}^{\ell}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W}) + \mathrm{E}\left[l'\left(\|\boldsymbol{\vartheta}_1^*\|^2\right)\left(g(\boldsymbol{\vartheta}_1^*) + \|\boldsymbol{\vartheta}_1^*\|^2\right)\right]$
$- \mathrm{E}\left[l'\left(\|\boldsymbol{\vartheta}_1^*\|^2\right)\|\boldsymbol{\vartheta}_1^*\|^2\right]$, then
$\mathrm{ADR}^{\ell}(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) \leqslant \mathrm{ADR}^{\ell}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W}) + \mathrm{ADR}^1(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) - \mathrm{ADR}^1(\hat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W})$,
where ADR^1 is given as in (4.47) with the expectation taken with respect to a probability

measure whose density is proportional to $f(\|\boldsymbol{x}\|^2)l'\left(\|\boldsymbol{x}\|^2\right)$. Hence, whenever $\mathrm{ADR}^1(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) - \mathrm{ADR}^1(\hat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W}) \leqslant 0$, we have

$$\mathrm{ADR}^\ell(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\theta};\boldsymbol{W}) \leqslant \mathrm{ADR}^\ell(\hat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W}).$$

This completes the proof. □

In the following lemma, we give a result which shows that if $l'(\cdot)$ is completely monotonic then the condition in Lemma 4.3.2 is fulfilled in the case where $f(.)$ is a pdf of some random tensors with elliptically contoured distribution as defined in Definition 2.2.1.

Lemma 4.3.3. *Suppose $l'(\cdot)$ is a completely monotonic function and let $\mathbb{X}$ be a $d-$dimensional random tensor such that $\mathbb{X} \sim \mathcal{E}_{q_1\times q_2\times\cdots\times q_d}(\mathbb{M},\Psi;g)$, where g is such that $(-1)^j g^{(j)}(t) \geqslant 0$, for $j = 1,2,3,\cdots$ and $t \geqslant 0$ and let $f_{\mathbb{X}}$ denote the pdf of $\mathbb{X}$. Then, $f^*_{\mathbb{X}}(x) = k\ f_{\mathbb{X}}(\|x\|^2)l'(\|x\|^2),\ k > 0$, is also a pdf belonging to an elliptically contoured family of distributions.*

Proof. Since $(-1)^k g^{(k)}(t) \geq 0$, then $\mathbb{X}$ has a scale mixture of normal distributions as in Gómez-Sánchez-Manzano et al. (2006). Hence, $f_{\mathbb{X}}(\cdot)$ is completely monotonic and since $l'(t)$ is also completely monotonic, we get $f_{\mathbb{X}}(x)l'(x)$ is a completely monotonic function (see Berger (1975) and Kubokawa et al. (2015)). Thus, $f^*_{\mathbb{X}}(x) = k\ f_{\mathbb{X}}(x)l'(x),\ k \geq 0$ is a scale mixture of normal distributions. This completes the proof. □

Remark 4.3.2. *From Corollary 4.2.1, it can be noted that f_{ϵ_1} is normally distributed and thus, is also elliptically contoured. Then, by Lemma 4.3.3, for some $k > 0$, $kf_{\epsilon_1}(\|t\|^2)l'(\|t\|^2)$ would be a pdf of a family of elliptically contoured distributions provided that $l'(t)$ is completely monotonic.*

To compare the asymptotic distributional risks of the proposed estimators, we use the identities about quadratic forms of elliptically contoured distributions in Section 2.2. These identities are crucial in deriving the ADR of the UE, RE and SEs. Thereafter, we work under the following assumption.

Assumption 4.3.1. *The loss function $l(\cdot)$ is non-negative, non-decreasing and concave with $l'(t)$ a completely monotonic function such that $k\, f_{\epsilon_1}(t) l'(t)$, $k > 0$, is a pdf of $\vartheta_1^{**} \sim \mathscr{E}_{p_1\times\cdots\times p_d}\left(\mathbf{0}, \bigotimes_{i=d}^{1} \bar{\boldsymbol{\Phi}}_i^{-1}; g\right)$, for non-random positive definite matrices $\bar{\boldsymbol{\Phi}}_i$, $i = 1, \cdots, d$.*

We derive below the main result of Section 4.3 which shows that for a suitable choice of the weight matrices, $\boldsymbol{W}_i, i = 1, \cdots, d$, the SEs always dominate the UE. To introduce some notation, let $c_1^* = \text{trace}\left(\boldsymbol{\delta}'_{(d)}\left(\boldsymbol{\delta}\Big(\bigtimes_{j=1}^{d}\Big)_j \boldsymbol{\Xi}_j^* \bar{\boldsymbol{\Phi}}_j^{-1} \boldsymbol{W}_j\right)_{(d)}\right)$, $c_2^* = \prod_{j=1}^{d} \text{trace}(\boldsymbol{W}_j \bar{\boldsymbol{\Phi}}_j^{-1})$, $c_3^* = \prod_{i=1}^{d} \text{trace}(\boldsymbol{W}_i \bar{\boldsymbol{\Upsilon}}_i^*)$, $c_4^* = \text{trace}\left(\boldsymbol{\delta}_{(d)}^{*'} \boldsymbol{\delta}_{(d)}^*\right)$, $\bar{\boldsymbol{\Upsilon}}_i^* = \boldsymbol{J}_i \bar{\boldsymbol{\Phi}}_i^{-1}$, $\boldsymbol{J}_i = \bar{\boldsymbol{\Phi}}_i^{-1} \boldsymbol{H}_i'(\mathbb{B}^*)(\boldsymbol{H}_i(\mathbb{B}^*) \bar{\boldsymbol{\Phi}}_i^{-1} \boldsymbol{H}_i'(\mathbb{B}^*))^{-1} \boldsymbol{H}_i(\mathbb{B}^*)$. We also define $\bar{\boldsymbol{\Pi}}^{**} = \left(\bar{\boldsymbol{\Pi}}^* + \bar{\boldsymbol{\Pi}}^{*'}\right)/2$ where $\bar{\boldsymbol{\Pi}}^* = \bigotimes_{i=d}^{1} \boldsymbol{\Xi}_i^{*1/2}\left(4 \bigotimes_{i=d}^{1} \bar{\boldsymbol{\Upsilon}}_i^* + (l-2) \bigotimes_{i=d}^{1} \boldsymbol{J}_i \boldsymbol{R}_i \bar{\boldsymbol{\Upsilon}}_i^*\right) \bigotimes_{i=d}^{1} \boldsymbol{W}_i \bar{\boldsymbol{\Upsilon}}_i^* \boldsymbol{R}_i' \boldsymbol{J}_i' \boldsymbol{\Xi}_i^{*1/2}$. Let $\text{Ch}_{\max}(\boldsymbol{A})$ denote the maximum eigenvalue of a matrix $\boldsymbol{A}$. In the following theorem, we present the main result of this section. Specifically, we derive some sufficient conditions for the SEs to dominate the UE.

Theorem 4.3.1. *Suppose that Assumption 4.3.1 holds along with the conditions of Corollary C.3.1 where $c_2^* \geqslant \max\left\{\frac{c_3^*}{2}, \frac{\text{Ch}_{\max}(\bar{\Pi}^*)}{4}\right\}$. Then,*

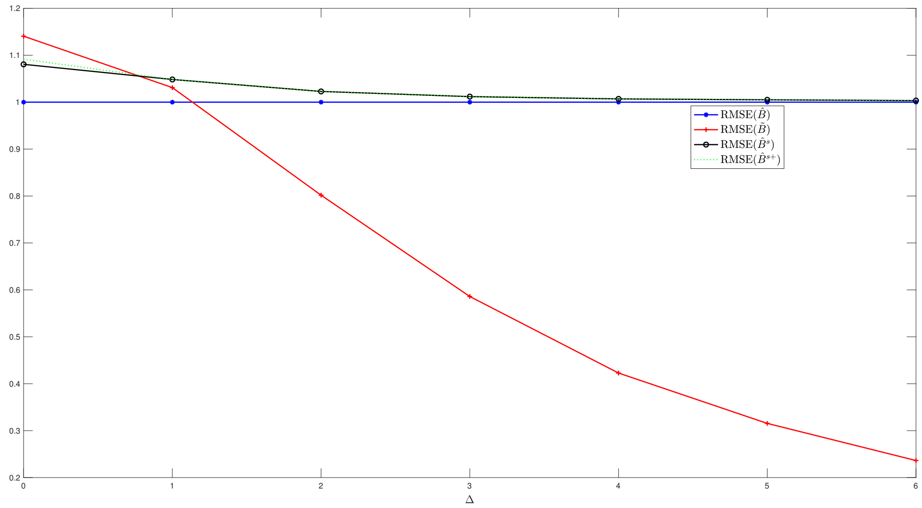
$$\text{ADR}^{\ell}(\hat{\mathbb{B}}^{sp}, \mathbb{B}^*; \boldsymbol{W}) \leqslant \text{ADR}^{\ell}(\hat{\mathbb{B}}^{s}, \mathbb{B}^*; \boldsymbol{W}) \leqslant \text{ADR}^{\ell}(\hat{\mathbb{B}}, \mathbb{B}^*; \boldsymbol{W}), \text{for all } \Delta \geqslant 0.$$

4.4 Simulation study and real data analysis

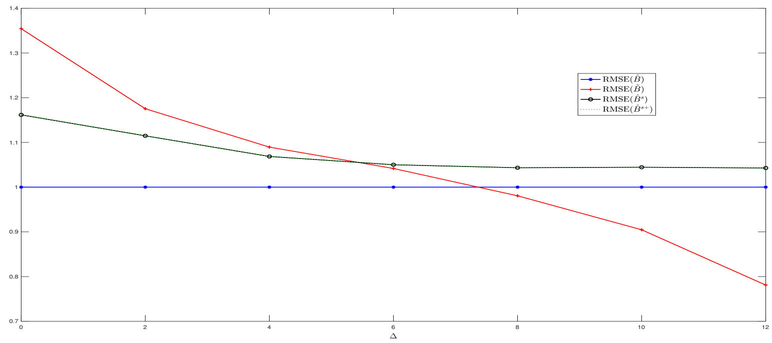
In this section, we present some simulation studies that illustrate the performance of the proposed methods. Further, we apply the proposed methods to two real datasets. In particular, we analyse a multi-relational network and a neuro-imaging datasets.

4.4.1 Simulation study

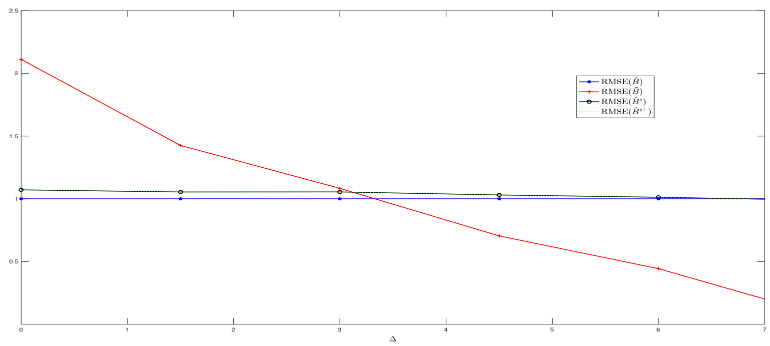
In this subsection, we present some simulation studies to support the theory. We consider the estimation of a square-centred image of ones surrounded by 0 signals. We set $d = 1$, $p_1 = 8$, $p_2 = 8$ and $p_3 = 1$. This gives an 8×8 parameter $\mathbb{B}$. We set the number of observations to be $n = 30$ and we set $q_1 = 20$, $q_2 = 20$, $q_3 = 30$. The mode-1 covariate matrix, X_1 is 20×8 generated from a uniform distribution on the interval (0,1). Similarly, the mode-2 covariate matrix is also 20×8 and the mode-3 covariate matrix is 30×1, both generated from a uniform distribution on the interval (0,1). The linear predictor of the model is then set as $\Theta = \mathbb{B} \times_1 X_1 \times_2 X_2 \times_3 X_3$ and stacked response $\mathbb{Y}$ is generated as Normal($\boldsymbol{\mu}$,1), Poisson($\boldsymbol{\mu}$) and Bin(1, $\boldsymbol{\mu}$) where $\boldsymbol{\mu}$ is set as Θ, e^{Θ} and $\frac{e^{\Theta_{ij}}}{1+e^{\Theta_{ij}}}$, $i = 1, \cdots, 8$, $j = 1, \cdots, 8$ for Normal, Poisson and Binomial distributions, respectively. We set the restriction matrices as $R_1 = I_8$, $R_2 = [I_3, 0_{3,5}]$ and $\boldsymbol{r_0}$ is set as the corresponding 8×2 matrix of zeros. This restriction sets the first three columns of the parameter to zero as it is suspected that the true parameter is a centred signal. In particular, R_1 selects all eight rows of the parameter and R_2 selects the first three columns of the parameter and this selection is set to zero through $\boldsymbol{r_0}$. We study the efficiency of the estimators by comparing the relative mean square error (RMSE) of each estimator with respect to the UE, $\hat{\mathbb{B}}$. The RMSE of some estimator, say $\hat{\mathbb{B}}^*$, with respect to $\hat{\mathbb{B}}$ is defined as $\text{RMSE}(\hat{\mathbb{B}}^*) = \text{ADR}\left(\hat{\mathbb{B}}, \mathbb{B}; \boldsymbol{W}\right) / \text{ADR}\left(\hat{\mathbb{B}}^*, \mathbb{B}; \boldsymbol{W}\right)$. We run the simulation with the $\boldsymbol{r}_0$ deviates away by units of $1/\sqrt{30}$, $1.5/\sqrt{30}$ and $2/\sqrt{30}$ for the Normal, Poisson and Bernoulli simulated data, respectively. For each restriction, we compute the UE, RE and SEs and replicate each run 1000 times to obtain the RMSE of each estimator. The results of the simulations are presented in Figure 4.1. As can be seen in Figure 4.1, as we deviate away from the restriction, the RE naturally fails at some point. However, the SEs still continue to perform at least as well as the UE despite the restriction being completely inaccurate. These results are confirmed for all the three distributions we have investigated and they corroborate the theory that SEs dominate the UE.



(a) Normal distribution



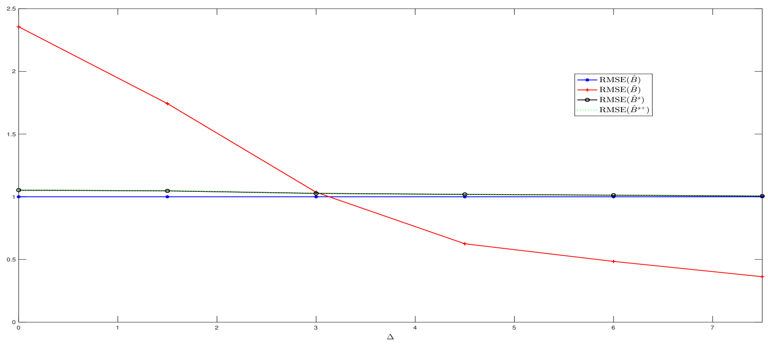
(b) Poisson distribution



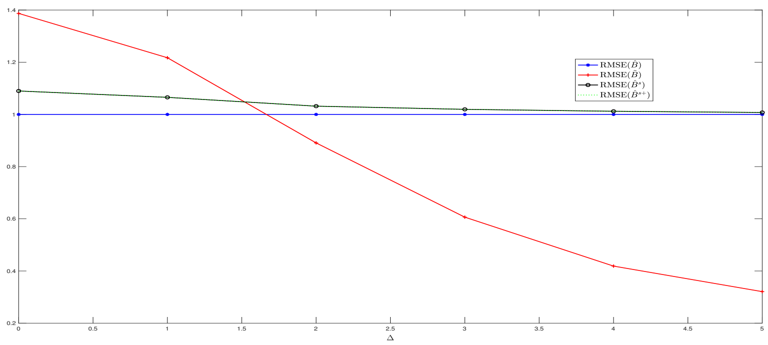
(c) Bernoulli distribution

Figure 4.1: RMSE versus Δ plot of the UE, the RE, and the SEs with square signal parameter under multi-mode covariates

Furthermore, we have noticed that the plot for the Bernoulli distribution is less dramatic than that of the Normal and Poisson data as we deviate away from the restriction. We performed some additional simulations to investigate a possible reason for the difference. Specifically, we ran the same simulation for Bernoulli with 3 covariate matrices but with $n = 100$ and we also ran a simulation for Bernoulli data with only one covariate matrix with $n = 30$. The results of these additional runs are presented in Figure 4.2. As can be seen, the case where $n = 100$ does not appear to improve the plot too much as compared to plot c) of Figure 4.1. However, the case where we estimate the Bernoulli data using only one covariate matrix results in a more clear descent as the restriction deviates. We suspect that this phenomenon may be affected by a certain interaction between the number of mode covariates in the model as well as the sample size. Perhaps as the number of mode covariates increases the sample size may need to also be increased. Overall, the simulation results corroborate our theoretical findings as established in Section 4.3.



(a) Bernoulli distribution with multimode covariates for $n = 100$



(b) Bernoulli distribution with mode 3 covariates for $n = 30$

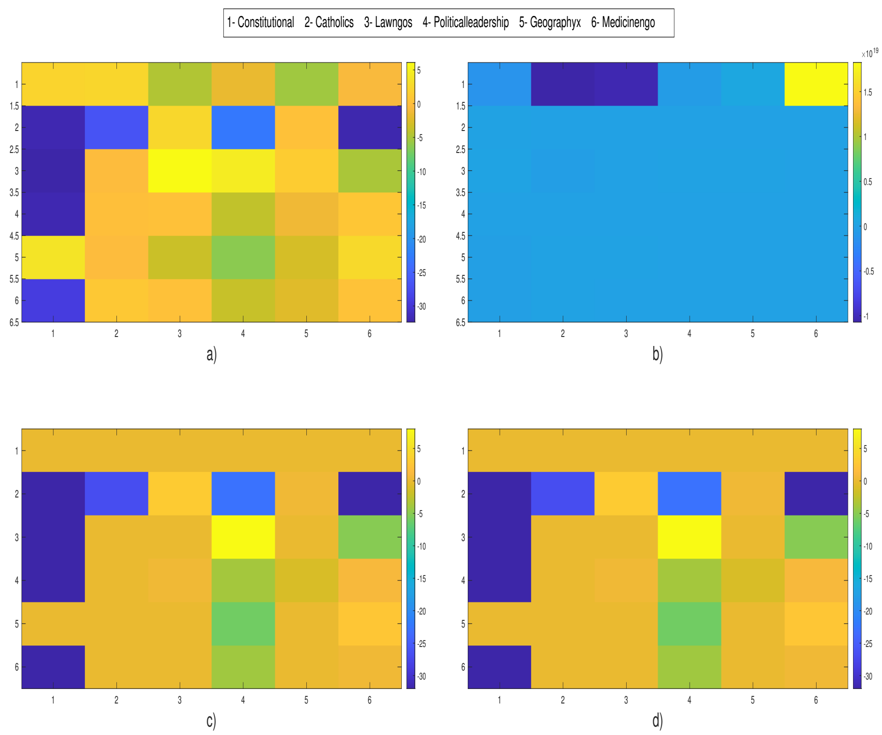
Figure 4.2: RMSE versus Δ plot of the UE, the RE, and the SEs with square signal parameter for Bernoulli data under multi-mode covariates and one-mode covariates

4.4.2 Real data analysis

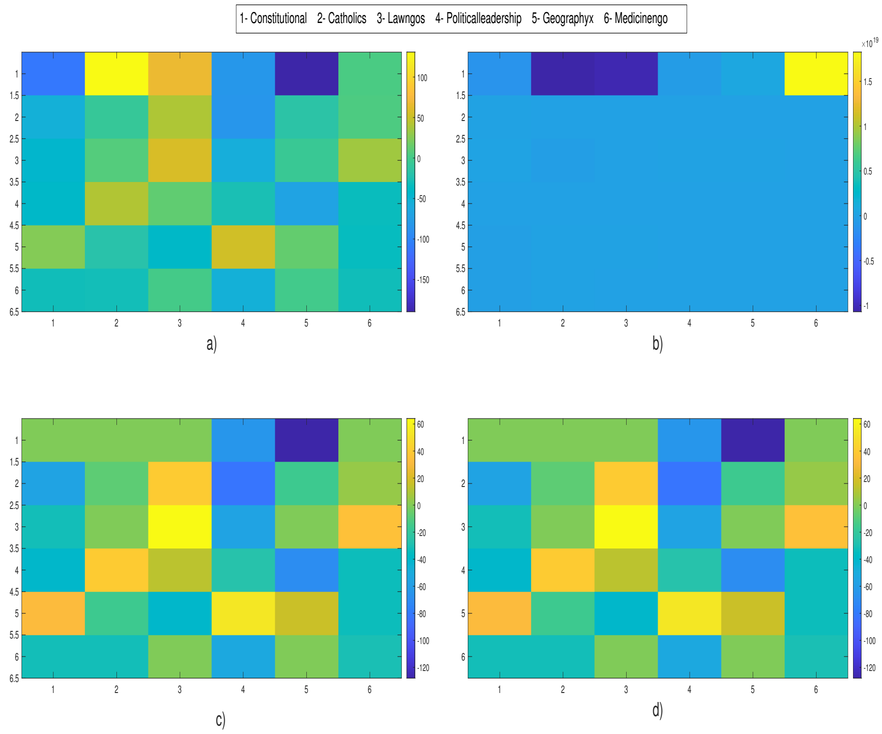
In this subsection, we apply the proposed methods to two datasets. First, we analyze a two dimensional parameter with three-dimensional data. Second, we analyze a neuro-imaging dataset which consists in observations collected on patients with schizophrenia disease.

The nations dataset

The nations dataset is composed of 56 recorded relations among 14 countries between 1950 and 1965. This relational network is summarized into a $14 \times 14 \times 56$ binary tensor where an entry of 1 indicates a connection and an entry of 0 indicates a lack of a connection between two countries for the particular relation. For more details on this dataset and relational networks, we refer the reader to Nickel et al. (2011). We take the $14 \times 14 \times 56$ binary data as the response tensor, $\mathbb{Y}$, and we take 6 country-level attributes as the mode-1 and mode-2 covariates. Namely, we take 'constitutional', 'catholics', 'lawngos','politicalleadership', 'geographyx' and 'medicinengo' as the covariates and form both the 14×6 mode-1 covariate and the 14×6 mode-2 covariate matrix, X_1 and X_2, respectively. These covariates are set as either 0 if the country does not have the attribute and 1 if it does. The tensor parameter $\mathbb{B}$ is thus a $6 \times 6 \times 56$, which exhibits the effects of a pair of covariates (i, j) on relation k. In addition, as Xu et al. (2019) noted, some relations belonging to a particular cluster exhibit similar covariate effects such as the 'economicaid' and 'warnings' relations, we add this using a restriction. Hence, to incorporate that possible prior knowledge, we set $R_1 = I_6$, $R_2 = I_6$, $R_3 = [1, 0_{1,9}, -1, 0_{1,45}]$ and $\boldsymbol{r_0}$ is a $6 \times 6 \times 1$ zero tensor. This sets the coefficients of the two relations to be equal. We run a bootstrap by resampling 500 times and at each iteration we obtain the UE, RE and SEs. The resulting RMSE with respect to the UE are 2.3307×10^{-38}, $1 + 0.12 \times 10^{-5}$ and $1 + 0.12 \times 10^{-5}$ for the RE, James-Stein estimator and Positive-rule James-Stein estimator, respectively. As the RMSE of the RE is very small, it can be concluded that the hypothesized restriction is inaccurate. However, even with such an incorrect prior information, the SEs still perform slightly better than the UE, further supporting our theory. We provide some visual plots of the effect estimation for each estimator in Figure 4.3 for the 'economicaid' and 'warnings' relations. From Figure 4.3, the coefficients of the RE are either very large or extremely small as compared to the effect maps of the other estimators. This indicates that the resulting estimation of the effects is questionable and that the restriction is most likely inaccurate. However, the effects of the SEs are more reasonable and indicate the covariate interactions with the strongest effects. For example, the interaction of 'medicinengo' with the 'catholics' covariate appears to have a strong positive effect on the 'economicaid' relation. In addition, the interactions of 'constitutional' covariate with the 'catholics' and 'lawngos' appears to have little effect on the 'warnings' relation.



(a) Economicaid covariate estimated coefficient



(b) Warnings covariate estimated coefficient

Figure 4.3: Plot of estimated coefficients for the economicaid and warnings covariates. a) UE, b) RE, c) James-Stein estimator, d) Positive Rule estimator

Schizophrenia dataset

The schizophrenia dataset is composed of a total of 174 preprocessed T1-weighted MRI images with 124 control subjects and 50 subjects diagnosed with schizophrenia. The dataset may be accessed at `https://openfmri.org/dataset/ds000030/`. Schizophrenia is a serious mental illness where an individual loses touch with reality. A rapidly growing interest in psychiatry and mental illness diagnosis is to be able to find regions of interest with respect to mental diseases and to potentially develop methods allowing for the diagnosis of psychiatric disorders by examining the brain of a patient. To estimate potential brain regions of interest, we set the stacked response, $\mathbb{Y}$ to be $30 \times 36 \times 36 \times 174$ tensor and we take diagnosis, gender and age to form the mode-3 covariates. We set the link function to be the identity. For schizophrenia, previous studies have suggested that anatomical abnormalities in the dorsolateral prefrontal cortex may play a role in an individual developing schizophrenia, see for example, Cheng et al. (2009). Hence, we use this prior information to set the restriction. In particular, we set $R_1 = [0_{2,3}, I_2, 0_{2,25}]$ be the 2×30 restriction on the first mode, $R_2 = [0_{7,23}, \boldsymbol{I}_7, 0_{7,6}]$ is the 7×36 restriction on the second mode and $R_3 = [0_{6,22}, \boldsymbol{I}_6, 0_{6,8}]$ is the 6×36 restriction on third mode, $R_4 = \boldsymbol{I}_3$ and $\boldsymbol{r}_0$ is the $2 \times 7 \times 6 \times 1$ zero tensor. The interpretation of the matrices is that R_1, R_2 and R_3 select the 3-dimensional location, in this case the dorsolateral prefrontal cortex, in the brain and R_4 signifies that we will set the effects of the covariates to zero. We present a visual representation of the restriction in Figure 4.4. We run the simulation using tensor glm under identity link by resampling the data 1000 times and at each iteration we obtain the estimators. The resulting RMSE of the RE, James-Stein estimator and Positive-rule James-Stein estimator are $1 + 0.174 \times 10^{-3}$, $1 + 0.119 \times 10^{-9}$ and $1 + 0.119 \times 10^{-9}$, respectively. As can be seen, the RMSE of the SEs is above 1, once again corroborating the established theoretical results. Moreover, the RMSE of the RE also performs better than the UE. This may suggest that there may not be much significant impact of the dorsolateral prefrontal cortex on having schizophrenia. However, in the future, further analysis would be recommended and the development of some inference techniques would be of great interest to study the significance of effects of the estimated brain region regions. We present the estimated regions of interest in Figure 4.5. As can be seen, the RE results in less regions that may affect the illness as compared with the UE. Furthermore, the difference between the UE and SEs is not clear at the surface of the skull, although internally there may be some interesting regions that the SEs cover.

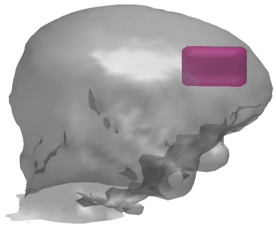
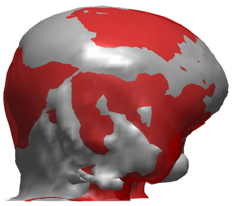
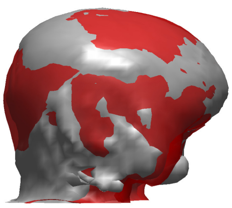


Figure 4.4: A visual representation of the restriction (pink). This is an approximate part of the dorsolateral prefrontal cortex

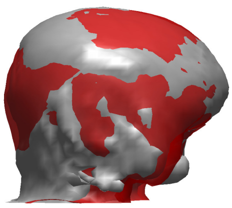


(a) UE estimated regions.

(b) RE estimated regions.



(c) James-Stein estimated regions.



(d) Positive James-Stein estimated regions.

Figure 4.5: Estimated regions (red) that may be associated with schizophrenia overlaid on a randomly-drawn subject (grey)

4.5 Conclusion

In this chapter, we studied an estimation problem about the tensor coefficient in a generalized tensor regression model with multi-mode covariates where the tensor parameter is suspected to satisfy some general constraint. We defined the estimating score function of the model and under some assumptions, we showed that estimating score function converges asymptotically to a normal vector. Using the estimating score function, we defined the proposed tensor estimators, the UE and RE. Moreover, we proved the existence and the consistency of the proposed estimators and under some weak dependence assumptions, we derived the joint asymptotic distribution of the UE and RE. In addition, we proposed the James-Stein and the Positive-rule Stein estimators in the context of this generalized model. We also defined the asymptotic distributional risk under a more general loss function which includes the ADR[1] of Chapter 3 as a special case. In a similar fashion to Chapter 3, we derived the conditions for the SEs to dominate the UE in terms of the ADR with respect to the general loss function. We corroborated those results by simulation studies on normal, binomial and Poisson data simulations and we analyzed a binary relational network and a schizophrenia neuro-imaging dataset.

Note that the results and methods of this chapter extend the methods of Chapter 3 when $m_0 = 0$. Namely, the link function of the model is no longer the identity link function on the tensor linear predictors and the dependence assumptions in Chapter 4 are weaker and imply the assumptions of Chapter 3. Further, we consider regressors on all modes of the tensor coefficient as opposed to just the mode $d + 1$ regressors of Chapter 3. This allows for studying the effects of covariates on different dimensions of the tensor coefficient. In addition, the restrictions, the SEs and the ADR of Chapter 3 are special cases of those in Chapter 4, respectively.

Chapter 5

Summary and Future Research

5.1 Summary

In this dissertation, we proposed two tensor regression models. The first model was a tensor regression model with multiple and possibly unknown number of change-points. Under the weakest dependence structure of the error and regressor terms, we generalized the assumptions and results of McLeish (1977) and Chen and Nkurunziza (2015, 2016). We also generalized the restriction imposed on the parameters of Chen and Nkurunziza (2015, 2016) by incorporating multi-mode restrictions on the tensor parameter. These restrictions allow for region selection, for example, and allow to set some prior knowledge or associations between brain regions and their effects on disease states. This was highlighted in the MRI data analyses of Chapters 3 and 4. We established a tensor $\mathcal{L}^2-$mixingale CLT-type theorem from which we derived the joint asymptotic distribution of the UE and RE. Using this joint asymptotic normality and the identities established in Section 2.1, we derived the joint asymptotic distributional risk of the proposed SEs, UE and RE. Shrinkage estimators are known to be especially advantageous in cases where the restriction is seriously violated as it still offers a good performance regardless of the validity of the restriction. On the other hand, the RE would not perform efficiently in the case where the restriction does not hold. Under the weak dependence structure, we derived some sufficient conditions for the SEs and the RE to dominate the UE. We also considered the case where the number of change-points is also unknown and proved that the methods give a consistent estimator of the number of change-points.

The second model that we consider in this dissertation is a generalized tensor regression model. This model focuses on data for which there are no change-points. However, this model is an extension of the model proposed in Chapter 3 as it replaces the identity link function in the change-point model of Chapter 3 with an arbitrary link function. Moreover, we consider a general constraint function on the tensor parameter of Chapter 4 which encloses the constraint of Chapter 3 as a special case. We set weaker assumptions on the error and regressor terms and show that the assumptions give the $\mathcal{L}^2-$mixingale assumption of Chapter 3. In addition, by using the estimating score function, we prove the existence and consistency of the resulting UE and RE as well as their joint asymptotic normality. Furthermore, we define the asymptotic distributional risk under a more general loss function than that considered in Chapter 3 and using the identities of Section 2.2.1, we prove that the SEs asymptotically dominate the UE with respect to the general loss function under some sufficient conditions. The general tensor regression model allows the extension of the methods to non-i.i.d, non-Gaussian tensor data such as binary tensor data. We corroborate the results and show examples of applications using the Nations relational network dataset (binary) and a schizophrenia MRI dataset (Gaussian).

5.2 Future research

In the future, we are interested in addressing the problem in the context where the number of change-points grows proportionally with T. Moreover, we seek to include some detailed data analysis in the case where the number of change-points is unknown. Specifically, we will investigate how our methods perform when the number of change-points is incorrectly estimated for complex data and what that would mean for the estimated locations and the performance of the proposed estimators. In addition, another point of interest would be to check how well our results perform in the case where the change-points are not clear-cut. Another improvement for our work would be to remove some redundant steps in the dynamic programming algorithm of Section B.4 as suggested by Chen et al. (2017), for example, in order to optimize the run-time.

An improvement is also to possibly investigate methods for a more precise restriction selection of brain regions for the neuro-imaging data analysis. A more precise region selection should further improve the efficiency of the restricted estimator as well as the shrinkage estimators.

Another point of interest is to incorporate methods which enable the model estimation to overcome the high-dimensionality problem. Common in tensor data, high-dimensionality is the case where there are far too many parameters to be estimated than the number of observations available. We wish to address this problem in the future and we look to study ways at which we may include some dimension reduction tools as in Xu et al. (2019) or tensor decomposition tools as in the block-relaxation algorithm of Zhou et al. (2013). The main challenge of such dimension-reduction tools is in establishing the joint asymptotic distributions of the resulting estimators. Another avenue to overcome this high-dimensionality problem is by using a procedure called the least absolute shrinkage and selection operator (Lasso) which was introduced by Tibshirani (1996). This procedure helps to identify the model with the most relevant parameters to explain underlying phenomena. We look to extend the results of Jandhyala et al. (2013) in the context of the proposed tensor regression model. Jandhyala et al. (2013) developed a Lasso-type method for a multiple change-point model for vector coefficients. Our extension would be interesting since we will be working with tensor coefficients under weak dependence assumptions.

Bibliography

J. Aitchison and S. Silvey. Maximum-likelihood estimation of parameters subject to restraints. *The annals of mathematical Statistics*, pages 813–828, 1958.

D. W. Andrews. Laws of large numbers for dependent non-identically distributed random variables. *Econometric theory*, 4(3):458–467, 1988.

J. A. Aston and C. Kirch. Evaluating stationarity via change-point alternatives with applications to fmri data. *The Annals of Applied Statistics*, pages 1906–1948, 2012.

J. Bai and P. Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78, 1998.

A. Batsidis. Robustness of the likelihood ratio test for detection and estimation of a mean change point in a sequence of elliptically contoured observations. *Statistics*, 44(1):17–24, 2010.

P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and R. C. Craddock. The neuro bureau adhd-200 preprocessed repository. *NeuroImage*, 144:275–286, 2017.

J. Berger. Minimax estimation of location vectors for a wide class of densities. *The Annals of Statistics*, pages 1318–1328, 1975.

P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 1968.

N. H. Bingham, R. Kiesel, et al. Semi-parametric modelling in finance: theoretical foundations. *Quantitative Finance*, 2(4):241–250, 2002.

B. B. Biswal, M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, et al. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107(10):4734–4739, 2010.

F. Chen and S. Nkurunziza. Optimal method in multiple regression with structural changes. *Bernoulli*, 21(4):2217–2241, 2015.

F. Chen and S. Nkurunziza. A class of stein-rules in multivariate regression model with structural changes. *Scandinavian Journal of Statistics*, 43(1):83–102, 2016. doi: 10.1111/sjos.12166.

F. Chen, R. Mamon, and M. Davison. Inference for a mean-reverting stochastic process with multiple change points. *Electronic Journal of Statistics*, 11(1):2199 – 2257, 2017.

D. Cheng, M. Bicego, U. Castellani, S. Cerruti, M. Bellani, G. Rambaldelli, M. Atzori, P. Brambilla, and V. Murino. Schizophrenia classification using regions of interest in brain mri. *Proceedings of Intelligent Data Analysis in Biomedicine and Pharmacology, IDAMAP*, 9:47–52, 2009.

K. C. Chu. Estimation and decision for linear systems with elliptical random processes. *IEEE Transactions on Automatic Control, AC*, 18:499–505, 1973.

J. Davidson. A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes. *Econometric theory*, pages 313–329, 1992.

E. Furman and Z. Landsman. Tail variance premium with applications for elliptical portfolio of risks. *ASTIN Bulletin: The Journal of the IAA*, 36(2):433–462, 2006.

C. Gallagher, R. Lund, and M. Robbins. Changepoint detection in daily precipitation data. *Environmetrics*, 23(5):407–419, 2012.

J. A. D. Garcia. Singular matrix variate skew-elliptical distribution and the distribution of general linear transformation. 2005.

M. Ghannam and S. Nkurunziza. The risk of tensor stein-rules in elliptically contoured distributions. *Statistics*, 56(2):421–454, 2022.

E. Gombay. Change detection in linear regression with time series errors. *The Canadian Journal of Statistics*, 38(1):65–79, 2010.

E. Gómez-Sánchez-Manzano, M. Gómez-Villegas, and J. Marín. Sequences of elliptical distributions and mixtures of normal distributions. *Journal of multivariate analysis*, 97 (2):295–310, 2006.

R. Guhaniyogi, S. Qamar, and D. B. Dunson. Bayesian tensor regression. *The Journal of Machine Learning Research*, 18(1):2733–2763, 2017.

A. Gupta and T. Varga. Normal mixture representations of matrix variate elliptically contoured distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 68–78, 1995.

P. D. Hoff. Multilinear tensor regression for longitudinal relational data. *The annals of applied statistics*, 9(3):1169, 2015.

S. Hossain, S. E. Ahmed, Y. Y. Grace, and C. B. Shrinkage and pretest estimators for longitudinal data analysis under partially linear models. *Journal of Nonparametric Statistics*, 28(3):531–549, 2016.

J. Jacod and A. Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media, 1987.

V. Jandhyala, S. Fotopoulos, I. MacNeill, and P. Liu. Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, 34(4):423–446, 2013.

G. G. Judge and M. E. Bock. *The statistical implications of pre-test and Stein-rule estimators in econometrics*, volume 25. North-Holland, 1978.

T. G. Kolda. Multilinear operators for higher-order decompositions. Technical Report SAND2006-2081, Sandia National Laboratories, April 2006.

T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51 (3):455–500, 2009.

T. Kubokawa, É. Marchand, and W. E. Strawderman. On improved shrinkage estimators for concave loss. *Statistics & Probability Letters*, 96:241–246, 2015.

Z. M. Landsman and E. A. Valdez. Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4):55–71, 2003.

B. Li, M. K. Kim, N. Altman, et al. On dimension folding of matrix-or array-valued statistical objects. *The Annals of Statistics*, 38(2):1094–1121, 2010.

L. Li and X. Zhang. Parsimonious tensor response regression. *Journal of the American Statistical Association*, 112(519):1131–1146, 2017.

X. Li, D. Xu, H. Zhou, and L. Li. Tucker tensor regression and neuroimaging analysis. *Statistics in Biosciences*, 10(3):520–545, 2018.

J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan. Generalized tensor regression for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2): 1244–1258, 2019.

J. S. Liu, W. C. Ip, and H. Wong. Predictive inference for singular multivariate elliptically contoured distributions. *Journal of multivariate analysis*, 100(7):1440–1446, 2009.

E. F. Lock. Tensor-on-tensor regression. *Journal of Computational and Graphical Statistics*, 27(3):638–647, 2018.

A. Mathai and S. Provost. *Quadratic Forms in Random Variables: Theory and Applications*, volume 87. Marcel Dekke: New York, 1992.

D. L. McLeish. On the invariance principle for nonstationary mixingales. *Ann. Probab.*, 5 (4):616–621, 08 1977. doi: 10.1214/aop/1176995772.

M. Nickel, V. Tresp, and H.-P. Kriegel. A three-way model for collective learning on multi-relational data. In *Icml*, 2011.

S. Nkurunziza. The risk of pretest and shrinkage estimators. *Statistics*, 46(3):305–312, 2012.

S. Nkurunziza. The bias and risk functions of some stein-rules in elliptically contoured distributions. *Mathematical Methods of Statistics*, 22(1):70–82, 2013.

S. Nkurunziza, K. Fu, et al. Improved inference in generalized mean-reverting processes with multiple change-points. *Electronic Journal of Statistics*, 13(1):1400–1442, 2019.

W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols. *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.

Z. Qu and P. Perron. Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502, 2007.

G. Raskutti, M. Yuan, and H. Chen. Convex regularization for high-dimensional multi-response tensor regression. *arXiv preprint arXiv:1512.01215*, 2015.

M. Robbins, R. Lund, and C. Gallagher. Changepoints in the north atlantic tropical cyclone record. *Journal of the American Statistical Association*, 106(493):89–99, 2011.

S. Roy, Y. Atchadé, and G. Michaelidis. Change point estimation in high dimensional markov random-field models. *J. R. Statist. Soc. B.*, 79(4):1187–1206, 2017.

A. M. E. Saleh. *Theory of preliminary test and Stein-type estimation with applications*, volume 517. John Wiley & Sons, 2006.

G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 03 1978. doi: 10.1214/aos/1176344136.

P. K. Sen and A. K. M. E. Saleh. On some shrinkage estimators of multivariate location. *Annals of Statistic*, 13:272–281, 1985.

M. Skup, H. Zhu, and H. Zhang. Multiscale adaptive marginal analysis of longitudinal neuroimaging data with time-varying covariates. *Biometrics*, 68(4):1083–1092, 2012.

M. V. Solanto, K. P. Schulz, J. Fan, C. Y. Tang, and J. H. Newcorn. Event-related fmri of inhibitory control in the predominantly inattentive and combined subtypes of adhd. *Journal of Neuroimaging*, 19(3):205–212, 2009.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

H. Tobing and C. McGilchrist. Recursive residuals for multivariate regression models. *Australian Journal of Statistics*, 34(2):217–232, 1992.

R. C. Wolf, M. M. Plichta, F. Sambataro, A. J. Fallgatter, C. Jacob, K.-P. Lesch, M. J. Herrmann, C. Schönfeldt-Lecuona, B. J. Connemann, G. Grön, et al. Regional brain activation changes and abnormal functional connectivity of the ventrolateral prefrontal cortex during working memory processing in adults with attention-deficit/hyperactivity disorder. *Human brain mapping*, 30(7):2252–2266, 2009.

J. Woody and R. B. Lund. A linear regression model with persistent level shifts: An alternative to infill asymptotics. *Statistics & Probability Letters*, 95(1):118–124, 2014.

Z. Xu, J. Hu, and M. Wang. Generalized tensor regression with covariates on multiple modes. *arXiv preprint arXiv:1910.09499*, 2019.

Z. Yu-Feng, H. Yong, Z. Chao-Zhe, C. Qing-Jiu, S. Man-Qiu, L. Meng, T. Li-Xia, J. Tian-Zi, and W. Yu-Feng. Altered baseline brain activity in children with adhd revealed by resting-state functional mri. *Brain and Development*, 29(2):83–91, 2007.

X. Zhang and L. Li. Tensor envelope partial least-squares regression. *Technometrics*, 59 (4):426–436, 2017.

H. Zhou, L. Li, and H. Zhu. Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552, 2013.

Appendix A

Some Useful Identities

Proof of Theorem 2.1.1. For $j = 1, \cdots, d+1$, let $\Xi_j^{1/2\dagger}$ be the Moore-Penrose pseudo-inverse of $\Xi_j^{1/2}$. Then, by the definition of the Moore-Penrose pseudo-inverse, we have $\Xi_j^{1/2\dagger}\Xi_j^{1/2}\Xi_j^{1/2\dagger} = \Xi_j^{1/2\dagger}$; $\Xi_j^{1/2}\Xi_j^{1/2\dagger}\Xi_j^{1/2} = \Xi_j^{1/2}$, $\left(\Xi_j^{1/2\dagger}\Xi_j^{1/2}\right)' = \Xi_j^{1/2\dagger}\Xi_j^{1/2}$; and $\left(\Xi_j^{1/2}\Xi_j^{1/2\dagger}\right)' = \Xi_j^{1/2}\Xi_j^{1/2\dagger}$. Further, $\Lambda_j\Xi_j$ are idempotent matrices for $j = 1, \cdots, d+1$ and note that $\Xi_j^{1/2}\Lambda_j\Xi_j^{1/2}$ is a symmetric and idempotent matrix for all j. Hence, the Kronecker product $\bigotimes_{j=d+1}^{1} \Xi_j^{1/2}\Lambda_j\Xi_j^{1/2}$ is also symmetric and idempotent. As a result, there exists an orthogonal matrix $\mathscr{O}$ such that $\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\Lambda_j\Xi_j^{1/2} = \mathscr{O}'\begin{pmatrix} I_{l_1\cdots l_{d+1}} & 0 \\ 0 & 0 \end{pmatrix}\mathscr{O}$. Let

$$V = \mathscr{O}\left(\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\right)\text{Vec}(X),$$

let $q = \prod_{i=1}^{d+1} q_i$ and let $l = \prod_{i=1}^{d+1} l_i$. Then, $V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim \mathcal{N}_{(m+1)q_1\cdots q_{d+1}}\left(\begin{pmatrix} \mathbb{M}_{V_1} \\ \mathbb{M}_{V_2} \end{pmatrix}, \begin{pmatrix} I_l & 0 \\ 0 & 0 \end{pmatrix}\right)$, where $\mathbb{M}_{V_1} = [I_l, 0]\, \mathscr{O} \left(\bigotimes_{i=d+1}^{1} \Xi_i^{1/2}\right) \mathrm{Vec}(\mathbb{M})$. Then, we have,

$$\|\mathbb{M}_{V_2}\|^2 = \mathbb{M}_{V_2}'\mathbb{M}_{V_2} = (\mathrm{Vec}(\mathbb{M}))' \bigotimes_{j=d+1}^{1} \Xi_j^{1/2} \mathscr{O}' \begin{pmatrix} 0 & 0 \\ 0 & I_{q-l} \end{pmatrix} \mathscr{O} \bigotimes_{j=d+1}^{1} \Xi_j^{1/2} (\mathrm{Vec}(\mathbb{M}))$$

$$= (\mathrm{Vec}(\mathbb{M}))' \bigotimes_{j=d+1}^{1} \Xi_j^{1/2} \left(I_q - \mathscr{O}' \begin{pmatrix} I_l & 0 \\ 0 & 0 \end{pmatrix} \mathscr{O}\right) \bigotimes_{j=d+1}^{1} \Xi_j^{1/2} (\mathrm{Vec}(\mathbb{M})).$$

Then, since $\Xi_j \Lambda_j \Xi_j = \Xi_j$, by multiplying through the parenthesis, we have

$$\bigotimes_{j=d+1}^{1} \Xi_j^{1/2} \left(I_q - \mathscr{O}' \begin{pmatrix} I_l & 0 \\ 0 & 0 \end{pmatrix} \mathscr{O}\right) \bigotimes_{j=d+1}^{1} \Xi_j^{1/2} = \bigotimes_{j=d+1}^{1} \Xi_j - \bigotimes_{j=d+1}^{1} \Xi_j \Lambda_j \Xi_j = 0.$$

Hence, $\|\mathbb{M}_{V_2}\| = 0$ which implies $\mathbb{M}_{V_2} = 0$. Then, since $\mathbb{M}_{V_2} = 0$ and $\Sigma_{V_2} = 0$, we get $V_2 = 0$ with probability 1. As a result, we have

$$\mathrm{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right) = (\mathrm{Vec}(\mathbb{X}))' \bigotimes_{j=d+1}^{1} \Xi_j^{1/2} \mathscr{O}'\mathscr{O} \bigotimes_{j=d+1}^{1} \left(\Xi_j^{1/2} \Lambda_j \Xi_j^{1/2}\right) \mathscr{O}'\mathscr{O} \bigotimes_{j=d+1}^{1} \Xi_j^{1/2} (\mathrm{Vec}(\mathbb{X})).$$

This gives

$$\mathrm{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right) = V' [I_l, 0]' [I_l, 0]\, V = V_1' V_1. \tag{A.1}$$

Further, one can verify that

$$X(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j = X(\bigtimes_{j=1}^{d+1})_j \left(\Xi_j^{1/2} \mathbb{W}_j^{*} \Xi_j^{1/2}\right) = X(\bigtimes_{j=1}^{d+1})_j \left(\Xi_j^{1/2} \mathbb{W}_j^{*} \Xi_j^{1/2} \Xi_j^{1/2\dagger} \Xi_j^{1/2}\right),$$

and then

$$\mathbb{X}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j = \mathbb{X}(\bigtimes_{j=1}^{d+1})_j \left(\Xi_j^{1/2} \mathbb{W}_j^{*} \Xi_j^{1/2}\right) = \mathbb{X}(\bigtimes_{j=1}^{d+1})_j \left(\mathbb{W}_j \Xi_j^{1/2\dagger} \Xi_j^{1/2}\right). \tag{A.2}$$

As such, we have $\mathrm{Vec}\left(\mathbb{X}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j\right) = \mathrm{Vec}\left(\mathbb{X}(\bigtimes_{j=1}^{d+1})_j \left(\mathbb{W}_j \Xi_j^{1/2\dagger} \Xi_j^{1/2}\right)\right)$, and

$$(\mathrm{Vec}(\mathbb{X}))' \bigotimes_{j=d+1}^{1} \mathbb{W}_j (\mathrm{Vec}(\mathbb{X}))' = (\mathrm{Vec}(\mathbb{X}))' \bigotimes_{j=d+1}^{1} \left(\Xi_j^{1/2} \Xi_j^{1/2\dagger} \mathbb{W}_j \Xi_j^{1/2} \Xi_j^{1/2\dagger}\right) (\mathrm{Vec}(\mathbb{X}))'.$$

Letting $\varsigma_1 = \mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}^{*\prime}_{(d)}\mathbb{X}^{*}_{(d)}\right)\right)\mathrm{Vec}\left(\mathbb{X}(\bigtimes_{j=1}^{d+1})_j\left(\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\right)\right)\right]$, we have $\varsigma_1 = \mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}^{*\prime}_{(d)}\mathbb{X}^{*}_{(d)}\right)\right)\bigotimes_{j=d+1}^{1}\left(\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\mathscr{O}'V\right]$. This gives

$$\varsigma_1 = \mathrm{E}\left[h\left(V_1'V_1\right)\left(\bigotimes_{j=d+1}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\mathscr{O}'[I_l,0]'V_1\right]. \tag{A.3}$$

Therefore, using Theorem 1 in Judge and Bock (1978), we get

$$\varsigma_1 = \mathrm{E}\left[h\left(\chi^2_{l+2}\left(\mathbb{M}_{V_1}'\mathbb{M}_{V_1}\right)\right)\bigotimes_{j=d+1}^{1}\left(\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\mathscr{O}'[I_l,0]'\mathbb{M}_{V_1}\right]. \tag{A.4}$$

We also have

$$\begin{aligned}\mathbb{M}_{V_1}'\mathbb{M}_{V_1} &= (\mathrm{Vec}(\mathbb{M}))'\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\mathscr{O}'\,[I_l,0]'\,[I_l,0]\,\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\,(\mathrm{Vec}(\mathbb{M}))\\ &= (\mathrm{Vec}(\mathbb{M}))'\left(\bigotimes_{j=d+1}^{1}\Xi_j\Lambda_j\Xi_j\right)(\mathrm{Vec}(\mathbb{M})) = (\mathrm{Vec}(\mathbb{M}))'\left(\bigotimes_{j=d+1}^{1}\Xi_j\right)(\mathrm{Vec}(\mathbb{M}))\,.\end{aligned}$$

This gives

$$\mathbb{M}_{V_1}'\mathbb{M}_{V_1} = \mathrm{trace}\left(\mathbb{M}^{*\prime}_{(d)}\mathbb{M}^{*}_{(d)}\right). \tag{A.5}$$

Moreover, since $\mathbb{W}_j = \Xi_j^{1/2}\mathbb{W}_j^{*}\Xi_j^{1/2}$, for $j = 1,\cdots,d+1$, we have

$$\bigotimes_{j=d+1}^{1}\left(\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\mathscr{O}'[I_l,0]'\mathbb{M}_{V_1} = \bigotimes_{j=d+1}^{1}\left(\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\Lambda_j\Xi_j^{1/2}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\mathrm{Vec}(\mathbb{M}).$$

This gives

$$\bigotimes_{j=d+1}^{1}\left(\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\mathscr{O}'[I_l,0]'\mathbb{M}_{V_1} = \mathrm{Vec}\left(\mathbb{M}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\Lambda_j\Xi_j\right). \tag{A.6}$$

By Part 1 of Theorem B.1.1,

$$\mathbb{M}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\Lambda_j\Xi_j = \left(\mathbb{M}\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\Lambda_j\Xi_j\right)(\bigtimes_{\substack{i=1\\ i\neq j}}^{d+1})_i\mathbb{W}_i\Xi_i^{1/2\dagger}\Xi_i^{1/2}\Lambda_i\Xi_i,$$

and by Part 2 of Theorem B.1.1, $\mathbb{M}\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\Lambda_j\Xi_j = \left(\mathbb{M}\times_j\Lambda_j\Xi_j\right)\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}$. Hence, $\mathbb{M}\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\Lambda_j\Xi_j = \mathbb{M}\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}$. Also, since $\mathbb{W}_j = \Xi_j^{1/2}\mathbb{W}_j^{*}\Xi_j^{1/2}$,

$$\mathbb{M}\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\Lambda_j\Xi_j = \mathbb{M}\times_j\Xi_j^{1/2}\mathbb{W}_j^{*}\Xi_j^{1/2}\Xi_j^{1/2\dagger}\Xi_j^{1/2} = \mathbb{M}\times_j\Xi_j^{1/2}\mathbb{W}_j^{*}\Xi_j^{1/2} = \mathbb{M}\times_j\mathbb{W}_j.$$

Therefore, (A.14) becomes

$$\left(\mathbb{W}_{d+1}\Xi_{d+1}^{1/2\dagger}\otimes\cdots\otimes\mathbb{W}_1\Xi_1^{1/2\dagger}\right)\mathcal{O}'[I_{l_1\cdots l_{d+1}},0]'\mathbb{M}_{V_1} = \text{Vec}\left(\mathbb{M}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j\right). \tag{A.7}$$

Therefore, by combining relations (A.11), (A.12) and (A.15), we get the stated result. □

Proof of Theorem 2.1.2. From the proof of Theorem 2.1.1, $\text{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right) = V_1'V_1$. Further, $\text{trace}\left(\mathbb{X}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**}\right)\left(\bigotimes_{j=d+1}^{1}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\left(\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\right)(\text{Vec}(\mathbb{X}))$

$$= (\text{Vec}(\mathbb{X}))'\left(\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\right)\mathcal{O}'\mathcal{O}\left(\bigotimes_{j=d+1}^{1}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\mathcal{O}'\mathcal{O}\left(\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\right)(\text{Vec}(\mathbb{X}))\,.$$

Then, $\text{trace}(\mathbb{X}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**}) = V'\mathcal{O}\left(\bigotimes_{j=d+1}^{1}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\mathcal{O}'V$. Since $V_2 = 0$ with probability 1, we get $\text{trace}(\mathbb{X}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**}) = V_1'\bar{\bar{W}}V_1$, with $\bar{\bar{W}} = [I_l,0]\mathcal{O}\left(\bigotimes_{j=d+1}^{1}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\mathcal{O}'[I_l,0]'$. Therefore,

$$\text{E}\left[h\left(\text{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right)\right)\text{trace}\left(\mathbb{X}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**}\right)\right] = \text{E}\left[h(V_1'V_1)V_1'\bar{\bar{W}}V_1\right],$$

with $V_1 \sim \mathcal{N}_l(\mathbb{M}_{V_1}, I_l)$. Then, by Theorem 2 in Judge and Bock (1978), we have

$$\text{E}\left[h(V_1'V_1)V_1'\bar{\bar{W}}V_1\right] = \text{E}\left[h\left(\chi^2_{l+2}(\mathbb{M}_{V_1}'\mathbb{M}_{V_1})\right)\right]\text{trace}(\bar{\bar{W}}) + \text{E}\left[h\left(\chi^2_{l+4}(\mathbb{M}_{V_1}'\mathbb{M}_{V_1})\right)\right]\mathbb{M}_{V_1}'\bar{\bar{W}}\mathbb{M}_{V_1}.$$

Then, from (A.13), it suffices to verify that $\text{trace}\left(\bar{\bar{W}}\right) = D_1$ and $\mathbb{M}_{V_1}'\bar{\bar{W}}\mathbb{M}_{V_1} = D_2$. We have $\text{trace}\left(\bar{\bar{W}}\right) = \text{trace}\left(\left(\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\Lambda_j\Xi_j^{1/2}\right)\left(\bigotimes_{j=d+1}^{1}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\right)$, then

$$\text{trace}\left(\bar{\bar{W}}\right) = \prod_{j=1}^{d+1}\text{trace}\left(\Xi_j^{1/2}\Lambda_j\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right). \tag{A.8}$$

Note that, for $j = 1,\cdots,d+1$,

$$\text{trace}\left(\Xi_j^{1/2}\Lambda_j\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right) = \text{trace}\left(\Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2}\Lambda_j\right) = \text{trace}\left(\mathbb{W}_j\Lambda_j\right),$$

and then $\text{trace}\left(\Xi_j^{1/2}\Lambda_j\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right) = \text{trace}\left(\Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2}\Lambda_j\right) = \text{trace}\left(\mathbb{W}_j\Lambda_j\right)$. Hence, together with (A.16), we have $\text{trace}\left(\bar{\bar{W}}\right) = \prod_{j=1}^{d+1}\text{trace}\left(\mathbb{W}_j\Lambda_j\right) = D_1$. Similarly,

$$\mathbb{M}_{V_1}'\bar{\bar{W}}\mathbb{M}_{V_1} = (\text{Vec}(\mathbb{M}))'\left(\bigotimes_{j=d+1}^{1}\Xi_j\Lambda_j\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\Lambda_j\Xi_j\right)(\text{Vec}(\mathbb{M}))\,.$$

Note that $\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2} = \Xi_j^{1/2}\Xi_j^{1/2\dagger}\Xi_j^{1/2}\mathbb{W}_j^{*}\Xi_j^{1/2}\Xi_j^{1/2\dagger}\Xi_j^{1/2} = \mathbb{W}_j$, for $j = 1, \cdots, d+1$. Hence,

$$\mathbb{M}_{V_1}'\bar{\bar{W}}\mathbb{M}_{V_1} = \left(\mathrm{Vec}(\mathbb{M}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\Lambda_j\Xi_j)\right)'\left(\mathrm{Vec}(\mathbb{M}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\Lambda_j\Xi_j)\right).$$

Note that for $j = 1, \cdots, d+1$, $\mathbb{M} \times_j \Lambda_j\Xi_j = \mathbb{M}$, then by Part 2 in Theorem B.1.1, $\mathbb{M} \times_j \mathbb{W}_j^{1/2}\Lambda_j\Xi_j = (\mathbb{M} \times_j \Lambda_j\Xi_j) \times_j \mathbb{W}_j^{1/2} = \mathbb{M} \times_j \mathbb{W}_j^{1/2}$. As such,

$$\mathbb{M}_{V_1}'\bar{\bar{W}}\mathbb{M}_{V_1} = \left(\mathrm{Vec}\left(\mathbb{M}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)\right)'\left(\mathrm{Vec}\left(\mathbb{M}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)\right) = \mathrm{trace}\left(\mathbb{M}_{11_{(d)}}^{*\prime}\mathbb{M}_{11_{(d)}}^{*}\right) = D_2,$$

this completes the proof. □

Proof of Theorem 2.1.3. As in Theorem 2.1.1, let $V = \mathscr{O} \bigotimes_{j=d+1}^{1} \Xi_j^{1/2}\mathrm{Vec}(X)$. Then,

$$\mathrm{E}\left[h\left(X_{(d)}^{*\prime}X_{(d)}^{*}\right)\mathrm{trace}\left(Y_{(d)}^{**\prime}X_{(d)}^{**}\right)\right] = \mathrm{E}\left[h(V_1'V_1)\mathrm{E}\left[\mathrm{Vec}(Y)|V_1\right]' \bigotimes_{j=d+1}^{1} \mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l, 0]'V1\right].$$

By using the fact that

$$(V_1', (\mathrm{Vec}(Y))')' = \begin{pmatrix} [I_l, 0]\,\mathscr{O} \bigotimes_{j=d+1}^{1} \Xi_j^{1/2} & 0 \\ 0 & I_q \end{pmatrix}((\mathrm{Vec}(X))', (\mathrm{Vec}(Y))')',$$

it can be shown that $\left(V_1', (\mathrm{Vec}(Y))'\right)' \sim \mathscr{N}_{lq}\left(\begin{pmatrix} \mathbb{M}_{V_1} \\ -\mathrm{Vec}\left(\mathbb{M}_X\right) \end{pmatrix}, \begin{pmatrix} I_l & \Pi_{21}^{*}{}' \\ \Pi_{21}^{*} & \Pi_{22} \end{pmatrix}\right)$, with

$$\Pi_{21}^{*} = \left(\bigotimes_{j=d+1}^{1} B_j - \bigotimes_{j=d+1}^{1} \Lambda_j\right)\bigotimes_{j=d+1}^{1} \Xi_j^{1/2}\mathscr{O}'\,[I_l, 0]'.$$

Hence,

$$\mathrm{E}\left[\mathrm{Vec}(Y)|V_1\right] = -\mathrm{Vec}(\mathbb{M}_X) + \left(\bigotimes_{j=d+1}^{1} B_j - \bigotimes_{j=d+1}^{1} \Lambda_j\right)\bigotimes_{j=d+1}^{1} \Xi_j^{1/2}\mathscr{O}'\,[I_l, 0]'\,(V_1 - \mathbb{M}_{V_1}).$$

Then, by Theorem 1 in Judge and Bock (1978), we have

$$\begin{aligned} &\mathrm{E}\left[h(V_1'V_1)\,(\mathrm{Vec}(\mathbb{M}_X))' \bigotimes_{j=d+1}^{1} \mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l, 0]'V1\right] \\ &= \mathrm{E}\left[h\left(\chi_{l+2}^2\left(\mathrm{trace}\left(\mathbb{M}_{X_{(d)}}^{**\prime}\mathbb{M}_{X_{(d)}}^{**}\right)\right)\right)\mathrm{trace}\left(\mathbb{M}_{X_{(d)}}^{*\prime}\mathbb{M}_{X_{(d)}}^{*}\right)\right]. \end{aligned}$$

Similarly,

$$\mathrm{E}\left[h(V_1'V_1)\mathbb{M}_{V_1}'\,[I_l,0]\,\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}B_j\bigotimes_{j=d+1}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l,0]'V1\right]=$$

$$\mathrm{E}\left[h\left(\chi_{l+2}^2\left(\mathrm{trace}\left(\mathbb{M}_{X_{(d)}}^{**\prime}\mathbb{M}_{X_{(d)}}^{**}\right)\right)\right)\mathbb{M}_{V_1}'\,[I_l,0]\,\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}B_j\bigotimes_{j=d+1}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l,0]'\mathbb{M}_{V1}\right].$$

Note that since $\mathbb{M}_{V_1}=[I_l,0]\,\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\mathrm{Vec}(\mathbb{M}_X)$, then

$$\mathbb{M}_{V_1}'\,[I_l,0]\,\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}B_j\bigotimes_{j=d+1}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l,0]'\mathbb{M}_{V1}=(\mathrm{Vec}(\mathbb{M}_X))'\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}$$

$$\times\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\Lambda_j\Xi_j^{1/2}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}B_j\mathbb{W}_j\Xi_j^{1/2\dagger}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\Lambda_j\Xi_j^{1/2}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\,(\mathrm{Vec}(\mathbb{M}_X))\,.$$

This gives

$$\begin{aligned}&\mathbb{M}_{V_1}'\,[I_l,0]\,\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}B_j\bigotimes\nolimits_{j=d+1}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l,0]'\mathbb{M}_{V1}\\&=(\mathrm{Vec}(\mathbb{M}_X))'\bigotimes_{j=d+1}^{1}\Xi_jB_j\mathbb{W}_j\Lambda_j\Xi_j\,(\mathrm{Vec}(\mathbb{M}_X))\end{aligned}$$

and then,

$$\begin{aligned}&\mathbb{M}_{V_1}'\,[I_l,0]\,\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}B_j\bigotimes_{j=d+1}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l,0]'\mathbb{M}_{V1}\\&=\;(\mathrm{Vec}(\mathbb{M}_X))'\bigotimes_{j=d+1}^{1}\Xi_jB_j\mathbb{W}_j\,(\mathrm{Vec}(\mathbb{M}_X)\,.\end{aligned}$$

Hence,

$$\mathrm{E}\left[h\left(V_1'V_1\right)\mathbb{M}_{V_1}'[I_l,0]\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\Lambda_j\mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l,0]'V_1\right]=$$

$$\mathrm{E}\left[h\left(\chi_{l+2}^2\left(\mathrm{trace}\left(\mathbb{M}_{X_{(d)}}^{**\prime}\mathbb{M}_{X_{(d)}}^{**}\right)\right)\right)\mathbb{M}_{V_1}'[I_l,0]\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\Lambda_j\mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l,0]'\mathbb{M}_{V_1}\right].$$

Since

$$\mathscr{O}'[I_l,0]'\mathbb{M}_{V_1}=\mathscr{O}'[I_l,0]'[I_l,0]\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\mathrm{Vec}(\mathbb{M}_X)=\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\Lambda_j\Xi_j\mathrm{Vec}(\mathbb{M}_X),$$

then using similar techniques as in Theorem 2.1.1

$$\mathrm{E}\left[h(V_1'V_1)V_1'\,[I_l,0]\,\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}B_j\bigotimes_{j=d+1}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l,0]'V1\right]$$
$$=\mathrm{E}\left[h\left(\chi^2_{l+2}\left(\mathrm{trace}\left(\mathbb{M}^{**\prime}_{X_{(d)}}\mathbb{M}^{**}_{X_{(d)}}\right)\right)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_jB_j)$$
$$+\,\mathrm{E}\left[h\left(\chi^2_{l+4}\left(\mathrm{trace}\left(\mathbb{M}^{**\prime}_{X_{(d)}}\mathbb{M}^{**}_{X_{(d)}}\right)\right)\right)\right]\mathrm{trace}\left(\mathbb{M}'_{X_{(d)}}\left(\mathbb{M}_X(\bigtimes_{j=1}^{d+1})_j\Xi_jB_j\mathbb{W}_j\right)_{(d)}\right),$$

and

$$\mathrm{E}\left[h(V_1'V_1)V_1'\,[I_l,0]\,\mathscr{O}\bigotimes_{j=d+1}^{1}\Xi_j^{1/2}\Lambda_j\bigotimes_{j=d+1}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\mathscr{O}'[I_l,0]'V1\right]$$
$$=\mathrm{E}\left[h\left(\chi^2_{l+2}\left(\mathrm{trace}\left(\mathbb{M}^{**\prime}_{X_{(d)}}\mathbb{M}^{**}_{X_{(d)}}\right)\right)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\Lambda_j\mathbb{W}_j\Lambda_j\Xi_j)$$
$$+\,\mathrm{E}\left[h\left(\chi^2_{l+4}\left(\mathrm{trace}\left(\mathbb{M}^{**\prime}_{X_{(d)}}\mathbb{M}^{**}_{X_{(d)}}\right)\right)\right)\right]\mathrm{trace}\left(\left(\mathbb{M}_X(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\Lambda_j\Xi_j\right)'_{(d)}\left(\mathbb{M}_X(\bigtimes_{j=1}^{d+1})_j\Lambda_j\Xi_j\right)_{(d)}\right)$$
$$=\mathrm{E}\left[h\left(\chi^2_{l+2}\left(\mathrm{trace}\left(\mathbb{M}^{**\prime}_{X_{(d)}}\mathbb{M}^{**}_{X_{(d)}}\right)\right)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j\Lambda_j)$$
$$+\,\mathrm{E}\left[h\left(\chi^2_{l+4}\left(\mathrm{trace}\left(\mathbb{M}^{**\prime}_{X_{(d)}}\mathbb{M}^{**}_{X_{(d)}}\right)\right)\right)\right]\mathrm{trace}\left(\left(\mathbb{M}_X(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mathbb{M}_X(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\right)_{(d)}\right).$$

Therefore, the proof follows from algebraic computations. □

Proof of Theorem 2.2.1. For $j=1,\cdots,d$, let $\Xi_j^{1/2\dagger}$ be the Moore-Penrose pseudo-inverse of $\Xi_j^{1/2}$. Then, by the definition of the Moore-Penrose pseudo-inverse, we have $\Xi_j^{1/2\dagger}\Xi_j^{1/2}\Xi_j^{1/2\dagger}=\Xi_j^{1/2\dagger}$; $\Xi_j^{1/2}\Xi_j^{1/2\dagger}\Xi_j^{1/2}=\Xi_j^{1/2}$; $\left(\Xi_j^{1/2\dagger}\Xi_j^{1/2}\right)'=\Xi_j^{1/2\dagger}\Xi_j^{1/2}$; and $\left(\Xi_j^{1/2}\Xi_j^{1/2\dagger}\right)'=\Xi_j^{1/2}\Xi_j^{1/2\dagger}$. Further, since $\Xi_j^{1/2}\boldsymbol{\Lambda}_j\Xi_j^{1/2}$ are idempotent, then the Kronecker product $\bigotimes_{j=d}^{1}\Xi_j^{1/2}\boldsymbol{\Lambda}_j\Xi_j^{1/2}$ is also idempotent. Hence, there exists an orthogonal matrix Q such that $\boldsymbol{Q}\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2}\boldsymbol{\Lambda}_j\Xi_j^{1/2}\right)\boldsymbol{Q}'=\begin{pmatrix}I_p & 0\\ 0 & 0\end{pmatrix}$, with $p=\prod_{j=1}^{d}p_j$. Let $\mathbb{V}=\boldsymbol{Q}\bigotimes_{j=d}^{1}\Xi_j^{1/2}\mathrm{Vec}(\mathbb{X})$. Hence,

$$\mathbb{V}=\begin{pmatrix}V_1\\ V_2\end{pmatrix}\sim\mathscr{E}_q^p\left(\begin{pmatrix}\mathbb{M}_1\\ 0\end{pmatrix},\boldsymbol{\Sigma}_v;g\right),\tag{A.9}$$

with

$$\mathbb{M}_1 = \begin{bmatrix} I_p & 0 \end{bmatrix} \boldsymbol{Q} \bigotimes_{j=d}^{1} \Xi_j^{1/2} \mathrm{Vec}(\mathbb{M}), \qquad \boldsymbol{\Sigma}_v = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix}. \tag{A.10}$$

From (A.9), $V_2 = 0$ with probability 1. Thus,
$\mathrm{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right) = (\mathrm{Vec}(\mathbb{X}))' \bigotimes_{j=d}^{1} \Xi_j^{1/2} \boldsymbol{Q}'\boldsymbol{Q} \bigotimes_{j=d}^{1} (\Xi_j^{1/2}\boldsymbol{\Lambda}_j\Xi_j^{1/2})\boldsymbol{Q}'\boldsymbol{Q} \bigotimes_{j=d}^{1} \Xi_j^{1/2} \,(\mathrm{Vec}(\mathbb{X}))$. This gives

$$\mathrm{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right) = \mathbb{V}'[I_p, 0]'[I_p, 0]\mathbb{V} = V_1'V_1.$$

Further, it can be verified that

$$\mathbb{X}\Big(\bigtimes_{j=1}^{d}\Big)_j \mathbb{W}_j = \mathbb{X}\Big(\bigtimes_{j=1}^{d}\Big)_j \left(\Xi_j^{1/2}\mathbb{W}_j^{*}\Xi_j^{1/2}\right) = \mathbb{X}\Big(\bigtimes_{j=1}^{d}\Big)_j \left(\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\right).$$

As such, we have $\mathrm{Vec}\left(\mathbb{X}(\bigtimes_{j=1}^{d})_j \mathbb{W}_j\right) = \mathrm{Vec}\left(\mathbb{X}(\bigtimes_{j=1}^{d})_j \left(\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\right)\right)$, and

$$(\mathrm{Vec}\,(\mathbb{X}))' \bigotimes_{j=d}^{1} \mathbb{W}_j \,(\mathrm{Vec}\,(\mathbb{X})) = (\mathrm{Vec}\,(\mathbb{X}))' \bigotimes_{j=d}^{1} \left(\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2}\Xi_j^{1/2\dagger}\right)(\mathrm{Vec}\,(\mathbb{X}))\,.$$

Set $\varsigma_1 = \mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right)\right)\mathrm{Vec}\left(\mathbb{X}(\bigtimes_{j=1}^{d})_j \left(\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\right)\right)\right]$.
We have $\varsigma_1 = \mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right)\right)\bigotimes_{j=d}^{1}\left(\mathbb{W}_j\Xi_j^{1/2\dagger}\right)Q'\mathbb{V}\right]$. This gives

$$\varsigma_1 = \mathrm{E}\left[h\,(V_1'V_1)\left(\bigotimes_{j=d}^{1} \mathbb{W}_j\Xi_j^{1/2\dagger}\right)Q'[I_p, 0]'V_1\right]. \tag{A.11}$$

Therefore, by using Lemma A.2 of Nkurunziza (2013), we get

$$\varsigma_1 = \psi_{1,p+2}^{(1)}(\mathbb{M}_1'\mathbb{M}_1)\boldsymbol{Q}'[I_p, 0]'\mathbb{M}_1. \tag{A.12}$$

We also have $\mathbb{M}_1'\mathbb{M}_1 = (\mathrm{Vec}(\mathbb{M}))' \bigotimes_{j=d}^{1} \Xi_j^{1/2}\boldsymbol{Q}'\left[I_p, 0\right]'\left[I_p, 0\right]\boldsymbol{Q}\bigotimes_{j=d}^{1} \Xi_j^{1/2}\,(\mathrm{Vec}(\mathbb{M}))$. Then,

$$\mathbb{M}_1'\mathbb{M}_1 = (\mathrm{Vec}(\mathbb{M}))'\left(\bigotimes_{j=d}^{1} \Xi_j\boldsymbol{\Lambda}_j\Xi_j\right)(\mathrm{Vec}(\mathbb{M})) = (\mathrm{Vec}(\mathbb{M}))'\left(\bigotimes_{j=d}^{1} \Xi_j\right)(\mathrm{Vec}(\mathbb{M}))\,.$$

This gives

$$\mathbb{M}_1'\mathbb{M}_1 = \|\mathbb{M}\|^2_{\{\Xi_i, i=1,\cdots,d\}}. \tag{A.13}$$

Moreover, since $\mathbb{W}_j = \Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2}$, for $j = 1, \cdots, d$, we have

$$\bigotimes_{j=d}^{1}\left(\mathbb{W}_j\Xi_j^{1/2\dagger}\right)Q'[I_p,0]'\mathbb{M}_1 = \bigotimes_{j=d}^{1}\left(\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\bigotimes_{j=d}^{1}\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j^{1/2}\bigotimes_{j=d}^{1}\Xi_j^{1/2}\mathrm{Vec}(\mathbb{M}).$$

This gives

$$\bigotimes_{j=d}^{1}\left(\mathbb{W}_j\Xi_j^{1/2\dagger}\right)Q'[I_p,0]'\mathbb{M}_1 = \mathrm{Vec}\left(\mathbb{M}\Big(\bigtimes_{j=1}^{d}\Big)_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j\right). \tag{A.14}$$

Then,

$$\mathbb{M}(\bigtimes_{j=1}^{d})_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j = \left(\mathbb{M}\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j\right)\Big(\bigtimes_{\substack{i=1\\ i\neq j}}^{d}\Big)_i\mathbb{W}_i\Xi_i^{1/2\dagger}\Xi_i^{1/2}\mathbf{\Lambda}_i\Xi_i,$$

and $\mathbb{M}\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j = \left(\mathbb{M}\times_j\mathbf{\Lambda}_j\Xi_j\right)\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}$.
Hence, $\mathbb{M}\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j = \mathbb{M}\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}$. Also, since $\mathbb{W}_j = \Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2}$,

$$\mathbb{M}\times_j\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j = \mathbb{M}\times_j\Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2}\Xi_j^{1/2\dagger}\Xi_j^{1/2} = \mathbb{M}\times_j\Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2} = \mathbb{M}\times_j\mathbb{W}_j.$$

Therefore, (A.14) becomes

$$\left(\mathbb{W}_d\Xi_d^{1/2\dagger}\otimes\cdots\otimes\mathbb{W}_1\Xi_1^{1/2\dagger}\right)Q'[I_p,0]'\mathbb{M}_1 = \mathrm{Vec}\left(\mathbb{M}\Big(\bigtimes_{j=1}^{d}\Big)_j\mathbb{W}_j\right). \tag{A.15}$$

Therefore, by combining relations (A.11), (A.12) and (A.15), we get the stated result. □

Proof of Theorem 2.2.2. As in Theorem 2.2.1, let $\mathbb{V} = \boldsymbol{Q}\bigotimes_{j=d}^{1}\Xi_j^{1/2}\mathrm{Vec}(\mathbb{X})$. Then,

$$\mathrm{E}\left[h\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right)\mathrm{trace}\left(\mathbb{Y}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**}\right)\right] = \mathrm{E}\left[h(V_1'V_1)\mathrm{E}\left[\mathrm{Vec}(\mathbb{Y})|V_1\right]'\bigotimes_{j=d}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\boldsymbol{Q}'[I_p,0]'V_1\right].$$

By using the fact that

$$(V_1',(\mathrm{Vec}(\mathbb{Y}))')' = \begin{pmatrix}\left[I_p,0\right]\boldsymbol{Q}\bigotimes_{j=d}^{1}\Xi_j^{1/2} & 0\\ 0 & I_q\end{pmatrix}((\mathrm{Vec}(\mathbb{X}))',(\mathrm{Vec}(\mathbb{Y}))')',$$

it can be shown that $\left(V_1', (\mathrm{Vec}(\mathbb{Y}))'\right)' \sim \mathscr{E}_{pq}\left(\begin{pmatrix} \mathbb{M}_1 \\ -\mathrm{Vec}\,(\mathbb{M}_1) \end{pmatrix}, \begin{pmatrix} I_p & \mathbf{\Pi}_{21}^{*'} \\ \mathbf{\Pi}_{21}^{*} & \mathbf{\Pi}_{22} \end{pmatrix}\right)$,

with $\mathbf{\Pi}_{21}^* = \left(\bigotimes_{j=d}^{1} \boldsymbol{B}_j - \bigotimes_{j=d}^{1} \mathbf{\Lambda}_j\right) \bigotimes_{j=d}^{1} \Xi_j^{1/2} \boldsymbol{Q}' \left[I_p, 0\right]'$. Hence,

$$\mathrm{E}\left[\mathrm{Vec}(\mathbb{Y})|V_1\right] = \mathrm{Vec}\,(\mathbb{M}_2) + \left(\bigotimes_{j=d}^{1} \boldsymbol{B}_j - \bigotimes_{j=d}^{1} \mathbf{\Lambda}_j\right) \bigotimes_{j=d}^{1} \Xi_j^{1/2} \boldsymbol{Q}' \left[I_p, 0\right]' (V_1 - \mathbb{M}_1)\,.$$

Then, by Lemma A.2 of Nkurunziza (2013),

$$\mathrm{E}\left[h(V_1'V_1)\,(\mathrm{Vec}(\mathbb{M}_2))' \bigotimes_{j=d}^{1} \mathbb{W}_j \Xi_j^{1/2\dagger} \boldsymbol{Q}'[I_p, 0]' V_1\right] = \psi_{1,p+2}^{(1)}\left(||\mathbb{M}||^2_{\{\Xi_i, i=1,\cdots,d\}}\right)$$
$$\times \mathrm{trace}\left(\left(\mathbb{M}_2 \Big(\bigtimes_{j=1}^{d} \Big)_j \mathbb{W}_j^{1/2}\right)'_{(d)} \left(\mathbb{M} \Big(\bigtimes_{j=1}^{d} \Big)_j \mathbb{W}_j^{1/2}\right)_{(d)}\right)$$

Similarly,

$$\mathrm{E}\left[h(V_1'V_1)\mathbb{M}_1' \left[I_p, 0\right] \boldsymbol{Q} \bigotimes_{j=d}^{1} \Xi_j^{1/2} \boldsymbol{B}_j \bigotimes_{j=d}^{1} \mathbb{W}_j \Xi_j^{1/2\dagger} \boldsymbol{Q}'[I_p, 0]' V1\right] =$$
$$\psi_{1,p+2}^{(1)}\left(||\mathbb{M}||^2_{\{\Xi_i, i=1,\cdots,d\}}\right)\left(\mathbb{M}_1' \left[I_p, 0\right] Q \bigotimes_{j=d}^{1} \Xi_j^{1/2} \boldsymbol{B}_j \bigotimes_{j=d} \mathbb{W}_j \Xi_j^{1/2\dagger} \boldsymbol{Q}'[I_p, 0]' \mathbb{M}_1\right).$$

Note that since $\mathbb{M}_1 = \left[I_p, 0\right] \boldsymbol{Q} \bigotimes_{j=d}^{1} \Xi_j^{1/2} \mathrm{Vec}(\mathbb{M})$, then

$$\mathbb{M}_1' \left[I_p, 0\right] \boldsymbol{Q} \bigotimes_{j=d}^{1} \Xi_j^{1/2} \boldsymbol{B}_j \bigotimes_{j=d}^{1} \mathbb{W}_j \Xi_j^{1/2\dagger} Q'[I_p, 0]' \mathbb{M}_1 = (\mathrm{Vec}(\mathbb{M}))' \bigotimes_{j=d}^{1} \Xi_j^{1/2} \bigotimes_{j=d}^{1} \Xi_j^{1/2} \mathbf{\Lambda}_j \Xi_j^{1/2}$$
$$\times \bigotimes_{j=d}^{1} \Xi_j^{1/2} \boldsymbol{B}_j \mathbb{W}_j \Xi_j^{1/2\dagger} \bigotimes_{j=d}^{1} \Xi_j^{1/2} \mathbf{\Lambda}_j \Xi_j^{1/2} \bigotimes_{j=d}^{1} \Xi_j^{1/2} (\mathrm{Vec}(\mathbb{M}))\,.$$

This gives

$$\mathbb{M}_1' \left[I_p, 0\right] \boldsymbol{Q} \bigotimes_{j=d}^{1} \Xi_j^{1/2} \boldsymbol{B}_j \bigotimes_{j=d}^{1} \mathbb{W}_j \Xi_j^{1/2\dagger} Q'[I_p, 0]' \mathbb{M}_1 = (\mathrm{Vec}(\mathbb{M}))' \bigotimes_{j=d}^{1} \Xi_j \boldsymbol{B}_j \mathbb{W}_j \mathbf{\Lambda}_j \Xi_j\, (\mathrm{Vec}(\mathbb{M}))$$

and then,

$$\mathbb{M}_1' \left[I_p, 0\right] \boldsymbol{Q} \bigotimes_{j=d}^{1} \Xi_j^{1/2} \boldsymbol{B}_j \bigotimes_{j=d}^{1} \mathbb{W}_j \Xi_j^{1/2\dagger} \boldsymbol{Q}'[I_p, 0]' \mathbb{M}_1 = (\mathrm{Vec}(\mathbb{M}))' \bigotimes_{j=d}^{1} \Xi_j \boldsymbol{B}_j \mathbb{W}_j \mathrm{Vec}\,(\mathbb{M})\,.$$

Hence,

$$\mathrm{E}\left[h\left(V_1'V_1\right)\mathbb{M}_1'[I_p,0]\boldsymbol{Q}\bigotimes_{j=d}^{1}\Xi_j^{1/2}\boldsymbol{B}_j\mathbb{W}_j\Xi_j^{1/2\dagger}\boldsymbol{Q}'[I_p,0]'V_1\right]=$$

$$\psi_{1,p+2}^{(1)}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\mathbb{M}_{(d)}'\left(\mathbb{M}\left(\bigtimes_{j=1}^{d}\right)_j\Xi_j\boldsymbol{B}_j\mathbb{W}_j\right)_{(d)}\right).$$

Similarly,

$$\mathrm{E}\left[h\left(V_1'V_1\right)\mathbb{M}_1'[I_p,0]\boldsymbol{Q}\bigotimes_{j=d}^{1}\Xi_j^{1/2}\boldsymbol{\Lambda}_j\mathbb{W}_j\Xi_j^{1/2\dagger}\boldsymbol{Q}'[I_p,0]'V_1\right]$$

$$=\psi_{1,p+2}^{(1)}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\mathbb{M}_{(d)}'\left(\mathbb{M}(\bigtimes_{j=1}^{d})_j\Xi_j\boldsymbol{\Lambda}_j\mathbb{W}_j\right)_{(d)}\right)$$

$$=\psi_{1,p+2}^{(1)}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\left(\mathbb{M}(\bigtimes_{j=1}^{d})_j\mathbb{W}_j^{1/2}\right)_{(d)}'\left(\mathbb{M}(\bigtimes_{j=1}^{d})_j\mathbb{W}_j^{1/2}\right)_{(d)}\right).$$

Further, $\boldsymbol{Q}'[I_p,0]'\mathbb{M}_1=\boldsymbol{Q}'[I_p,0]'[I_p,0]Q\bigotimes_{j=d}^{1}\Xi_j^{1/2}\mathrm{Vec}(\mathbb{M}_1)=\bigotimes_{j=d}^{1}\Xi_j^{1/2}\boldsymbol{\Lambda}_j\Xi_j\mathrm{Vec}(\mathbb{M})$. Then, as in proof of Theorem 2.2.1,

$$\mathrm{E}\left[h(V_1'V_1)V_1'\left[I_p,0\right]\boldsymbol{Q}\bigotimes_{j=d}^{1}\Xi_j^{1/2}\boldsymbol{B}_j\bigotimes_{j=d}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\boldsymbol{Q}'[I_p,0]'V1\right]$$

$$=\psi_{1,p+2}^{(2)}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\prod_{j=1}^{d}\mathrm{trace}(\mathbb{W}_j\boldsymbol{B}_j)+\psi_{1,p+4}^{(1)}\mathrm{trace}\left(\mathbb{M}_{(d)}'\left(\mathbb{M}\left(\bigtimes_{j=1}^{d}\right)_j\Xi_j\boldsymbol{B}_j\mathbb{W}_j\right)_{(d)}\right).$$

We also have

$$\mathrm{E}\left[h(V_1'V_1)V_1'\left[I_p,0\right]\boldsymbol{Q}\bigotimes_{j=d}^{1}\Xi_j^{1/2}\boldsymbol{\Lambda}_j\bigotimes_{j=d}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\boldsymbol{Q}'[I_p,0]'V1\right]$$

$$=\psi_{1,p+2}^{(2)}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\prod_{j=1}^{d}\mathrm{trace}(\boldsymbol{\Lambda}_j\mathbb{W}_j\boldsymbol{\Lambda}_j\Xi_j)$$

$$+\psi_{1,p+4}^{(1)}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\left(\mathbb{M}\left(\bigtimes_{j=1}^{d}\right)_j\mathbb{W}_j^{1/2}\boldsymbol{\Lambda}_j\Xi_j\right)_{(d)}'\left(\mathbb{M}\left(\bigtimes_{j=1}^{d}\right)_j\boldsymbol{\Lambda}_j\Xi_j\right)_{(d)}\right).$$

Then,

$$\begin{aligned}
&\mathrm{E}\left[h(V_1'V_1)V_1'\left[I_p,0\right]\boldsymbol{Q}\bigotimes_{j=d}^{1}\Xi_j^{1/2}\boldsymbol{\Lambda}_j\bigotimes_{j=d}^{1}\mathbb{W}_j\Xi_j^{1/2\dagger}\boldsymbol{Q}'[I_p,0]'V1\right]\\
&=\psi_{1,p+2}^{(2)}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\prod_{j=1}^{d}\mathrm{trace}(\mathbb{W}_j\boldsymbol{\Lambda}_j)\\
&+\psi_{1,p+4}^{(1)}\left(||\mathbb{M}||^2_{\{\Xi_i,i=1,\cdots,d\}}\right)\mathrm{trace}\left(\left(\mathbb{M}\Big(\bigtimes_{j=1}^{d}\Big)_j\mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mathbb{M}\Big(\bigtimes_{j=1}^{d}\Big)_j\mathbb{W}_j^{1/2}\right)_{(d)}\right).
\end{aligned}$$

Therefore, the proof follows from algebraic computations. □

Proof of Theorem 2.2.3. From the proof of Theorem 2.2.1, $\mathrm{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right)=V_1'V_1$. Further,

$$\begin{aligned}
\mathrm{trace}\left(\mathbb{X}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**}\right)&=(\mathrm{Vec}(\mathbb{X}))'\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2}\right)\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2}\right)(\mathrm{Vec}(\mathbb{X}))\\
&=(\mathrm{Vec}(\mathbb{X}))'\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2}\right)\boldsymbol{Q}'\boldsymbol{Q}\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\boldsymbol{Q}'\boldsymbol{Q}\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2}\right)(\mathrm{Vec}(\mathbb{X}))\,.
\end{aligned}$$

Then, $\mathrm{trace}(\mathbb{X}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**})=V'\boldsymbol{Q}\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\boldsymbol{Q}'V$. Since $V_2=0$ with probability 1, we get

$\mathrm{trace}(\mathbb{X}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**})=V_1'\bar{\bar{\boldsymbol{W}}}V_1$, with $\bar{\bar{\boldsymbol{W}}}=[I_l,0]\boldsymbol{Q}\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\boldsymbol{Q}'[I_l,0]'$. Therefore,

$$\mathrm{E}\left[h\left(\mathrm{trace}\left(\mathbb{X}_{(d)}^{*\prime}\mathbb{X}_{(d)}^{*}\right)\right)\mathrm{trace}\left(\mathbb{X}_{(d)}^{**\prime}\mathbb{X}_{(d)}^{**}\right)\right]=\mathrm{E}\left[h(V_1'V_1)V_1'\bar{\bar{W}}V_1\right],$$

with $V_1\sim\mathscr{E}_q^p\left(\mathbb{M}_1,I_p\right)$. Then, by Lemma A.3 in Nkurunziza (2013), we have

$$\mathrm{E}\left[h(V_1'V_1)V_1'\bar{\bar{\boldsymbol{W}}}V_1\right]=\psi_{1,p+2}^{(1)}(\mathbb{M}_1'\mathbb{M}_1)\mathrm{trace}(\bar{\bar{\boldsymbol{W}}})+\psi_{1,p+4}^{(1)}(\mathbb{M}_1'\mathbb{M}_1)\mathbb{M}_1'\bar{\bar{\boldsymbol{W}}}\mathbb{M}_1.$$

Then, it suffices to verify that $\mathrm{trace}\left(\bar{\bar{\boldsymbol{W}}}\right)=D_1$ and $\mathbb{M}_1'\bar{\bar{\boldsymbol{W}}}\mathbb{M}_1=D_2$. We have $\mathrm{trace}\left(\bar{\bar{\boldsymbol{W}}}\right)=\mathrm{trace}\left(\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2}\boldsymbol{\Lambda}_j\Xi_j^{1/2}\right)\left(\bigotimes_{j=d}^{1}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right)\right)$, then

$$\mathrm{trace}\left(\bar{\bar{\boldsymbol{W}}}\right)=\prod_{j=1}^{d}\mathrm{trace}\left(\Xi_j^{1/2}\boldsymbol{\Lambda}_j\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right).\tag{A.16}$$

Note that, for $j = 1, \cdots, d$, $\text{trace}\left(\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right) = \text{trace}\left(\Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2}\mathbf{\Lambda}_j\right) = \text{trace}\left(\mathbb{W}_j\mathbf{\Lambda}_j\right)$, and then

$\text{trace}\left(\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\right) = \text{trace}\left(\Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2}\mathbf{\Lambda}_j\right) = \text{trace}\left(\mathbb{W}_j\mathbf{\Lambda}_j\right)$. Hence, together with (A.16), we get

$$\text{trace}\left(\bar{\bar{W}}\right) = \prod_{j=1}^{d} \text{trace}\left(\mathbb{W}_j\mathbf{\Lambda}_j\right) = D_1.$$

Similarly,

$$\mathbb{M}_1'\bar{\bar{\boldsymbol{W}}}\mathbb{M}_1 = (\text{Vec}(\mathbb{M}))'\left(\bigotimes_{j=d}^{1}\Xi_j\mathbf{\Lambda}_j\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2}\mathbf{\Lambda}_j\Xi_j\right)(\text{Vec}(\mathbb{M})).$$

Note that $\Xi_j^{1/2}\Xi_j^{1/2\dagger}\mathbb{W}_j\Xi_j^{1/2\dagger}\Xi_j^{1/2} = \Xi_j^{1/2}\Xi_j^{1/2\dagger}\Xi_j^{1/2}\mathbb{W}_j^*\Xi_j^{1/2}\Xi_j^{1/2\dagger}\Xi_j^{1/2} = \mathbb{W}_j$, for $j = 1, \cdots, d$. Hence,

$$\mathbb{M}_1'\bar{\bar{W}}\mathbb{M}_1 = \left(\text{Vec}(\mathbb{M}\Big(\bigtimes_{j=1}^{d}\Big)_j\mathbb{W}_j^{1/2}\mathbf{\Lambda}_j\Xi_j)\right)'\left(\text{Vec}(\mathbb{M}\Big(\bigtimes_{j=1}^{d}\Big)_j\mathbb{W}_j^{1/2}\mathbf{\Lambda}_j\Xi_j)\right).$$

Note that for $j = 1, \cdots, d$, $\mathbb{M}\times_j\mathbf{\Lambda}_j\Xi_j = \mathbb{M}$, then, $\mathbb{M}\times_j\mathbb{W}_j^{1/2}\mathbf{\Lambda}_j\Xi_j = (\mathbb{M}\times_j\mathbf{\Lambda}_j\Xi_j)\times_j\mathbb{W}_j^{1/2} = \mathbb{M}\times_j\mathbb{W}_j^{1/2}$. As such,

$$\begin{aligned}
\mathbb{M}_1'\bar{\bar{\boldsymbol{W}}}\mathbb{M}_1 &= \left(\text{Vec}\left(\mathbb{M}(\bigtimes_{j=1}^{d})_j\mathbb{W}_j^{1/2}\right)\right)'\left(\text{Vec}\left(\mathbb{M}(\bigtimes_{j=1}^{d})_j\mathbb{W}_j^{1/2}\right)\right) \\
&= \text{trace}\left(\left(\mathbb{M}(\bigtimes_{j=1}^{d})_j\mathbb{W}_j^{1/2}\right)_{(d)}'\left(\mathbb{M}(\bigtimes_{j=1}^{d})_j\mathbb{W}_j^{1/2}\right)_{(d)}\right) \\
&= D_2,
\end{aligned}$$

this completes the proof. □

Appendix B

Tensor Regression with Multiple Change-points

B.1 Properties of tensors and definitions

Theorem B.1.1. *Let* $\mathbb{S} \in \mathbb{R}^{q_1 \times q_2 \times \cdots \times q_d}$ *be a* $d-$*dimensional tensor and let* $A \in \mathbb{R}^{I_m \times q_m}$ *and* $B \in \mathbb{R}^{I_n \times q_n}$ *be two matrices. Then,*

1.
$$\mathbb{S} \times_m A \times_n B = (\mathbb{S} \times_m A) \times_n B = (\mathbb{S} \times_n B) \times_m A, \quad (m \neq n),$$

2. *If* $C \in \mathbb{R}^{K \times I_m}$ *is a given matrix, then*
$$\mathbb{S} \times_m A \times_m C = \mathbb{S} \times_m (CA).$$

The proof of this result is given in Kolda and Bader (2009).

To simplify some notations, let $\Lambda^* = \bigotimes_{i=d+1}^{1} \Lambda_i$, $\Sigma_{11}^* = \Gamma^{-1}\Lambda_{d+1}\Gamma^{-1} \otimes \Lambda^*$, $\Omega = \bigotimes_{i=d+1}^{1} \Omega_i$, $P^* = \mathbb{G}_{d+1}^* \otimes \bigotimes_{i=d}^{1} \mathbb{G}_i$, and let $G_0 = \left(\Gamma^{*-1} \otimes I_{q_d} \cdots \otimes I_{q_1}\right)$.

Proposition B.1.1. *Let $\mathbb{X}$ be a $(d+1)-$dimensional tensor random variate with $\mathbb{X} \sim \mathcal{N}_{q_1\times\cdots\times q_d\times(m_0+1)q_{d+1}}(0,\Lambda^*)$, and let Γ^* be an $(m_0+1)q_{d+1}\times(m_0+1)q_{d+1}$ symmetric, non-random matrix. Then, $\mathbb{X}\times_{d+1}\Gamma^{*-1}\sim\mathcal{N}_{q_1\times\cdots\times q_d\times(m_0+1)q_{d+1}}\left(0,\Sigma_{11}^*\right)$.*

Proof. Note that $\text{Vec}\left(\mathbb{X}\times_{d+1}\Gamma^{*-1}\right)=\left(\Gamma^{*-1}\otimes I_{q_d}\cdots\otimes I_{q_1}\right)\text{Vec}(\mathbb{X})=G_0\text{Vec}(\mathbb{X})$. Now, since $\text{Vec}(\mathbb{X})\sim\mathcal{N}_{(m_0+1)q_1\cdots q_{d+1}}(0,\Lambda^*)$, then $G_0\text{Vec}(\mathbb{X})\sim\mathcal{N}_{(m_0+1)q_1\cdots q_{d+1}}\left(0,\Sigma_{11}^*\right)$, with $\Sigma_{11}^*=G_0\Lambda^*G_0$. Then, the proof follows from Kronecker product rules. □

Proposition B.1.2. *Let $\epsilon_1^*\sim\mathcal{N}_{q_1\times\cdots\times q_d\times q_{d+1}}\left(0,\Sigma_{11}^*\right)$. Then,*

$$\left(I,I-\Omega',\Omega'\right)'\text{Vec}(\epsilon_1^*)+(0,-1,1)'\otimes P^*\text{Vec}(r_0)\sim\mathcal{N}_{q_1\cdots q_d3(m_0+1)q_{d+1}}\left(\text{Vec}(\mu),\Sigma\right),$$

where $\mu=0\boxplus_{(d+1)}(-r_0\times_1\mathbb{G}_1\times_2\cdots\times_d\mathbb{G}_d\times_{d+1}\mathbb{G}_{d+1}^*)\boxplus_{(d+1)}(r_0\times_1\mathbb{G}_1\times_2\cdots\times_d\mathbb{G}_d\times_{d+1}\mathbb{G}_{d+1}^*)$,

$$\Sigma=\begin{pmatrix}\Sigma_{11}^* & \Sigma_{11}^*-\Sigma_{11}^*\Omega' & \Sigma_{11}^*\Omega'\\ \Sigma_{11}^*-J^*\Sigma_{11}^* & \Sigma_{11}^*-J^*\Sigma_{11}^*-\Sigma_{11}^*\Omega'+J^*\Sigma_{11}^*\Omega' & \Sigma_{11}^*\Omega'-J^*\Sigma_{11}^*\Omega'\\ J^*\Sigma_{11}^* & J^*\Sigma_{11}^*-J^*\Sigma_{11}^*\Omega' & J^*\Sigma_{11}^*\Omega'\end{pmatrix}.$$

Proof. We have

$$\begin{pmatrix}I\\ \\ I-J^*\\ \\ J^*\end{pmatrix}\text{Vec}(\epsilon_1^*)+\begin{pmatrix}0\\ \\ -P^*\text{Vec}(r_0)\\ \\ P^*\text{Vec}(r_0)\end{pmatrix}\sim\mathcal{N}_{q_1\cdots q_d3(m_0+1)q_{d+1}}\left(0+\mu^*,\Sigma\right)$$

with

$$\mu^*=\begin{pmatrix}0\\ \\ -P^*\text{Vec}(r_0)\\ \\ P^*\text{Vec}(r_0)\end{pmatrix},\Sigma=\begin{pmatrix}I\\ \\ I-J^*\\ \\ J^*\end{pmatrix}\Lambda^*\left(I\ \vdots\ I-J^{*\prime}\ \vdots\ J^{*\prime}\right).$$

Then, by some algebraic computations, we have

$$\Sigma = \begin{pmatrix} \Lambda^* & \Lambda^*(I - J^{*\prime}) & \Lambda^* J^{*\prime} \\ (I - J^*)\Lambda^* & (I - J^*)\Lambda^*(I - J^{*\prime}) & (I - J^*)\Lambda^* J^{*\prime} \\ J^*\Lambda^* & J^*\Lambda\Lambda^*(I - J^{*\prime}) & J^*\Lambda^* J^{*\prime} \end{pmatrix},$$

this completes the proof. □

B.2 Some proofs of technical results in Chapter 3

Proof of Proposition 3.1.1. By vectorizing the model in (3.1), we have

$$\text{Vec}(\mathbb{Y}) = Q\text{Vec}(\boldsymbol{\delta}) + \text{Vec}(\mathbb{U}).$$

If $\boldsymbol{\tau_1}, \cdots, \boldsymbol{\tau_{m_0}}$ and m_0 are known, we minimize with respect to Vec($\boldsymbol{\delta}$), the optimization function

$$\mathscr{L}_1(\boldsymbol{\delta}) = [\text{Vec}(\mathbb{Y}) - Q\text{Vec}(\boldsymbol{\delta})]'[\text{Vec}(\mathbb{Y}) - Q\text{Vec}(\boldsymbol{\delta})],$$

with respect to Vec($\boldsymbol{\delta}$). Taking the derivative of $\mathscr{L}_1(\boldsymbol{\delta})$ with respect to Vec($\boldsymbol{\delta}$) and setting it equal to 0, we have

$$-2\text{Vec}(\mathbb{Y})'Q + 2\text{Vec}(\boldsymbol{\delta})'Q'Q = 0. \tag{B.1}$$

Solving for Vec($\boldsymbol{\delta}$), we get

$$\text{Vec}(\boldsymbol{\delta}) = (Q'Q)^{-1}Q'\text{Vec}(\mathbb{Y}).$$

Further, we have

$$\frac{\partial^2 \mathscr{L}_1(\delta)}{\partial \text{Vec}(\delta)\text{Vec}(\delta)'} = Q'Q,$$

which is a positive definite matrix. Then, by noting that $(Q'Q)^{-1}Q' = (\bar{Z}'\bar{Z})^{-1}\bar{Z}' \otimes \bigotimes_{i=d}^{1} I_{q_i}$, we have

$$\mathrm{Vec}(\hat{\boldsymbol{\delta}}(\tau)) = [(\bar{Z}'\bar{Z})^{-1}\bar{Z}' \otimes \bigotimes_{i=d}^{1} I_{q_i}]\mathrm{Vec}(\mathbb{Y}).$$

Therefore, returning to tensor form, we have

$$\mathrm{Vec}(\hat{\boldsymbol{\delta}}(\tau)) = \mathbb{Y} \times_{d+1} (\bar{Z}'\bar{Z})^{-1}\bar{Z}'.$$

Further, if $\boldsymbol{\tau}_1, \cdots, \boldsymbol{\tau}_{m_0}$ and m_0 are known, to find $\tilde{\boldsymbol{\delta}}(\tau)$, we minimize the Lagrangian, $\mathscr{L}_2(\boldsymbol{\delta})$, with respect to Vec($\boldsymbol{\delta}$)

$$\mathscr{L}_2(\boldsymbol{\delta}) = [\mathrm{Vec}(\mathbb{Y}) - Q\mathrm{Vec}(\boldsymbol{\delta})]'[\mathrm{Vec}(\mathbb{Y}) - Q\mathrm{Vec}(\boldsymbol{\delta})] + \lambda'(R\mathrm{Vec}(\boldsymbol{\delta}) - \mathrm{Vec}(\boldsymbol{r})),$$

where λ is an arbitrary $l_1 l_2 \cdots l_{d+1}$−column vector of Lagrangian multipliers.
To this end, taking the derivative of both sides with respect to Vec($\boldsymbol{\delta}$), we get

$$-2\mathrm{Vec}(\mathbb{Y})'Q + 2\mathrm{Vec}(\boldsymbol{\delta})'Q'Q + \lambda'R = 0. \tag{B.2}$$

Solving for Vec($\boldsymbol{\delta}$), we have

$$\mathrm{Vec}(\boldsymbol{\delta}) = (Q'Q)^{-1}Q'\mathrm{Vec}(\mathbb{Y}) - 1/2(Q'Q)^{-1}R'\lambda. \tag{B.3}$$

Multiplying both sides of (B.3) by R and subtracting Vec($\boldsymbol{r}$), we get

$$R\mathrm{Vec}(\boldsymbol{\delta}) - \mathrm{Vec}(\boldsymbol{r}) = R(Q'Q)^{-1}Q'\mathrm{Vec}(\mathbb{Y}) - 1/2R(Q'Q)^{-1}R'\lambda - \mathrm{Vec}(\boldsymbol{r}).$$

Hence, under the restriction in (3.2), we have,

$$R(Q'Q)^{-1}Q'\mathrm{Vec}(\mathbb{Y}) - 1/2R(Q'Q)^{-1}R'\lambda - \mathrm{Vec}(r) = 0.$$

Finally, solving for λ, we get

$$\lambda = 2[R(Q'Q)^{-1}R']^{-1}(R(Q'Q)^{-1}Q'\mathrm{Vec}(Y) - \mathrm{Vec}(\boldsymbol{r})) \tag{B.4}$$

Substituting λ from (B.4) into (B.3), we have

$$\mathrm{Vec}(\boldsymbol{\delta}) = (Q'Q)^{-1}Q'\mathrm{Vec}(\mathbb{Y}) - (Q'Q)^{-1}R'[R(Q'Q)^{-1}R']^{-1}[R(Q'Q)^{-1}Q'\mathrm{Vec}(\mathbb{Y}) - \mathrm{Vec}(\boldsymbol{r})].$$

Using the fact that $\mathrm{Vec}(\hat{\boldsymbol{\delta}}(\boldsymbol{\tau})) = (Q'Q)^{-1}Q'\mathrm{Vec}(\mathbb{Y})$, the above equation can be simplified as

$$\mathrm{Vec}(\boldsymbol{\delta}) = \mathrm{Vec}(\hat{\boldsymbol{\delta}}(\boldsymbol{\tau})) - (Q'Q)^{-1}R'[R(Q'Q)^{-1}R']^{-1}[R\mathrm{Vec}(\hat{\boldsymbol{\delta}}(\boldsymbol{\tau})) - \mathrm{Vec}(\boldsymbol{r})].$$

Using Kronecker product properties, we have

$$Q'Q = \left(\bar{Z} \otimes \bigotimes_{i=d}^{1} I_{q_i}\right)' \left(\bar{Z} \otimes \bigotimes_{i=d}^{1} I_{q_i}\right) = \left(\bar{Z}' \otimes \bigotimes_{i=d}^{1} I_{q_i}\right)\left(\bar{Z} \otimes \bigotimes_{i=d}^{1} I_{q_i}\right) = \left(\bar{Z}'\bar{Z} \otimes \bigotimes_{i=d}^{1} I_{q_i}\right).$$

Hence,

$$(Q'Q)^{-1} = (\bar{Z}'\bar{Z})^{-1} \otimes \bigotimes_{i=d}^{1} I_{q_i}.$$

Similarly, it can be shown that

$$R'[R(Q'Q)^{-1}R']^{-1}R = \bigotimes_{i=d+1}^{1} \mathbb{J}_i,$$

and

$$R'[R(Q'Q)^{-1}R']^{-1} = \bigotimes_{i=d+1}^{1} \mathbb{G}_i.$$

Hence, we have

$$\mathrm{Vec}(\tilde{\boldsymbol{\delta}}(\boldsymbol{\tau})) = \mathrm{Vec}(\hat{\boldsymbol{\delta}}(\boldsymbol{\tau})) - \left(\bigotimes_{i=d+1}^{1} \mathbb{J}_i\right)\mathrm{Vec}(\hat{\boldsymbol{\delta}}(\boldsymbol{\tau})) + \left(\bigotimes_{i=d+1}^{1} \mathbb{G}_i\right)\mathrm{Vec}(\boldsymbol{r}). \tag{B.5}$$

Therefore, converting back to tensor mode by "un-vecing", we have

$$\tilde{\boldsymbol{\delta}}(\boldsymbol{\tau}) = \hat{\boldsymbol{\delta}}(\boldsymbol{\tau}) - \hat{\boldsymbol{\delta}}(\boldsymbol{\tau})(\bigtimes_{i=1}^{d+1})_i \mathbb{J}_i + \boldsymbol{r}(\bigtimes_{i=1}^{d+1})_i \mathbb{G}_i,$$

this completes the proof. □

Proof of Proposition 3.2.7. We have by Corollary 3.2.5,

$$\sum_{k=-\infty}^{\infty} a_k^{-1}\mathrm{E}\left(V^2_{L_p,k,s_1,\cdots,s_{d+1}}(l)\right) = \sum_{i=l+1}^{l+L_p} [a_0^{-1}\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^i_{p,-\infty})) + a_1^{-1}\mathrm{E}(D^2_{i,0,s_1,\cdots,s_{d+1}})$$
$$+ \sum_{k=1}^{\infty}(a_{k+1}^{-1} - a_k^{-1})\mathrm{E}(D^2_{i,k,s_1,\cdots,s_{d+1}}) + \sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i-k}_{p,-\infty}))].$$

Thus,

$$\begin{aligned}\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\sum_{k=-\infty}^{\infty}a_k^{-1}\mathrm{E}\left(V^2_{L_p,k,s_1,\cdots,s_{d+1}}\right) &= a_1^{-1}\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}(D^2_{i,0,s_1,\cdots,s_{d+1}})\\ &+\sum_{i=l+1}^{l+L_p}a_0^{-1}\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^i_{p,-\infty}))\\ &+\sum_{k=1}^{\infty}(a_{k+1}^{-1}-a_k^{-1})\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}(D^2_{i,k,s_1,\cdots,s_{d+1}})\\ &+\sum_{k=1}^{\infty}(a_k^{-1}-a_{k-1}^{-1})\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i-k}_{p,-\infty})).\end{aligned}$$

Now, using mixingale properties, we have

$$\sum_{s_{d+1}=1}^{q_{d+1}}\sum_{s_d=1}^{q_d}\cdots\sum_{s_1=1}^{q_1}\mathrm{E}(\mathrm{E}^2(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i-k}_{p,-\infty}))\le c^2_{p,i}\psi^2(k),$$

and

$$\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}(D^2_{i,k,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i-k}_{p,-\infty})\le c^2_{p,i}\psi^2(k+1).$$

This gives

$$\begin{aligned}\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\sum_{k=-\infty}^{\infty}a_k^{-1}\mathrm{E}\left(V^2_{L_p,k,s_1,\cdots,s_{d+1}}(l)\right)&\le\left(\sum_{i=l+1}^{l+L_p}c^2_{p,i}\right)\{a_0^{-1}\left(\sum_{j=0}^{1}\psi^2(j)\right)\\ +(a_1^{-1}\psi^2(1)-a_0^{-1}\psi^2(1))+\sum_{j=2}^{\infty}(a_j^{-1}-a_{j-1}^{-1})\psi^2(j)&+\sum_{k=1}^{\infty}(a_k^{-1}-a_{k-1}^{-1})\psi^2(k)\}.\end{aligned}$$

Then, we have

$$\begin{aligned}\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\sum_{k=-\infty}^{\infty}a_k^{-1}\mathrm{E}\left(V^2_{L_p,k,s_1,\cdots,s_{d+1}}\right)&\le\left(\sum_{i=l+1}^{l+L_p}c^2_{p,i}\right)\{a_0^{-1}\left(\psi^2(0)+\psi^2(1)\right)\\ &+2\sum_{k=1}^{\infty}(a_k^{-1}-a_{k-1}^{-1})\psi^2(k)\}<\infty.\end{aligned}$$

This completes the proof. □

Proof of Lemma 3.2.3. Let $Q_{j,k}(l)=\sum_{i=l+1}^{l+j}\left[\mathrm{E}(\mathbb{X}_{p,i}|\mathscr{F}^{i+k}_{p,-\infty})-\mathrm{E}(\mathbb{X}_{p,i}|\mathscr{F}^{i+k-1}_{p,-\infty})\right]$. Then, by Corollary 3.2.2, we get $\sum_{i=l+1}^{l+j}\mathbb{X}_{p,i}=\sum_{k=-\infty}^{\infty}Q_{j,k}(l)$ *a.s.* Define a sequence $\{a_k\}_{-\infty}^{\infty}$ such that conditions in Lemma 3.2.2 hold. Then, by Cauchy-Schwarz's inequality, for $s_1=1,\cdots,q_1$,

$s_2 = 1, \cdots, q_2, \cdots s_{d+1} = 1, \cdots, q_{d+1}$, we have

$$\left(\sum_{i=l+1}^{l+j} X_{p,i,s_1,\cdots,s_{d+1}}\right)^2 = \left(\sum_{k=-\infty}^{\infty} Q_{j,k,s_1,\cdots,s_{d+1}}(l)\right)^2 \leqslant \left(\sum_{k=-\infty}^{\infty} a_k\right)\left(\sum_{k=-\infty}^{\infty} a_k^{-1} Q^2_{j,k,s_1,\cdots,s_{d+1}}\right).$$

Thus, $\mathrm{E}\left(\max_{j\leqslant L}\left(\sum_{i=1}^{j} X_{p,i,s_1,\cdots,s_{d+1}}\right)^2\right) \leqslant \left(\sum_{k=-\infty}^{\infty} a_k\right)\left(\sum_{k=-\infty}^{\infty} a_k^{-1}\mathrm{E}\left(\max_{j\leqslant L} Q^2_{j,k,s_1,\cdots,s_{d+1}}(l)\right)\right).$

For each k,s_1, $s_2, \cdots, s_{d+1}$, the sequence $\{Q_{i,s_1,\cdots,s_{d+1}}(l), \mathscr{F}^{i+k}_{p,-\infty}, 1+l \leqslant i \leqslant l+L\}$ is a martingale. Then, by Doob's inequality, we have

$$\mathrm{E}\left(\max_{j\leqslant L}\left(\sum_{i=l+1}^{l+j} X_{p,i,s_1,\cdots,s_{d+1}}\right)^2\right) \leqslant 4\left(\sum_{k=-\infty}^{\infty} a_k\right)\left(\sum_{k=-\infty}^{\infty} a_k^{-1}\mathrm{E}\left(Q^2_{L,k,s_1,\cdots,s_{d+1}}(l)\right)\right).$$

Using Proposition 3.2.7, we get

$$\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left(\max_{j\leqslant L}\left(\sum_{i=l+1}^{l+j} X_{p,i,s_1,\cdots,s_{d+1}}\right)^2\right) \leqslant 4\left(\sum_{k=-\infty}^{\infty} a_k\right)\left(\sum_{i=l+1}^{l+L} c^2_{p,i}\right)\mathcal{K}(\psi) < \infty,$$

where $\mathcal{K}(\psi) = a_0^{-1}\left(\psi^2(0) + \psi^2(1)\right) + 2\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\psi^2(k)$, this completes the proof. □

Proof of Proposition 3.2.8. 1) We have

$$\mathcal{J}_1(a,b,m) = \mathrm{E}\left(\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\max_{j\leq L}\frac{\bar{U}^2_{1,j,s_1,\cdots,s_{d+1}}}{\tilde{v}^2_L(l)}\mathbb{I}\left(\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\max_{j\leq L}\frac{\bar{U}^2_{1,j,s_1,\cdots,s_{d+1}}}{\tilde{v}^2_L(l)} > b/9\right)\right).$$

Since $\varphi^2(k)/\varphi^2(m) \geq 1$ for $k < m$, we have

$$\begin{aligned}
&\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}[\mathrm{E}^2(U_{1,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i-k}_{p,-\infty})]\\
&= \sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}[\mathrm{E}^2((\mathrm{E}_{i+m}X^a_{p,i,s_1,\cdots,s_{d+1}}) - \mathrm{E}_{i-m}X^a_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i-k}_{p,-\infty})]\\
&= \sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}[(\mathrm{E}(X^a_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i-k}_{p,-\infty}) - \mathrm{E}((\mathrm{E}_{i-m}(X^a_{p,i,s_1,\cdots,s_{d+1}})|\mathscr{F}^{i-k}_{p,-\infty}))^2].
\end{aligned}$$

Hence,

$$\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}[\mathrm{E}^2(U^*_{1,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i-k}_{p,-\infty})]\begin{cases}\leq \frac{2q_1\cdots q_d q_{d+1}a^2c^2_{p,i}}{\varphi^2(m)}\psi^2(k), & \text{if } k < m,\\ = 0, & \text{if} k \geq m.\end{cases}$$

Similarly,

$$\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}[(U_{1,i,s_1,\cdots,s_{d+1}} - \mathrm{E}(U_{1,i,s_1,\cdots,s_{d+1}}|\mathscr{F}_{p,-\infty}^{i+k}))^2]$$
$$= \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}[(\mathrm{E}_{i+m} X_{p,i,s_1,\cdots,s_{d+1}}^a - \mathrm{E}_{i-m} X_{p,i,s_1,\cdots,s_{d+1}}^a)$$
$$- \mathrm{E}((\mathrm{E}_{i+m} X_{p,i,s_1,\cdots,s_{d+1}}^a - \mathrm{E}_{i-m} X_{p,i,s_1,\cdots,s_{d+1}}^a)|\mathscr{F}_{p,-\infty}^{i+k})^2]$$
$$= \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}[(\mathrm{E}(\mathrm{E}_{i+m} X_{p,i,s_1,\cdots,s_{d+1}}^a|\mathscr{F}_{p,-\infty}^{i+k}) - \mathrm{E}(X_{p,i,s_1,\cdots,s_{d+1}}^a|\mathscr{F}_{p,-\infty}^{i+k}))^2].$$

Then, we have

$$\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left[(U_{1,i,s_1,\cdots,s_{d+1}} - \mathrm{E}(U_{1,i,s_1,\cdots,s_{d+1}}|\mathscr{F}_{p,-\infty}^{i+k}))^2\right]$$
$$\leq \frac{2q_1 \cdots q_d q_{d+1} a^2 c_{p,i}^2}{\varphi^2(m)} \psi^2(k+1)\mathbb{I}(k < m).$$

Hence, $\{\mathbb{U}_{1,i}, \mathscr{F}_{p,-\infty}^i\}$ is a mixingale of size $-1/2$, and it follows from Corollary 3.2.2 that $\bar{U}_{1,j} = \sum_{i=1}^{j} U_{1,i} = \sum_{k=-\infty}^{\infty} \hat{U}_{1,j,k}$ *a.s.*, where $\hat{U}_{1,j,k} = \sum_{i=1}^{j} [\mathrm{E}(\mathbb{U}_{1,i}|\mathscr{F}_{p,-\infty}^{i+k}) - \mathrm{E}(\mathbb{U}_{1,i}|\mathscr{F}_{p,-\infty}^{i+k-1})]$. Let $\{B_k\}_{k=-\infty}^{\infty}$ be a sequence of positive real numbers. By Jensen's inequality,

$$\left(\sum_{k=-\infty}^{\infty} \hat{U}_{1,j,k,s_1,\cdots,s_{d+1}}\right)^4 = \left(\sum_{k=-\infty}^{\infty} B_k\right)^4 \left(\sum_{k=-\infty}^{\infty} \frac{\hat{U}_{1,j,k,s_1,\cdots,s_{d+1}}}{B_k} \frac{B_k}{\sum_{k=-\infty}^{\infty} B_k}\right)^4$$
$$\leq \left(\sum_{k=-\infty}^{\infty} B_k\right)^4 \sum_{k=-\infty}^{\infty} \frac{\hat{U}_{1,j,k,s_1,\cdots,s_{d+1}}^4}{B_k^4} \frac{B_k}{\sum_{k=-\infty}^{\infty} B_k},$$

which results in

$$\left(\sum_{k=-\infty}^{\infty} \hat{U}_{1,j,k,s_1,\cdots,s_{d+1}}\right)^4 \leq \left(\sum_{k=-\infty}^{\infty} B_k\right)^3 \sum_{k=-\infty}^{\infty} \frac{\hat{U}_{1,j,k,s_1,\cdots,s_{d+1}}^4}{B_k^3}.$$

Note that

$$\begin{aligned}
&\mathrm{E}(U_{1,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i+k}_{p,-\infty}) - \mathrm{E}(U_{1,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i+k-1}_{p,-\infty})\\
&= \mathrm{E}(\mathrm{E}_{i+m}X^a_{p,i,s_1,\cdots,s_{d+1}} - \mathrm{E}_{i-m}X^a_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i+k}_{p,-\infty})\\
&- \mathrm{E}(\mathrm{E}_{i+m}X^a_{p,i,s_1,\cdots,s_{d+1}} - \mathrm{E}_{i-m}X^a_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i+k-1}_{p,-\infty}).
\end{aligned}$$

Hence,

$$|\mathrm{E}(U_{1,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i+k}_{p,-\infty}) - \mathrm{E}(U_{1,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^{i+k-1}_{p,-\infty})| \begin{cases} \leq 4a, & \text{if } k \leq m,\\ = 0, & \text{if } k > m. \end{cases}$$

Thus, we have

$$\begin{aligned}
\mathrm{E}\left(\max_{j\leq L} \bar{U}^4_{1,j,s_1,\cdots,s_{d+1}}\right) &\leq \left(\sum_{k=-\infty}^{\infty} B_k\right)^3 \sum_{k=-\infty}^{\infty} \frac{\mathrm{E}(\max_{j\leq L} \hat{U}^4_{1,j,k,s_1,\cdots,s_{d+1}})}{B_k^3}\\
&= \left(\sum_{k=-\infty}^{\infty} B_k\right)^3 \sum_{k=-m}^{m} \frac{\mathrm{E}(\max_{j\leq L} \hat{U}^4_{1,j,k,s_1,\cdots,s_{d+1}})}{B_k^3},
\end{aligned}$$

and hence, by Doob's inequaltiy, we get

$$\mathrm{E}\left(\max_{j\leq L} \bar{U}^4_{1,j,s_1,\cdots,s_{d+1}}\right) \leq (4/3)^4 \left(\sum_{k=-\infty}^{\infty} B_k\right)^3 \sum_{k=-m}^{m} \frac{\mathrm{E}(\hat{U}^4_{1,L,k,s_1,\cdots,s_{d+1}})}{B_k^3}.$$

Thus, by Lemma 3.1 of McLeish (1977),

$$\mathrm{E}(\hat{U}^4_{L,k,s_1,\cdots,s_{d+1}}) \leq 10(4a)^4(\tilde{v}^2_L(l))^2.$$

Therefore,

$$\mathrm{E}\left(\max_{j\leq L} \bar{U}^4_{1,j,s_1,\cdots,s_{d+1}}\right) \leq (4/3)^4 \left(\sum_{k=-\infty}^{\infty} B_k\right)^3 10(4a)^4\tilde{v}^4_L(l) \sum_{k=-m}^{m} B_k^{-3}.$$

Hence, we choose $B_j = 1$ if $|j| \leq m$ and $B_j = |j|^{-1}\boldsymbol{\kappa}^{-1}(|j|)$, otherwise. Thus,

$$\mathrm{E}\left(\max_{j\leq L} \bar{U}^4_{1,j,s_1,\cdots,s_{d+1}}\right) \leq 10(4/3)^4 \left(2m+1+2\sum_{j=m+1}^{\infty} j^{-1}\boldsymbol{\kappa}^{-1}(j)\right)^3 (2m+1)(4a)^4\tilde{v}^4_L(l).$$

Then, by Cauchy-Schwarz's inequality,

$$\mathcal{J}_1(a,b,m) \leq \sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}^{1/2}\left(\max_{j\leq L} \bar{U}^4_{1,j,s_1,\cdots,s_{d+1}}/\tilde{v}^4_L(l)\right)$$
$$\times\, \mathrm{E}^{1/2}\left(\mathbb{I}(\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}})\max_{j\leq L} \bar{U}^2_{1,j,s_1,\cdots,s_{d+1}}/\tilde{v}^2_L(l) > b/9)^2\right).$$

Therefore, for fixed (m,a) and any $\epsilon > 0$, we can pick b such that $\mathcal{J}_1(a,b,m) \leq \epsilon$, proving the first statement.

2) One can verify that

$$\begin{aligned}||\mathrm{E}(U_{2,i}|\mathscr{F}^{i-k}_{p,-\infty})||_2 &= ||\mathrm{E}((\mathbb{X}_{p,i} - \mathrm{E}_{i+m}\mathbb{X}_{p,i} + \mathrm{E}_{i-m}\mathbb{X}_{p,i})|\mathscr{F}^{i-k}_{p,-\infty})||_2\\ &= ||\mathrm{E}(\mathrm{E}_{i-m}|\mathscr{F}^{i-k}_{p,-\infty})||_2 \leq c_{p,i}\psi(m\vee k),\end{aligned}$$

and similarly,

$$\begin{aligned}||\mathbb{U}_{2,i} - \mathrm{E}(\mathbb{U}_{2,i}|\mathscr{F}^{i+k}_{p,-\infty})||_2 &= ||(\mathbb{X}_{p,i} - \mathrm{E}_{i+m}\mathbb{X}_{p,i} + \mathrm{E}_{i-m}\mathbb{X}_{p,i})\\ &\quad - \mathrm{E}((\mathbb{X}_{p,i} - \mathrm{E}_{i+m}\mathbb{X}_{p,i} + \mathrm{E}_{i-m}\mathbb{X}_{p,i})|\mathscr{F}^{i+k}_{p,-\infty})||_2\\ &= ||(\mathbb{X}_{p,i} - \mathrm{E}_{i+m}\mathbb{X}_{p,i}) - \mathrm{E}((\mathbb{X}_{p,i} - \mathrm{E}_{i+m}\mathbb{X}_{p,i})|\mathscr{F}^{i+k}_{p,-\infty})||_2.\end{aligned}$$

This gives

$$||\mathrm{E}(\mathbb{U}_{2,i}|\mathscr{F}^{i-k}_{p,-\infty})||_2 \leq c_{p,i}\psi(m\vee k+1).$$

Then, $\{\mathbb{U}_{2,i}, \mathscr{F}^i_{p,-\infty}\}$ is a mixingale with mixing function $\hat{\psi}(k) = \psi(m\vee k)$.

Therefore, it follows from Lemma 3.2.3 that

$$\mathcal{J}_2(m) \leq 4\left(\sum_{k=-\infty}^{\infty} a_k\right)\left(\sum_{i=1}^{L^*} c^2_{p,i}\right)\{a_0^{-1}(\hat{\psi}^2(0)+\hat{\psi}^2(1)) + 2\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\hat{\psi}^2(k)\}$$

where

$$a_0^{-1}(\hat{\psi}^2(0)+\hat{\psi}^2(1)) + 2\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\hat{\psi}(k)$$
$$= a_0^{-1}(\psi^2(m)+\psi^2(m)) + 2\sum_{k=1}^{m}(a_k^{-1} - a_{k-1}^{-1})\psi^2(m) + 2\sum_{k=m+1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\hat{\psi}^2(k).$$

Noting that

$\{a_0^{-1}(\hat{\psi}^2(0) + \hat{\psi}^2(1)) + 2\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\hat{\psi}^2(k)\}$

$= a_0^{-1}(\psi^2(m) + \psi^2(m) + 2m(O(\boldsymbol{\kappa}(m))\psi^2(m) + 2\sum_{j=m+1}^{\infty}(O(\boldsymbol{\kappa}(j))\psi^2(j))$

$= O(m^{-1}\boldsymbol{\kappa}^{-1}(m)) + O(\boldsymbol{\kappa}^{-1}(m)) + O(\sum_{j=m+1}^{\infty} j^{-1}\boldsymbol{\kappa}^{-1}(k)),$

we conclude that we can choose m such that $\mathcal{J}_2(m) \le \epsilon$, this completes the proof of the second statement.

3) Since $\frac{\psi(k)}{\varphi(m)} \ge 1$ for $k < m$, then by Jensen's inequality, we have

$$\begin{aligned}||\mathrm{E}(\mathbb{U}_{3,j}\mathscr{F}_{p,-\infty}^{i-k})||_2 = ||\mathrm{E}(\mathrm{E}_{i+m}(X_{p,i,s_1,\cdots,s_{d+1}} - X^a_{p,i,s_1,\cdots,s_{d+1}})|\mathscr{F}_{p,-\infty}^{i-k})\\ - \mathrm{E}_{i-m}(X_{p,i,s_1,\cdots,s_{d+1}} - X^a_{p,i,s_1,\cdots,s_{d+1}})|\mathscr{F}_{p,-\infty}^{i-k})||_2.\end{aligned}$$

This gives

$$||\mathrm{E}(\mathbb{U}_{3,j}|\mathscr{F}_{p,-\infty}^{i-k})||_2 \le ||\mathrm{E}_{i+m}(X_{p,i,s_1,\cdots,s_{d+1}} - X^a_{p,i,s_1,\cdots,s_{d+1}})||_2.$$

Then,

$$||\mathrm{E}(\mathbb{U}_{3,j}|\mathscr{F}_{p,-\infty}^{i-k})||_2 \le c_{p,i}\sqrt{\mathrm{E}(X^2_{p,i,s_1,\cdots,s_{d+1}}/c^2_{p,i}\mathbb{I}(X^2_{p,i,s_1,\cdots,s_{d+1}}/c^2_{p,i} > a^2))}.$$

This gives

$$||\mathrm{E}(\mathbb{U}_{3,j}|\mathscr{F}_{p,-\infty}^{i-k})||_2 \le \frac{\psi(k)}{\varphi(m)}c_{p,i}\sqrt{\sup_j \mathrm{E}(X^2_{p,j,s_1,\cdots,s_{d+1}}/c^2_{p,j}\mathbb{I}(X^2_{p,j,s_1,\cdots,s_{d+1}}/c^2_{p,j} > a^2))},$$

if $k < m$, and is equal to 0 for $k \le m$. Similarly,

$$\begin{aligned}&||\mathbb{U}_{3,j} - \mathrm{E}(\mathbb{U}_{3,j}|\mathscr{F}_{p,-\infty}^{i+k})||_2 \le ||\mathrm{E}_{i+m}(X_{p,i,s_1,\cdots,s_{d+1}} - X^a_{p,i,s_1,\cdots,s_{d+1}})||_2\\ &\le \frac{\varphi(k+1)}{\psi(m)}c_{p,i}\sqrt{\sup_j \mathrm{E}(X^2_{p,j,s_1,\cdots,s_{d+1}}/c^2_{p,j}\mathbb{I}(X^2_{p,j,s_1,\cdots,s_{d+1}}/c^2_{p,j} > a^2))}\mathbb{I}(k < m).\end{aligned}$$

Therefore, $\{\mathbb{U}_{3,i}, \mathscr{F}^i_{p,-\infty}\}$ is also a L^2-mixingale with functions $\hat{\psi}(k) = \psi(k)$ and $c^2_{p,i}$ becomes $\frac{c^2_{p,i}}{\psi^2(m)}\sup_j \mathrm{E}(X^2_{p,j,s_1,\cdots,s_{d+1}}/c^2_{p,j}\mathbb{I}(X^2_{p,j,s_1,\cdots,s_{d+1}}/c^2_{p,j} > a^2)).$

Hence, for a fixed m with $\mathcal{J}_2(m) \leq \epsilon$, we can choose an a significantly large enough such that $\mathcal{J}_3(a, m) \leq \epsilon$, this completes the proof of the last statement.

□

Proof of Proposition 3.2.9. 1. We have

$$S\mathbb{I}(S > b) \leq p\sum_{j=1}^{p} A_j \mathbb{I}\left(p\sum_{j=1}^{p} A_j > b\right).$$

Further,

$$\begin{aligned}p\sum_{j=1}^{p} A_j \mathbb{I}\left(p\sum_{j=1}^{p} A_j > b\right) &= pA_1\mathbb{I}\left(\sum_{j=1}^{p} A_j > b/p\right) + p\sum_{j=2}^{p} A_j\mathbb{I}\left(\sum_{j=1}^{p} A_j > b/p\right)\\ &= pA_1\mathbb{I}\left(\sum_{j=1}^{p} A_j > b/p, A_1 > b/p^2\right) + pA_1\mathbb{I}\left(\sum_{j=1}^{p} A_j > b/p, A_1 \leq b/p^2\right)\\ &+ p\sum_{j=2}^{p} A_j\mathbb{I}\left(\sum_{j=1}^{p} A_j > b/p\right).\end{aligned}$$

This gives

$$\begin{aligned}p\sum_{j=1}^{p} A_j \mathbb{I}\left(p\sum_{j=1}^{p} A_j > b\right) &\leq pA_1\mathbb{I}\left(A_1 > b/p^2\right) + pA_1\mathbb{I}\left(\sum_{j=1}^{p} A_j > b/p, A_1 \leq b/p^2\right)\\ &+ p\sum_{j=2}^{p} A_j\mathbb{I}\left(\sum_{j=1}^{p} A_j > b/p\right)\end{aligned}$$

Then, since $p \geq 2$, we have

$$pA_1\mathbb{I}\left(\sum_{j=1}^{p} A_j > b/p, A_1\right) \leq p\sum_{j=2}^{p} A_j\mathbb{I}\left(\sum_{j=1}^{p} A_j > (p-1)b/p^2\right).$$

Then,

$$\begin{aligned}p\sum_{j=1}^{p} A_j \mathbb{I}\left(p\sum_{j=1}^{p} A_j > b\right) &\leq pA_1\mathbb{I}\left(A_1 > b/p^2\right) + p\sum_{j=2}^{p} A_j\mathbb{I}\left(\sum_{j=1}^{p} A_j > (p-1)b/p^2\right)\\ &+ p\sum_{j=2}^{p} A_j\mathbb{I}\left(\sum_{j=1}^{p} A_j > b/p\right).\end{aligned}$$

Hence, we have

$$p\sum_{j=1}^{p} A_j \mathbb{I}\left(p\sum_{j=1}^{p} A_j > b\right) \leq pA_1\mathbb{I}\left(A_1 > b/p^2\right) + 2p\sum_{j=2}^{p} A_j.$$

Since $p \geq 2$, we have

$$p\sum_{j=1}^{p} A_j \mathbb{I}\left(p\sum_{j=1}^{p} A_j > b\right) \leq p^2 A_1\mathbb{I}\left(A_1 > b/p^2\right) + p^2\sum_{j=2}^{p} A_j.$$

Therefore, taking the expected value through both sides we have,

$$\mathrm{E}\left[p\sum_{j=1}^{p} A_j \mathbb{I}\left(p\sum_{j=1}^{p} A_j > b\right)\right] \leq p^2\mathrm{E}\left[A_1\mathbb{I}\left(A_1 > b/p^2\right)\right] + p^2\mathrm{E}\left[\sum_{j=2}^{p} A_j\right].$$

This proves Part 1.

2. The proof of Part 2 follows from Part 1 along with the fact that if $q \geq 1$, the function $f(x) = x^q$ is convex on $(0, +\infty)$. This completes the proof.

□

Proof of Lemma 3.2.4. Let $\bar{U}_{t,j,s_1,\cdots,s_{d+1}}(l) = \sum_{i=l+1}^{l+j} U_{t,i,s_1,\cdots,s_{d+1}}$, $t = 1, 2, 3$. Using the convexity of the quadratic function, we have

$$\left(S_{l+j,s_1,\cdots,s_{d+1}} - S_{l,s_1,\cdots,s_{d+1}}\right)^2 = \left(\sum_{i=l+1}^{l+j} X_{p,i,s_1,\cdots,s_{d+1}}\right)^2$$
$$= \left[\sum_{i=l+1}^{l+j} (U_{1,i,s_1,\cdots,s_{d+1}} + U_{2,i,s_1,\cdots,s_{d+1}} + U_{3,i,s_1,\cdots,s_{d+1}})\right]^2$$
$$\leq 3(\bar{U}^2_{1,j,s_1,\cdots,s_{d+1}}(l) + \bar{U}^2_{2,j,s_1,\cdots,s_{d+1}}(l) + \bar{U}^2_{3,j,s_1,\cdots,s_{d+1}}(l)).$$

Then,

$$\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}}\left(S_{l+j,s_1,\cdots,s_{d+1}} - S_{l,s_1,\cdots,s_{d+1}}\right)^2 \leq 3\left(\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}} \bar{U}^2_{1,j,s_1,\cdots,s_{d+1}}(l)\right)$$
$$+ 3\left(\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}} \bar{U}^2_{2,j,s_1,\cdots,s_{d+1}}(l)\right) + 3\left(\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}} \bar{U}_{3,j,s_1,\cdots,s_{d+1}}(l)\right).$$

This implies

$$\max_{j\leq L}\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}}\left(S_{l+j,s_1,\cdots,s_{d+1}} - S_{l,s_1,\cdots,s_{d+1}}\right)^2 / \tilde{v}^2_L(l) \leq 3(A_p(a,m) + B_p(m) + C_p(a,m)),$$

where $A_p(a,m;L)$, $B_p(m;L)$ and $C_p(a,m;L)$ are as defined in Proposition 3.2.8. Then, from Proposition 3.2.9, we have

$$\mathrm{E}\left(\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}}\mathscr{S}_{j,s_1,\cdots,s_{d+1}}(l)\mathbb{I}\left(\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}}\mathscr{S}_{j,s_1,\cdots,s_{d+1}}(l)>b\right)\right)$$
$$\leq 9(\mathcal{J}_1(a,b,m)+\mathcal{J}_2(m)+\mathcal{J}_3(a,m)),$$

where $\mathscr{S}_{j,s_1,\cdots,s_{d+1}}(l)=\max_{j\leq L}\left(S_{l+j,s_1,\cdots,s_{d+1}}-S_{l,s_1,\cdots,s_{d+1}}\right)^2\tilde{v}_L^2(l)$ and $\mathcal{J}_1(a,b,m),\mathcal{J}_2(m),\mathcal{J}_3(a,m)$ are as defined in Proposition 3.2.8. Thus, from Proposition 3.2.8, for any ϵ' we can choose an m,a,b such that $I_1(a,b,m)<\epsilon'/27, I_2(m)<\epsilon'/27,$ and $I_3(a,m)<\epsilon'/27$, this completes the proof. □

Proof of Proposition 3.2.11. We have

$$\sum_{i=1}^{r_p}\mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}^*_{i-1})=\sum_{i=1}^{r_p}\sum_{t=(i-1)b_p+l_p+1}^{ib_p}\mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}^*_{i-1})=\sum_{t\in S_1}\mathrm{E}(\mathbb{X}_{p,t}|\mathscr{F}^*_{i-1}),$$

where $S_1=\{t: t\in\cup_{i=1}^{r_p}[(i-1)b_p+l_p+1, ib_p]\}$. It follows from Proposition 3.2.10 that $\{\mathrm{E}(\mathbb{X}_{p,i}|\mathscr{F}^*_{i-1}),\mathscr{F}^i_{p,-\infty}\}$ is an L^2-mixingale of size $-1/2$. Hence, by Lemma 3.2.3

$$\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}\left(\sum_{t\in S_1}\mathrm{E}(X_{p,t,s_1,\cdots,s_{d+1}}|\mathscr{F}^*_{i-1})\right)^2=O\left(\sum_{t\in S_2}c^2_{p,t}\psi(l_p)^{2\eta}\right)=O(T^{-\alpha}\psi(l_p)^{2\eta}),$$

and then

$$\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}\left(\sum_{t\in S_1}\mathrm{E}(X_{p,t,s_1,\cdots,s_{d+1}}|\mathscr{F}^*_{i-1})\right)^2=o(1).$$

Hence, $\sum_{i=1}^{r_p}\mathrm{E}(\mathbb{V}_{p,i}\mathscr{F}^*_{i-1})\xrightarrow[L_p\to\infty]{P}0$, this proves the first statement. Similarly, by the triangle inequality it can be shown that $\{\mathbb{X}_{p,i}-\mathrm{E}(\mathbb{X}_{p,i}|\mathscr{F}^*_{i-1}),\mathscr{F}^i_{p,-\infty}\}$ is an L^2-mixingale of size $-1/2$. Using Lemma 3.2.3, we have

$$\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}\left[\left(\sum_{i=1}^{r_p}(V_{p,i,s_1,\cdots,s_{d+1}}-\mathrm{E}(V_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^*_{i-1}))\right)^2\right]$$
$$=\sum_{s_1=1}^{q_1}\cdots\sum_{s_d=1}^{q_d}\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}\left[\left(\sum_{t\in S_1}(X_{p,i,s_1,\cdots,s_{d+1}}-\mathrm{E}(X_{p,i,s_1,\cdots,s_{d+1}}|\mathscr{F}^*_{i-1}))\right)^2\right]=o(1),$$

this implies that $\sum_{i=1}^{r_p}(\mathbb{V}_{p,i} - \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}^*_{i-1})) \xrightarrow[L_p\to\infty]{P} 0.$ □

Proof of Proposition 3.2.12. For $n = 1, \cdots, d+1$, we have

$$\sum_{i=1}^{r_p} \mathbb{W}_{p,i_{(n)}}\mathbb{W}'_{p,i_{(n)}} = \sum_{i=1}^{r_p} \mathbb{V}_{p,i_{(n)}}\mathbb{V}'_{p,i_{(n)}} - \sum_{i=1}^{r_p}[\mathbb{V}_{p,i_{(n)}}\mathbb{V}'_{p,i_{(n)}} - \mathbb{W}_{p,i_{(n)}}\mathbb{W}'_{p,i_{(n)}}] \tag{B.6}$$

By combining Proposition 3.2.11, Condition ($\mathscr{C}_6$) and Slutsky's theorem, we have

$$\sum_{i=1}^{r_p} \mathbb{W}_{p,i_{(n)}}\mathbb{W}'_{p,i_{(n)}} \xrightarrow[L_p\to\infty]{P} \Sigma_{p,n}, n = 1, \cdots, d+1,$$

which proves the first statement. To prove the second statement, note that $\{\mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}^*_i) - \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}^*_{i-1}), \mathscr{F}^*_i\}$ is an L^2-mixingale array of size $-1/2$ with mixingale magnitude indices $2c_{p,i}$. Also, let $\tilde{v}_i^2 = \sum_{t=(i-1)b_p+l_p+1}^{ib_p} c^2_{p,t}$, then,

$$\begin{aligned}
&\sum_{i=1}^{r_p}\sum_{s_1}^{q_1}\cdots\sum_{s_{d+1}}^{q_{d+1}} \mathrm{E}\left(W^2_{p,i,s_1,\cdots,s_{d+1}}\mathbb{I}\left(\sum_{s_1}^{q_1}\cdots\sum_{s_{d+1}}^{q_{d+1}} W^2_{p,i,s_1,\cdots,s_{d+1}} > \epsilon\right)\right)\\
&= \sum_{i=1}^{r_p}\sum_{s_1}^{q_1}\cdots\sum_{s_{d+1}}^{q_{d+1}} \mathrm{E}\left(W^2_{p,i,s_1,\cdots,s_{d+1}}/\tilde{v}_i^2\mathbb{I}\left(\|\mathbb{W}_{p,i}\|^2_{\mathrm{F}}/\tilde{v}_i^2 > \epsilon/\tilde{v}_i^2\right)\right)\tilde{v}_i^2\\
&\le \sum_{i=1}^{r_p}\sum_{s_1}^{q_1}\cdots\sum_{s_{d+1}}^{q_{d+1}} \mathrm{E}\left(W^2_{p,i,s_1,\cdots,s_{d+1}}/\tilde{v}_i^2\mathbb{I}\left(\|\mathbb{W}_{p,i}\|^2_{\mathrm{F}}/\tilde{v}_i^2 > \frac{\epsilon}{\max\limits_{1\le i\le r_p}\tilde{v}_i^2}\right)\right)\tilde{v}_i^2\\
&\le \sum_{i=1}^{r_p}\sum_{s_1}^{q_1}\cdots\sum_{s_{d+1}}^{q_{d+1}} \mathrm{E}\left(W^2_{p,i,s_1,\cdots,s_{d+1}}/\tilde{v}_i^2\mathbb{I}\left(\max_{1\le i\le r_p}\|\mathbb{W}_{p,i}\|^2_{\mathrm{F}}/\tilde{v}_i^2 > \frac{\epsilon}{\max\limits_{1\le i\le r_p}\tilde{v}_i^2}\right)\right)\tilde{v}_i^2.
\end{aligned}$$

So we have,

$$\begin{aligned}
&\sum_{i=1}^{r_p}\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left(W^2_{p,i,s_1,\cdots,s_{d+1}}\mathbb{I}\left(\|\mathbb{W}_{p,i}\|^2_{\mathrm{F}} > \epsilon\right)\right)\\
&\le \max_{1\le i\le r_p}\left[\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left(W^2_{p,i,s_1,\cdots,s_{d+1}}/\tilde{v}_i^2\mathbb{I}\left(\max_{1\le i\le r_p}\|\mathbb{W}_{p,i}\|^2_{\mathrm{F}}/\tilde{v}_i^2 > \frac{\epsilon}{\max\limits_{1\le i\le r_p}\tilde{v}_i^2}\right)\right)\right]\sum_{i=1}^{r_p}\tilde{v}_i^2.
\end{aligned}$$

Under Conditions ($\mathscr{C}_5$) and ($\mathscr{C}_6$), we have

$$\sum_{i=1}^{r_p}\tilde{v}_i^2 = \sum_{i=1}^{r_p}\sum_{t=(i-1)b_p+l_p+1}^{ib_p} c^2_{p,t} \le \sum_{i=1}^{r_p} b_p\left(\max_{(i-1)b_p+1\le t\le ib_p} c^2_{p,t}\right)^2 = O(T^{-\alpha}) = o(1).$$

Hence,

$$\sum_{i=1}^{r_p}\sum_{s_1}^{q_1}\cdots\sum_{s_{d+1}}^{q_{d+1}}\mathrm{E}\left(W^2_{p,i,s_1,\cdots,s_{d+1}}\mathbb{I}(\sum_{s_1}^{q_1}\cdots\sum_{s_{d+1}}^{q_{d+1}}W^2_{p,i,s_1,\cdots,s_{d+1}}>\epsilon)\right)$$
$$\leq \max_{1\leq i\leq r_p}\left[\sum_{s_1}^{q_1}\cdots\sum_{s_{d+1}}^{q_{d+1}}\mathrm{E}\left(W^2_{p,i,s_1,\cdots,s_{d+1}}/\tilde{v}_i^2\mathbb{I}\left(\max_{1\leq i\leq r_p}\|\mathbb{W}_{p,i}\|_{\mathrm{F}}^2/\tilde{v}_i^2>\frac{\epsilon}{\max\limits_{1\leq i\leq r_p}\tilde{v}_i^2}\right)\right)\right]\sum_{i=1}^{r_p}\left(\max_{(i-1)b_p+1\leq t\leq ib_p}c_{p,t}\right)^2 b_p$$
$$=O\left(\max_{1\leq i\leq r_p}\left[\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}\left(W^2_{p,i,s_1,\cdots,s_{d+1}}/\tilde{v}_i^2\mathbb{I}\left(\max_{1\leq i\leq r_p}\|\mathbb{W}_{p,i}\|_{\mathrm{F}}^2/\tilde{v}_i^2>\frac{\epsilon}{\max\limits_{1\leq i\leq r_p}\tilde{v}_i^2}\right)\right)\right]\right).$$

It follows from Corollary 3.2.6 that $\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}W^2_{p,i,s_1,\cdots,s_{d+1}}/\tilde{v}_i^2$ is uniformly integrable and $\lim_{L_p\to\infty}\max_{1\leq i\leq r_p}\tilde{v}_i=0$ which implies that the last term above converges to 0 for any $\epsilon>0$. This completes the proof. □

Proof of Lemma 3.2.5. First, note that $T\to\infty$ if and only if $L_p\to\infty$. Further, we have

$$\sum_{i=1}^{L_p}\mathbb{X}_{p,i}=\sum_{i=1}^{r_p}\left(\sum_{t=(i-1)b_p+l_p+1}^{ib_p}\mathbb{X}_{p,t}\right)+\sum_{i=1}^{r_p}\left(\sum_{t=(i-1)b_p+1}^{(i-1)b_p+l_p}\mathbb{X}_{p,t}\right)+\sum_{t=r_pb_p+1}^{L_p}\mathbb{X}_{p,t}. \tag{B.7}$$

Then, using Lemma 3.2.3, we have that

$$\sum_{s_1}^{q_1}\cdots\sum_{s_{d+1}}^{q_{d+1}}\mathrm{E}\left(\sum_{t=r_pb_p+1}^{L_p}X_{p,t,s_1,\cdots,s_{d+1}}\right)^2=O\left(\sum_{t=r_pb_p+1}^{L_p}c^2_{p,t}\right)=O\left(b_p\max_{r_pb_p+1\leq t\leq L_p}c^2_{p,t}\right)=o(1).$$

To study the convergence of $\sum_{i=1}^{r_p}\sum_{t=(i-1)b_p+1}^{(i-1)b_p+l_p}\mathbb{X}_{p,t}$, let $S_2=\{t:t\in\cup_{i=1}^{r_p}[(i-1)b_p+1,(i-1)b_p+l_p]\}$. Then, by combining Lemma 3.2.3 and Conditions ($\mathscr{C}_5$) and ($\mathscr{C}_6$), we have

$$\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}\left[\left(\sum_{i=1}^{r_p}\sum_{t=(i-1)b_p+1}^{(i-1)b_p+l_p}X_{p,t,s_1,\cdots,s_{d+1}}\right)^2\right]=\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\mathrm{E}\left[\left(\sum_{t\in S_2}X_{p,t,s_1,\cdots,s_{d+1}}\right)^2\right]$$
$$=O\left(\sum_{i=1}^{r_p}\sum_{t=(i-1)b_p+1}^{(i-1)b_p+l_p}c^2_{p,t}\right)=O\left(\sum_{i=1}^{r_p}\left(\max_{(i-1)b_p+1\leq t\leq ib_p}c^2_{p,t}\right)^2 l_p\right)=O(T^{-\alpha}l_p/b_p)=o(1).$$

Since the second and third terms of (B.7) converge in probability to 0 as $L_p\to\infty$, it remains to show that $\sum_{i=1}^{r_p}\sum_{t=(i-1)b_p+l_p+1}^{ib_p}\mathbb{X}_{p,t}$ converges in distribution to a random tensor which is distributed as $\mathcal{N}_{q_1\times\cdots\times q_{d+1}}(0,\Sigma_{p,d+1}\otimes\cdots\otimes\Sigma_{p,1})$. To this end, let $\mathbb{V}_{p,i}=\sum_{t=(i-1)b_p+l_p+1}^{ib_p}\mathbb{X}_{p,t}$ and let $\mathscr{F}_i^*$ be the $\sigma-$field generated by $\{U_{ib_p}, U_{ib_p-1,\cdots}\}$, with U_i random variables defined on $(\Omega, \mathscr{F}, P)$ such that $\mathscr{F}_{i-1}^* \subseteq \mathscr{F}_{p,-\infty}^{i-j}$. Then,

$$\sum_{i=1}^{r_p} \sum_{t=(i-1)b_p+l_p+1}^{ib_p} \mathbb{X}_{p,t} = \sum_{i=1}^{r_p}(\mathbb{V}_{p,i} - \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_i^*)) + \sum_{i=1}^{r_p}(\mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_i^*) - \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_{i-1}^*)) + \sum_{i=1}^{r_p} \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_{i-1}^*).$$

By Proposition 3.2.12, we have

$$\sum_{i=1}^{r_p} \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_i^*) \xrightarrow[L_p\to\infty]{P} 0, \quad \sum_{i=1}^{r_p}(\mathbb{V}_{p,i} - \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_{i-1}^*)) \xrightarrow[L_p\to\infty]{P} 0.$$

Since $\mathrm{E}((\mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_i^*) - \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_{i-1}^*))|\mathscr{F}_{i-1}^*) = 0$, then $\mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_i^*) - \mathrm{E}(\mathbb{V}_{p,i}|\mathscr{F}_{i-1}^*)$ is a martingale difference array with respect to $\mathscr{F}_{i-1}^*$. Hence, from Proposition 3.2.12

$$\sum_{i=1}^{r_p} \mathbb{W}_{p,i_{(n)}} \mathbb{W}'_{p,i_{(n)}} \xrightarrow[L_p\to\infty]{P} \Sigma_{p,n}, \ n = 1, \cdots, d+1,$$

and

$$\sum_{i=1}^{r_p} \mathrm{Vec}(\mathbb{W}_{p,i_{(n)}})\mathrm{Vec}(\mathbb{W}_{p,i_{(n)}})' \xrightarrow[L_p\to\infty]{P} \bigotimes_{j=d+1}^{1} \Sigma_{p,j}.$$

Also, by Proposition 3.2.12,

$$\sum_{i=1}^{r_p} \mathrm{E}\left[\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} (W_{p,i,s_1,\cdots,s_{d+1}})^2 \mathbb{I}\left(\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} (W_{p,i,s_1,\cdots,s_{d+1}})^2 > \epsilon\right)\right] \xrightarrow[L_p\to\infty]{P} 0,$$

for any $\epsilon > 0$. Hence, by the martingale difference sequence central limit theorem,

$$\sum_{i=1}^{L_p} \mathbb{X}_{p,i} \xrightarrow[L_p\to\infty]{d} \mathscr{N}_{q_1\times\cdots\times q_{d+1}}\left(0, \bigotimes_{j=d+1}^{1} \Sigma_{p,j}\right),$$

which completes the proof. □

Proof of Proposition 3.2.13. For the first equation above, we have

$$\begin{aligned}\sum_{i=1}^{r_{\min}} \|\zeta_{1,a,b,i,n}\|_1 \le & \sum_{i=1}^{r_{\min}} \|(\mathbb{V}_{a,i_{(n)}} - \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_i^*) + \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_{i-1}^*))\|_2 \\ & \times \|(\mathbb{V}_{b,i_{(n)}} + \mathrm{E}(\mathbb{V}_{b,i_{(n)}}|\mathscr{F}_i^*) - \mathrm{E}(\mathbb{V}_{b,i_{(n)}}|\mathscr{F}_{i-1}^*))'\|_2 \\ \le & \sum_{i=1}^{r_{\min}} \|(\mathbb{V}_{a,i_{(n)}} - \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_i^*) + \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_{i-1}^*))\|_2 \\ & \times (\|(\mathbb{V}_{b,i_{(n)}}\|_2 + \|\mathrm{E}(\mathbb{V}_{b,i_{(n)}}|\mathscr{F}_i^*)\|_2 + \|\mathrm{E}(\mathbb{V}_{b,i_{(n)}}|\mathscr{F}_{i-1}^*))'\|_2).\end{aligned}$$

This gives

$$\begin{aligned}\sum_{a=1}^{m+1}\sum_{b=1}^{m+1}\sum_{i=1}^{r_{\min}} \|\zeta_{1,a,b,i,n}\|_1 \le & \ 3\sum_{i=1}^{r_{\min}} \|(\mathbb{V}_{a,i_{(n)}} - \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_i^*) + \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_{i-1}^*))\|_2\|\mathbb{V}_{b,i_{(n)}}\|_2 \\ \le & \ 3\sum_{i=1}^{r_{\min}} (\|(\mathbb{V}_{a,i_{(n)}} - \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_i^*)\|_2 + \|\mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_{i-1}^*))\|_2)\|\mathbb{V}_{b,i_{(n)}}\|_2.\end{aligned}$$

It can be shown that, for some $\eta \in (0,1)$

$$\|(\mathbb{V}_{a,i_{(n)}} - \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_i^*)\|_2 = O\left(\sum_{i=1}^{r_p}\left(\sum_{t=(i-1)b_p+l_p+1}^{ib_p} c_{p,t}^2\psi(l_p+1)^{2\eta}\right)^{1/2}\right), \tag{B.8}$$

$$\|\mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_{i-1}^*))\|_2 = O\left(\sum_{i=1}^{r_p}\left(\sum_{t=(i-1)b_p+l_p+1}^{ib_p} c_{p,t}^2\psi(l_p)^{2\eta}\right)^{1/2}\right), \tag{B.9}$$

and

$$\|\mathbb{V}_{b,i_{(n)}}\|_2 = O\left(\left(\sum_{t=(i-1)b_p+l_p+1}^{ib_p} c_{p,t}^2\right)^{1/2}\right). \tag{B.10}$$

Then, combining the relations (B.8), (B.9) and (B.10), we have as a result

$$\sum_{i=1}^{r_{\min}} \|\zeta_{1,a,b,i,n}\|_1 = O\left(\sum_{i=1}^{r_{\min}}\left(\sum_{t=(i-1)b_a+l_a+1}^{ib_a} c_{a,t}^2\psi(l_{p,a})^{2\eta}\right)^{1/2}\left(\sum_{t=(i-1)b_b+l_b+1}^{ib_b} c_{b,t}^2\right)^{1/2}\right).$$

Thus, we have

$$\sum_{i=1}^{r_{\min}} \|\zeta_{1,a,b,i,n}\|_1 = O\left(\sum_{i=1}^{r_{\min}}(b_aM_{a,i})^{1/2}\psi(l_a)^{\eta}(b_bM_{b,i})^{1/2}\right),$$

where $M_{j,i} = (\max_{(i-1)b_j+l_j+1j}\, c_{j,t}^2)$, $j = a, b$ for each $a, b = 1, \cdots, m+1$.

Therefore, using $(\mathscr{C}_5)$ and $(\mathscr{C}_6)$, we have

$\sum_{i=1}^{r_{\min}} \|\zeta_{1,a,b,i,n}\|_1 = O(T^{-\alpha}\psi(l_a)^{\eta}) = o(1)$, which proves the first statement in (3.6).

To prove the second statement in (3.6), we have

$$\sum_{i=1}^{r_{\min}} \|\zeta_{2,a,b,i,n}\|_1 \leq \sum_{i=1}^{r_{\min}} (\|\mathbb{V}_{a,i_{(n)}} - \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_i^*) + \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_{i-1}^*))\|_2 \|V_{b,i_n}\|_2$$
$$+ \|\mathbb{V}_{a,i_{(n)}}\|_2 \|\mathbb{V}_{b,i_{(n)}} - \mathrm{E}(\mathbb{V}_{b,i_{(n)}}|\mathscr{F}_i^*) + \mathrm{E}(\mathbb{V}_{b,i_{(n)}}|\mathscr{F}_{i-1}^*))\|_2).$$

Again, by combining the relations (B.8), (B.9) and (B.10), we get

$\sum_{i=1}^{r_{\min}} \|\zeta_{2,a,b,i,n}\|_1 = o(1) + o(1) = o(1)$, this proves the second statement in (3.6). □

Proof of Proposition 3.2.14. We have

$$\left\|\sum_{i=1}^{r_{\min}} [\mathbb{V}_{i_{(n)}}\mathbb{V}'_{i_{(n)}} - \mathbb{W}_{i_{(n)}}\mathbb{W}'_{i_{(n)}}]\right\|_1 \leq \sum_{i=1}^{r_{\min}} \left\|[\mathbb{V}_{i_{(n)}}\mathbb{V}'_{i_{(n)}} - \mathbb{W}_{i_{(n)}}\mathbb{W}'_{i_{(n)}}]\right\|_1$$
$$\leq \sum_{a=1}^{m+1}\sum_{b=1}^{m+1}\sum_{i=1}^{r_{\min}} \left\|\mathbb{V}_{a,i_{(n)}}\mathbb{V}'_{b,i_{(n)}} - \mathbb{W}_{a,i_{(n)}}\mathbb{W}'_{b,i_{(n)}}\right\|_1 .$$

Using the same techniques as in Proposition 3.2.13, we have that for any $a, b = 1, \cdots, m+1$,

$$\|\mathbb{V}_{a,i_{(n)}}\mathbb{V}'_{b,i_{(n)}} - \mathbb{W}_{a,i_{(n)}}\mathbb{W}'_{b,i_{(n)}}\|_1 \leq \|(\mathbb{V}_{a,i_{(n)}} - \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_i^*) + \mathrm{E}(\mathbb{V}_{a,i_{(n)}}|\mathscr{F}_{i-1}^*))$$
$$\times (\mathbb{V}_{b,i_{(n)}} + \mathrm{E}(\mathbb{V}_{b,i_{(n)}}|\mathscr{F}_i^*) - \mathrm{E}(V_{-,i_{(n)}}|\mathscr{F}_{i-1}^*))' - \zeta_{2,a,b,i,n}\|_1.$$

So, we have

$$\|\mathbb{V}_{a,i_{(n)}}\mathbb{V}'_{b,i_{(n)}} - \mathbb{W}_{a,i_{(n)}}\mathbb{W}'_{b,i_{(n)}}\|_1 \leq \|\zeta_{1,a,b,i,n}\|_1 + \zeta_{2,a,b,i,n}\|_1.$$

Then, the first statement follows from Proposition 3.2.13. To prove the relation (3.8), note that $\|\mathbb{X}\|_{\mathrm{F}} = \sqrt{\mathrm{Vec}(\mathbb{X})'\mathrm{Vec}(\mathbb{X})}$ for any random tensor $\mathbb{X}$. Therefore, the proof of (3.8) follows from similar steps to that of (3.7). □

Proof of Proposition 3.2.15. First, from Proposition 3.2.12,

$$\sum_{i=1}^{r_a}\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left[(W_{a,i,s_1,\cdots,s_{d+1}})^2 \mathbb{I}\left(\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}} W^2_{a,i,s_1,\cdots,s_{d+1}} > \epsilon\right)\right] \xrightarrow[L_{\min}\to\infty]{P} 0,$$

for each $a = 1, \cdots, m+1$, and thus, their summation also converges to 0, this proves (3.9). The proof of the relation in (3.10) follows from a more generalized version of the proof of Proposition 3.2.11 with $\sum_{i=1}^{r_p} \mathbb{V}_{p,i_{(n)}} \mathbb{V}'_{p,i_{(n)}}$ and $\sum_{i=1}^{r_p} \mathbb{W}_{p,i_{(n)}} \mathbb{W}'_{p,i_{(n)}}$ extended to $\sum_{i=1}^{r_{\min}} \mathbb{V}_{i_{(n)}} \mathbb{V}'_{i_{(n)}}$ and $\sum_{i=1}^{r_{\min}} \mathbb{W}_{i_{(n)}} \mathbb{W}'_{i_{(n)}}$, $n = 1, \cdots, d+1$, respectively. Indeed,

$$\sum_{i=1}^{r_{\min}} \mathbb{W}_{i_{(n)}} \mathbb{W}'_{i_{(n)}} = \sum_{i=1}^{r_{\min}} \mathbb{V}_{i_{(n)}} \mathbb{V}'_{i_{(n)}} - \sum_{i=1}^{r_{\min}} \left[\mathbb{V}_{i_{(n)}} \mathbb{V}'_{i_{(n)}} - \mathbb{W}_{i_{(n)}} \mathbb{W}'_{i_{(n)}} \right].$$

Hence, combining ($\mathscr{C}_6$), Proposition 3.2.14 and Slutsky's theorem, we have

$$\sum_{i=1}^{r_{\min}} \mathbb{W}_{i_{(n)}} \mathbb{W}'_{i_{(n)}} \xrightarrow[L_{\min}]{P} \Lambda_n, \qquad \text{for } n = 1, \cdots, d+1.$$

Moreover, note that

$\sum_{i=1}^{r_{\min}} \mathrm{Vec}(\mathbb{W}_i)\mathrm{Vec}(\mathbb{W}_i)' = \sum_{i=1}^{r_{\min}} \mathrm{Vec}(\mathbb{V}_i)\mathrm{Vec}(\mathbb{V}_i)' - \sum_{i=1}^{r_{\min}} [\mathrm{Vec}(\mathbb{V}_i)\mathrm{Vec}(\mathbb{V}_i)' - \mathrm{Vec}(\mathbb{W}_i)\mathrm{Vec}(\mathbb{W}_i)']$.
Then, the proof of (3.8) follows from Proposition 3.2.14 along with Condition ($\mathscr{C}_6$). □

Proof of Lemma 3.2.6. We have

$$T^{-1/2}\mathbb{U} \times_{(d+1)} Z^{0'} = \zeta + \sum_{i=1}^{r_{\min}} \mathbb{W}_i$$
$$+ \left(\sum_{i=r_{\min}+1}^{r_1} \left(\sum_{t=(i-1)b_1+1}^{ib_1} \mathbb{X}_{1,t} \right) \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \sum_{i=r_{\min}+1}^{r_{m+1}} \left(\sum_{t=(i-1)b_{m+1}+1}^{ib_{m+1}} \mathbb{X}_{m+1,t} \right) \right). \tag{B.11}$$

Then, it follows from Lemma 3.2.5 that

$$\zeta \xrightarrow[L_{\min\to\infty}]{p} 0. \tag{B.12}$$

Moreover, for each $j = 1, \cdots, m+1$, we have

$$\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left[\left(\sum_{i=r_{\min}}^{r_j} \sum_{t=(i-1)b_j+1}^{ib_j} X_{j,t,s_1,\cdots,s_{d+1}} \right)^2 \right] = \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left[\left(\sum_{t \in S_j^*} X_{j,t,s_1,\cdots,s_{d+1}} \right)^2 \right],$$

where $S_j^* = \{t : t \in \bigcup_{i=r_{\min+1}}^{r_j} [(i-1)b_j+1, ib_j]\}$. Then, by using Lemma 3.2.3, we have

$$\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left[\left(\sum_{i=r_{\min}}^{r_j} \sum_{t=(i-1)b_j+1}^{ib_j} X_{j,t,s_1,\cdots,s_{d+1}}\right)^2\right] = O\left(\sum_{i=r_{\min}+1}^{r_j} \sum_{t=(i-1)b_j+1}^{ib_j} c_{j,t}^2\right)$$
$$= O\left(\sum_{i=r_{\min}+1}^{r_j} \left(\max_{(i-1)b_j+1 \le t \le ib_j} c_{j,t}\right)^2 b_j\right) = o(1).$$

This implies that

$$\sum_{i=r_{\min}+1}^{r_1} \left(\sum_{t=(i-1)b_1+1}^{ib_1} X_{1,t}\right) \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \sum_{i=r_{\min}+1}^{r_{m+1}} \left(\sum_{t=(i-1)b_{m+1}+1}^{ib_{m+1}} X_{m+1,t}\right) \xrightarrow[L_{\min}\to\infty]{P} 0. \tag{B.13}$$

In addition, by Proposition 3.2.15, we have

$$\sum_{i=1}^{r_{\min}} \mathbb{W}_{i_{(n)}} \mathbb{W}_{i_{(n)}}' \xrightarrow[L_{\min}\to\infty]{P} \Lambda_n, n = 1, \cdots, d+1,$$

$$\sum_{i=1}^{r_{\min}} \mathrm{Vec}(\mathbb{W}_i)\mathrm{Vec}(\mathbb{W}_i)' \xrightarrow[T\to\infty]{P} \Lambda_1 \otimes \cdots \otimes \Lambda_{d+1},$$

and

$$\sum_{a+1}^{m+1} \sum_{i=1}^{r_a} \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \mathrm{E}\left[(W_{a,i,s_1,\cdots,s_{d+1}})^2 \mathbb{I}\left(\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} W_{a,i,s_1,\cdots,s_{d+1}}^2 > \epsilon\right)\right] \xrightarrow[L_{\min}\to\infty]{} 0,$$

for all $\epsilon > 0$. Hence, by the martingale difference sequence central limit theorem,

$$\sum_{i=1}^{r_{\min}} \mathbb{W}_i \xrightarrow[T\to\infty]{d} \mathcal{N}_{q_1\times\cdots\times q_d\times(m+1)q_{d+1}}(0, \Lambda_{d+1} \otimes \cdots \otimes \Lambda_1).$$

Therefore, together with (B.11), (B.12), (B.13) along with Slutsky's theorem, we establish the stated result. □

Proof of Proposition 3.2.16. From (3.1) and Proposition 3.1.1, we have

$$\hat{\boldsymbol{\delta}}(\tau) = \mathbb{Y} \times_{d+1} (Z'Z)^{-1}ZZ' = (\boldsymbol{\delta} \times_{d+1} Z + \mathbb{U}) \times_{d+1} (Z'Z)^{-1}Z'.$$

Using tensor mode multiplication properties, this becomes

$$\hat{\boldsymbol{\delta}}(\tau) = \boldsymbol{\delta} \times_{d+1} Z \times_{d+1} (Z'Z)^{-1}Z' + \mathbb{U} \times_{d+1} (Z'Z)^{-1}Z'$$
$$= \boldsymbol{\delta} \times_{d+1} ((Z'Z)^{-1}Z'Z) + (\mathbb{U} \times_{d+1} Z') \times_{d+1} (Z'Z)^{-1}.$$

Thus,

$$\hat{\boldsymbol{\delta}}(\tau) = \boldsymbol{\delta} + (\mathbb{U} \times_{d+1} Z') \times_{d+1} (Z'Z)^{-1}. \tag{B.14}$$

From (B.14), we get

$$\sqrt{T}(\hat{\boldsymbol{\delta}}(\tau) - \boldsymbol{\delta}) = \sqrt{T}(\mathbb{U} \times_{d+1} Z') \times_{d+1} (Z'Z)^{-1} = T^{-1/2}(\mathbb{U} \times_{d+1} Z') \times_{d+1} (T(Z'Z)^{-1}).$$

Then,

$$\sqrt{T}(\hat{\boldsymbol{\delta}}(\tau) - \boldsymbol{\delta}) = (T^{-1/2}\mathbb{U} \times_{d+1} Z') \times_{d+1} (T^{-1}Z'Z)^{-1}.$$

Therefore, by using Lemma 3.2.6 and Slutsky's theorem along with the fact that $T^{-1}Z'Z \xrightarrow[T\to\infty]{p} \Gamma$, we have

$$\sqrt{T}(\hat{\boldsymbol{\delta}}(\tau) - \boldsymbol{\delta}) \xrightarrow[T\to\infty]{d} \epsilon_1^{*0} \times_{d+1} \Gamma^{-1},$$

where $\epsilon_1^{*0} \sim \mathcal{N}_{q_1,\cdots,q_d,(m+1)q_{d+1}}(0, \Lambda_{d+1} \otimes \cdots \otimes \Lambda_1)$. Hence, using Proposition B.1.1,

$$\epsilon_1^* = \epsilon_1^{*0} \times_{d+1} \Gamma^{-1} \sim \mathcal{N}_{q_1,\cdots,q_d,(m+1)q_{d+1}}(0, \Gamma^{-1}\Lambda_{d+1}\Gamma^{-1} \otimes \Lambda_d \otimes \cdots \otimes \Lambda_1). \tag{B.15}$$

This completes the proof. □

Proof of Proposition 3.2.18. Note that

$$\begin{aligned}
\mathrm{Vec}(\epsilon_{1,T}^* \boxplus_{d+1} \epsilon_{2,T}^* \boxplus_{d+1} \epsilon_{3,T}^*) &= \left(\mathrm{Vec}(\epsilon_{1,T}^*)', \quad \mathrm{Vec}(\epsilon_{2,T}^*)', \quad \mathrm{Vec}(\epsilon_{3,T}^*)'\right)' \\
&= \begin{pmatrix} \mathrm{Vec}(\epsilon_{1,T}^*) \\ (I - \mathbb{J}(T))\mathrm{Vec}(\epsilon_{1,T}^*) - \mathbb{G}(T)\mathrm{Vec}(r_0) \\ \mathbb{J}(T)\mathrm{Vec}(\epsilon_{3,T}^*) + \mathbb{G}(T)\mathrm{Vec}(r_0) \end{pmatrix} = \begin{pmatrix} I \\ I - \mathbb{J}(T) \\ \mathbb{J}(T) \end{pmatrix} \mathrm{Vec}(\epsilon_{1,T}^*) + \begin{pmatrix} 0 \\ -\mathbb{G}(T)\mathrm{Vec}(r_0) \\ \mathbb{G}(T)\mathrm{Vec}(r_0) \end{pmatrix}.
\end{aligned}$$

Then, using Proposition 3.2.16 along with Slutsky's Theorem, we have

$$\mathrm{Vec}(\epsilon_{1,T}^*(\tau, m) \boxplus_{d+1} \epsilon_{2,T}^*(\tau, m) \boxplus_{d+1} \epsilon_{3,T}^*(\tau, m)) \xrightarrow[T\to\infty]{d} \begin{pmatrix} \mathrm{Vec}(\epsilon_1^*) \\ \mathrm{Vec}(\epsilon_2^*) \\ \mathrm{Vec}(\epsilon_3^*) \end{pmatrix}$$

with

$$\Big((\mathrm{Vec}(\epsilon_1^*))', (\mathrm{Vec}(\epsilon_2^*))', (\mathrm{Vec}(\epsilon_3^*))'\Big)' = \Big(I, I-\Omega', \Omega'\Big)' (\mathrm{Vec}(\epsilon_1^*))' + \Big(0, -(\mathbb{G}^*\mathrm{Vec}(r_0))', (\mathbb{G}^*\mathrm{Vec}(r_0))'\Big)'.$$

Then, by using Proposition B.1.2, we have

$$\Big((\mathrm{Vec}(\epsilon_1^*))', (\mathrm{Vec}(\epsilon_2^*))', (\mathrm{Vec}(\epsilon_3^*))'\Big)' \sim \mathscr{N}_{3(m+1)q_1\cdots q_{d+1}}\left(\mathrm{Vec}(\mu), \Sigma\right),$$

where $\mathrm{Vec}(\mu) = \begin{pmatrix} 0 \\ \mathrm{Vec}(\mu_1^*) \\ \mathrm{Vec}(-\mu_1^*) \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* & \Sigma_{13}^* \\ \Sigma_{21}^* & \Sigma_{22}^* & \Sigma_{23}^* \\ \Sigma_{31}^* & \Sigma_{32}^* & \Sigma_{33}^* \end{pmatrix}$, this proves the first statement. The last statement follows from the first statement along with Proposition B.3.7 and Proposition B.3.8. This completes the proof. □

Proof of Proposition 3.3.1. Note that, for $j = 1, \cdots, d$,

$$\begin{aligned}\Lambda_{Xj}A_j &= \mathbb{G}_j R_j \Lambda_j R_j' \mathbb{G}_j' R_j' (R_j \Lambda_j R_j')^{-1} R_j = R_j'(R_jR_j')^{-1} R_j \Lambda_j R_j' (R_jR_j')^{-1} R_j R_j' (R_j\Lambda_j R_j')^{-1} R_j \\ &= R_j'(R_jR_j')^{-1}R_j,\end{aligned}$$

and

$$(R_j'(R_jR_j')^{-1}R_j)^2 = R_j'(R_jR_j')^{-1}R_jR_j'(R_jR_j')^{-1}R_j = R_j'(R_jR_j')^{-1}R_j.$$

Therefore, $\Lambda_{Xj}A_j$ are idempotent for $j = 1, \cdots, d$. Moreover,

$$A_j\Lambda_{Xj}A_j = R_j'(R_j\Lambda_jR_j')^{-1}R_jR_j'(R_jR_j')^{-1}R_j = R_j'(R_j\Lambda_jR_j')^{-1}R_j = A_j,$$

and

$$\begin{aligned}\Lambda_{Xj}A_j\Lambda_{Xj} &= R_j'(R_jR_j')^{-1}R_jR_j'(R_jR_j')^{-1}R_jR_j'(R_jR_j')^{-1}R_j\Lambda_jR_j'(R_jR_j')^{-1}R_j \\ &= R_j'(R_jR_j')^{-1}R_j\Lambda_jR_j'(R_jR_j')^{-1}R_j = \Lambda_{Xj}.\end{aligned}$$

Also, for $j = 1, \cdots, d$,

$$\mu^{**} \times_j \Lambda_{Xj} A_j = \left(\boldsymbol{r}_0 \bigtimes_{\substack{i=1 \\ i \neq j}}^{d} \mathbb{G}_i \times_{d+1} \mathbb{G}^*_{d+1} \right) \times_j (R'_j (R_j R'_j)^{-1} R_j)$$

$$= \boldsymbol{r}_0 \bigtimes_{\substack{i=1 \\ i \neq j}}^{d} \mathbb{G}_i \times_j (R'_j (R_j R'_j)^{-1} R_j) \mathbb{G}_j \times_{d+1} \mathbb{G}^*_{d+1}$$

$$= \boldsymbol{r}_0 \bigtimes_{\substack{i=1 \\ i \neq j}}^{d} \mathbb{G}_i \times_j R'_j (R_j R'_j)^{-1} R_j R'_j (R_j R'_j)^{-1} \times_{d+1} \mathbb{G}^*_{d+1}$$

$$= \boldsymbol{r}_0 \bigtimes_{\substack{i=1 \\ i \neq j}}^{d} \mathbb{G}_i \times_j R'_j (R_j R'_j)^{-1} \times_{d+1} \mathbb{G}^*_{d+1} = \boldsymbol{r}_0 \bigtimes_{\substack{i=1 \\ i \neq j}}^{d} \mathbb{G}_i \times_j \mathbb{G}_j \times_{d+1} \mathbb{G}^*_{d+1} = \mu^{**},$$

and $\mathbb{W}_j = A_j^{1/2} \mathbb{W}^*_j A_j^{1/2}$ is symmetric and non-negative definite matrix since A_j is non-negative with rank l_j. Similarly,

$$\Lambda_{Xd+1} A_{d+1} = (\mathbb{G}^*_{d+1} R_{d+1} \Gamma^{-1} \Lambda_{d+1} \Gamma^{-1} R'_{d+1} {\mathbb{G}^*_{d+1}}') R'_{d+1} (R_{d+1} \Gamma^{-1} \Lambda_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1}$$

$$= (\Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1} \Gamma^{-1} \Lambda_{d+1} \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1} \Gamma^{-1})$$

$$\times R'_{d+1} (R_{d+1} \Gamma^{-1} \Lambda_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1},$$

this gives

$$\Lambda_{Xd+1} A_{d+1} = \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1}$$

and

$$\Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1} = \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1} \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1}$$

$$= \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1}.$$

Therefore, $\Lambda_{Xd+1} A_{d+1}$ is idempotent. Moreover,

$$A_{d+1} \Lambda_{Xd+1} A_{d+1} = R'_{d+1} (R_{d+1} \Gamma^{-1} \Lambda_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1} \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1} = A_{d+1},$$

and

$$\Lambda_{Xd+1} A_{d+1} \Lambda_{Xd+1} = \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1} \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1}$$
$$\Gamma^{-1} \Lambda_{d+1} \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1} \Gamma^{-1}.$$

This gives

$$\Lambda_{Xd+1} A_{d+1} \Lambda_{Xd+1} = \Lambda_{Xd+1}.$$

Also,

$$\mu^{**} \times_{d+1} \Lambda_{Xd+1} A_{d+1} = \left(\boldsymbol{r}_0 \bigtimes_{j=1}^{d} \mathbb{G}_j \times_{d+1} \mathbb{G}^*_{d+1} \right) \times_{d+1} \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1}$$
$$= \boldsymbol{r}_0 \bigtimes_{j=1}^{d} \mathbb{G}_j \times_{d+1} \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1} \mathbb{G}^*_{d+1}$$
$$= \boldsymbol{r}_0 \bigtimes_{j=1}^{d} \mathbb{G}_j \times_{d+1} \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} R_{d+1} \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1}$$
$$= \boldsymbol{r}_0 \bigtimes_{j=1}^{d} \mathbb{G}_j \times_{d+1} \Gamma^{-1} R'_{d+1} (R_{d+1} \Gamma^{-1} R'_{d+1})^{-1} = \boldsymbol{r}_0 \bigtimes_{j=1}^{d} \mathbb{G}_j \times_{d+1} \mathbb{G}^*_{d+1}.$$

This gives $\mu^{**} \times_{d+1} \Lambda_{Xd+1} A_{d+1} = \mu^{**}$, and since A_{d+1} is non-negative with rank l_{d+1}, $\mathbb{W}_{d+1} = A^{1/2}_{d+1} \mathbb{W}^*_{d+1} A^{1/2}_{d+1}$ is a symmetric and non-negative definite matrix . Hence, the assumptions of Theorem 2.1.1 hold and it follows that

$$\mathrm{E}\left[h\left(\mathrm{trace}\left(\epsilon^{*\prime}_{31_{(d)}} \epsilon^*_{31_{(d)}} \right) \right) \epsilon^*_3 (\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j \right] = \mathrm{E}\left[h\left(\chi^2_{l+2} \left(\mathrm{trace}\left(\mu^{*\prime}_{1_{(d)}} \mu^*_{1_{(d)}} \right) \right) \right) \right] \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j \right).$$

□

Proof of Lemma 3.3.1. By Slutsky's theorem, we have

$$\sqrt{T} \left(\hat{\boldsymbol{\vartheta}}(h, \hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}) - \boldsymbol{\delta} \right) = \sqrt{T} \left(\tilde{\boldsymbol{\delta}}(\hat{\tau}) - \boldsymbol{\delta} \right) + (h(\psi)) \sqrt{T} \left(\hat{\boldsymbol{\delta}}(\hat{\tau}) - \tilde{\boldsymbol{\delta}}(\hat{\tau}) \right) \xrightarrow[T \to \infty]{d} \boldsymbol{\vartheta}^*(h),$$

where $\boldsymbol{\vartheta}^*(h) = \epsilon^*_2 + h\left(\mathrm{trace}\left(\epsilon^{*\prime}_{31_{(d)}} \epsilon^*_{31_{(d)}} \right) \right) \epsilon^*_3$. Then,

$$\mathrm{ADR}^1 \left(\hat{\boldsymbol{\vartheta}}(h, \hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}), \boldsymbol{\delta}, \mathbb{W} \right) = \mathrm{E}\left[\mathrm{trace}\left(\left(\boldsymbol{\vartheta}^*(h) (\bigtimes_{j=1}^{d+1})_j \mathbb{W}^{1/2}_j \right)'_{(d)} \left(\boldsymbol{\vartheta}^*(h) (\bigtimes_{j=1}^{d+1})_j \mathbb{W}^{1/2}_j \right)_{(d)} \right) \right]$$
$$= \mathrm{E}\left[\mathrm{trace}\left(\left(\epsilon^*_{21} + h\left(\mathrm{trace}\left(\epsilon^{*\prime}_{31_{(d)}} \epsilon^*_{31_{(d)}} \right) \right) \epsilon^*_{32} \right)'_{(d)} \left(\epsilon^*_{21} + h\left(\mathrm{trace}\left(\epsilon^{*\prime}_{31_{(d)}} \epsilon^*_{31_{(d)}} \right) \right) \epsilon^*_{32} \right)_{(d)} \right) \right].$$

Then,

$$\begin{aligned}
\mathrm{ADR}^1\left(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\delta},\mathbb{W}\right) &= \mathrm{E}\left[\mathrm{trace}\left(\epsilon^{*\prime}_{21_{(d)}}\epsilon^{*}_{21_{(d)}}\right)\right]\\
&+ 2\mathrm{E}\left[\mathrm{trace}\left(\epsilon^{*\prime}_{21_{(d)}}\left(h\left(\mathrm{trace}\left(\epsilon^{*\prime}_{31_{(d)}}\epsilon^{*}_{31_{(d)}}\right)\right)\epsilon^{*}_{32}\right)_{(d)}\right)\right]\\
&+ \mathrm{E}\left[\mathrm{trace}\left(\left(h\left(\mathrm{trace}\left(\epsilon^{*\prime}_{31_{(d)}}\epsilon^{*}_{31_{(d)}}\right)\right)\epsilon^{*}_{32}\right)'_{(d)}\left(h\left(\mathrm{trace}\left(\epsilon^{*\prime}_{31_{(d)}}\epsilon^{*}_{31_{(d)}}\right)\right)\epsilon^{*}_{32}\right)_{(d)}\right)\right],
\end{aligned}$$

and then

$$\begin{aligned}
&\mathrm{ADR}^1\left(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\delta},\mathbb{W}\right) = \mathrm{E}\left[\mathrm{trace}\left(\epsilon^{*\prime}_{21_{(d)}}\epsilon^{*}_{21_{(d)}}\right)\right]\\
&+ 2\mathrm{E}\left[h\left(\mathrm{trace}\left(\epsilon^{*\prime}_{31_{(d)}}\epsilon^{*}_{31_{(d)}}\right)\right)\mathrm{trace}\left(\epsilon^{*\prime}_{21_{(d)}}\epsilon^{*}_{32_{(d)}}\right)\right] + \mathrm{E}\left[h^2\left(\epsilon^{*\prime}_{31_{(d)}}\epsilon^{*}_{31_{(d)}}\right)\mathrm{trace}\left(\epsilon^{*\prime}_{32_{(d)}}\epsilon^{*}_{32_{(d)}}\right)\right]\\
&= \mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}),\boldsymbol{\delta},\mathbb{W}\right) + 2\mathrm{E}\left[h\left(\mathrm{trace}\left(\epsilon^{*\prime}_{31_{(d)}}\epsilon^{*}_{31_{(d)}}\right)\right)\mathrm{trace}\left(\epsilon^{*\prime}_{21_{(d)}}\epsilon^{*}_{32_{(d)}}\right)\right]\\
&+ \mathrm{E}\left[h^2\left(\epsilon^{*\prime}_{31_{(d)}}\epsilon^{*}_{31_{(d)}}\right)\mathrm{trace}\left(\epsilon^{*\prime}_{32_{(d)}}\epsilon^{*}_{32_{(d)}}\right)\right].
\end{aligned}$$

Hence, by Proposition 3.3.2 and Proposition 3.3.3, we have

$$\begin{aligned}
&\mathrm{ADR}^1\left(\hat{\boldsymbol{\vartheta}}(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}),\boldsymbol{\delta},\mathbb{W}\right) = \mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}),\boldsymbol{\delta},\mathbb{W}\right)\\
&- 2\mathrm{E}\left[h\left(\chi^2_{l+2}(\Delta)\right)\mathrm{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B^*_j \mathbb{W}_j\right)_{(d)}\right)\right]\\
&+ 2\mathrm{E}\left[h\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j B^*_j) - 2\mathrm{E}\left[h\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j \Upsilon^*_j)\\
&+ 2\mathrm{E}\left[h\left(\chi^2_{l+4}(\Delta)\right)\right]\mathrm{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B^*_j \mathbb{W}_j\right)_{(d)}\right) + \mathrm{E}\left[h^2(\chi^2_{l+2}(\Delta))\right]\prod_{j=1}^{d+1}\mathrm{trace}\left(\mathbb{W}_j \Upsilon^*_j\right)\\
&- 2\mathrm{E}\left[h\left(\chi^2_{l+4}(\Delta)\right)\right]\mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}^{1/2}_j\right)'_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}^{1/2}_j\right)_{(d)}\right)\\
&+ \mathrm{E}\left[h^2(\chi^2_{l+4}(\Delta))\right]\mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}^{1/2}_j\right)'_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}^{1/2}_j\right)_{(d)}\right).
\end{aligned}$$

This completes the proof. □

Proof of Lemma 3.3.2. Since $\epsilon_1^* \sim \mathcal{N}_{q_1\times\cdots q_d\times(m_0+1)q_{d+1}}\left(0, \Sigma_{11}^*\right)$, then,

$$\mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}), \boldsymbol{\delta}, \mathbb{W}\right) = \mathrm{E}\left[\mathrm{trace}\left(\left(\epsilon_1^*(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}' \left(\epsilon_1^*(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right)\right]$$
$$= \mathrm{E}\left[(\mathrm{Vec}(\epsilon_1^*))' \bigotimes_{j=d+1}^{1} \mathbb{W}_j \left(\mathrm{Vec}(\epsilon_1^*)\right)\right] = \mathrm{trace}\left(\bigotimes_{j=d+1}^{1} \mathbb{W}_j \Sigma_{11}^*\right)$$
$$= \mathrm{trace}\left(\left(\bigotimes_{j=d+1}^{1} \mathbb{W}_j\right)\left(\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1} \otimes \bigotimes_{j=d}^{1} \Lambda_j\right)\right)$$
$$= \mathrm{trace}\left(W_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\right)\prod_{j=1}^{d} \mathrm{trace}\left(\mathbb{W}_j\Lambda_j\right) = \prod_{j=1}^{d+1} \mathrm{trace}\left(\mathbb{W}_j C_j^*\right).$$

Similarly, since $\epsilon_2^* \sim \mathcal{N}_{q_1\times\cdots q_d\times(m_0+1)q_{d+1}}\left(\mu^{**}, \Sigma_{22}^*\right)$, then,

$$\mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}), \boldsymbol{\delta}, \mathbb{W}\right) = \mathrm{E}\left[\mathrm{trace}\left(\left(\epsilon_2^*(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}' \left(\epsilon_2^*(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right)\right]$$
$$= \mathrm{E}\left[(\mathrm{Vec}(\epsilon_2^*))' \bigotimes_{j=d+1}^{1} \mathbb{W}_j \left(\mathrm{Vec}(\epsilon_2^*)\right)\right]$$
$$= \mathrm{trace}\left(\bigotimes_{j=d+1}^{1} \mathbb{W}_j \Sigma_{22}^*\right) + \left(\mathrm{Vec}\left(\mu^{**}\right)\right)' \bigotimes_{j=d+1}^{1} \mathbb{W}_j \left(\mathrm{Vec}\left(\mu^{**}\right)\right)$$
$$= \mathrm{trace}\left(\bigotimes_{j=d+1}^{1} \mathbb{W}_j \left(\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1} \otimes \bigotimes_{j=d}^{1} \Lambda_j\right)\right)$$
$$- \mathrm{trace}\left(\bigotimes_{j=d+1}^{1} \mathbb{W}_j \left(\Omega_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1} \otimes \bigotimes_{j=d}^{1} \Omega_j\Lambda_j\right)\right)$$
$$- \mathrm{trace}\left(\bigotimes_{j=d+1}^{1} \mathbb{W}_j \left(\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}{}' \otimes \bigotimes_{j=d}^{1} \Lambda_j\Omega_j'\right)\right)$$
$$+ \mathrm{trace}\left(\bigotimes_{j=d+1}^{1} \mathbb{W}_j \left(\Omega_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}{}' \otimes \bigotimes_{j=d}^{1} \Omega_j\Lambda_j\Omega_j'\right)\right)$$
$$+ \left(\mathrm{Vec}\left(\mu^{**}\right)\right)' \bigotimes_{j=d+1}^{1} \mathbb{W}_j \left(\mathrm{Vec}\left(\mu^{**}\right)\right).$$

Thus, we have,

$$\begin{aligned}
\mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}), \boldsymbol{\delta}, \mathbb{W}\right) &= \operatorname{trace}\left(W_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1} \otimes \bigotimes_{j=d}^{1} \mathbb{W}_j\Lambda_j\right)\\
&- 2\operatorname{trace}\left(W_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}{}' \otimes \bigotimes_{j=d}^{1} \mathbb{W}_j\Lambda_j\Omega_j'\right)\\
&+ \operatorname{trace}\left(W_{d+1}\Omega_{d+1}\Gamma^{-1}\Lambda_{d+1}\Gamma^{-1}\Omega_{d+1}{}' \otimes \bigotimes_{j=d}^{1} \mathbb{W}_j\Omega_j\Lambda_j\Omega_j'\right)\\
&+ \operatorname{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\right)_{(d)}' \left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\right)_{(d)}\right),
\end{aligned}$$

and the result follows. □

Proof of Corollary 3.3.1. The proof follows directly from Lemma 3.3.1 by taking $h(x) = 1 - ((l-2)/x)$ and $h(x) = h_2(x) = (1 - ((l-2)/x))\,\mathbb{I}(x > l-2)$, respectively. For the convenience of the reader, we give details on the derivation of $\mathrm{ADR}\left(\hat{\boldsymbol{\delta}}^{s+}, \boldsymbol{\delta}, \mathbb{W}\right)$. Using Lemma 3.3.1, we have

$$\begin{aligned}
\mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^{s+}, \boldsymbol{\delta}, \mathbb{W}\right) &= \mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}), \boldsymbol{\delta}, \mathbb{W}\right)\\
&- 2\mathrm{E}\left[h_2\left(\chi^2_{l+2}(\Delta)\right)\operatorname{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^*\mathbb{W}_j\right)_{(d)}\right)\right]\\
&+ 2\mathrm{E}\left[h_2\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\operatorname{trace}(\mathbb{W}_j B_j^*) - 2\mathrm{E}\left[h_2\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\operatorname{trace}(\mathbb{W}_j\Upsilon_j^*)\\
&+ 2\mathrm{E}\left[h_2\left(\chi^2_{l+4}(\Delta)\right)\right]\operatorname{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^*\mathbb{W}_j\right)_{(d)}\right)\\
&- 2\mathrm{E}\left[h_2\left(\chi^2_{l+4}(\Delta)\right)\right]\operatorname{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\right)_{(d)}' \left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\right)_{(d)}\right)\\
&+ \mathrm{E}\left[h_2^2(\chi^2_{l+2}(\Delta))\right]\prod_{j=1}^{d+1}\operatorname{trace}\left(\mathbb{W}_j\Upsilon_j^*\right)\\
&+ \mathrm{E}\left[h_2^2(\chi^2_{l+4}(\Delta))\right]\operatorname{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\right)_{(d)}' \left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j\mathbb{W}_j^{1/2}\right)_{(d)}\right),
\end{aligned}$$

where $h_2(X) = \left(1 - \frac{l-2}{X}\right)\mathbb{I}(X > l-2)$. Using Corollary 3.3.1, note that

$$
\begin{aligned}
& \mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\boldsymbol{\tau}}), \boldsymbol{\delta}, \mathbb{W}\right) = \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^s, \boldsymbol{\delta}, \mathbb{W}\right) \\
+ \quad & 2\mathrm{E}\left[h_1\left(\chi^2_{l+2}(\Delta)\right)\mathrm{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right)\right] \\
- \quad & 2\mathrm{E}\left[h_1\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j B_j^*) + 2\mathrm{E}\left[h_1\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j \Upsilon_j^*) \\
- \quad & 2\mathrm{E}\left[h_1\left(\chi^2_{l+4}(\Delta)\right)\right]\mathrm{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right) \\
+ \quad & 2\mathrm{E}\left[h_1\left(\chi^2_{l+4}(\Delta)\right)\right]\mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right) \\
- \quad & \mathrm{E}\left[h_1^2(\chi^2_{l+2}(\Delta))\right]\prod_{j=1}^{d+1}\mathrm{trace}\left(\mathbb{W}_j \Upsilon_j^*\right) \\
- \quad & \mathrm{E}\left[h_1^2(\chi^2_{l+4}(\Delta))\right]\mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right).
\end{aligned}
$$

Hence, we get

$$
\begin{aligned}
\mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^{s+}, \boldsymbol{\delta}, \mathbb{W}\right) &= \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^{s}, \boldsymbol{\delta}, \mathbb{W}\right) \\
&+ 2\mathrm{E}\left[h_1\left(\chi^2_{l+2}(\Delta)\right)\mathrm{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right)\right] \\
&- 2\mathrm{E}\left[h_1\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j B_j^*) + 2\mathrm{E}\left[h_1\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j \Upsilon_j^*) \\
&- 2\mathrm{E}\left[h_1\left(\chi^2_{l+4}(\Delta)\right)\right]\mathrm{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right) \\
&+ 2\mathrm{E}\left[h_1\left(\chi^2_{l+4}(\Delta)\right)\right]\mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right) \\
&- \mathrm{E}\left[h_1^2(\chi^2_{l+2}(\Delta))\right]\prod_{j=1}^{d+1}\mathrm{trace}\left(\mathbb{W}_j \Upsilon_j^*\right) + 2\mathrm{E}\left[h_2\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j B_j^*) \\
&- \mathrm{E}\left[h_1^2(\chi^2_{l+4}(\Delta))\right]\mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right) \\
&- 2\mathrm{E}\left[h_2\left(\chi^2_{l+2}(\Delta)\right)\mathrm{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right)\right] \\
&+ 2\mathrm{E}\left[h_2\left(\chi^2_{l+4}(\Delta)\right)\right]\mathrm{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right) \\
&- 2\mathrm{E}\left[h_2\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j \Upsilon_j^*) + \mathrm{E}\left[h_2^2(\chi^2_{l+2}(\Delta))\right]\prod_{j=1}^{d+1}\mathrm{trace}\left(\mathbb{W}_j \Upsilon_j^*\right) \\
&- 2\mathrm{E}\left[h_2\left(\chi^2_{l+4}(\Delta)\right)\right]\mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right) \\
&+ \mathrm{E}\left[h_2^2(\chi^2_{l+4}(\Delta))\right]\mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)'_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right).
\end{aligned}
$$

Since $h_1(X) - h_2(X) = (1 - \frac{l-2}{X})\mathbb{I}(X < l-2)$, then,

$$\begin{aligned}
\mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^{s+}, \boldsymbol{\delta}, \mathbb{W}\right) &= \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^{s}, \boldsymbol{\delta}, \mathbb{W}\right) \\
&+ 2\mathrm{E}\left[h_3\left(\chi^2_{l+2}(\Delta)\right)\mathrm{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B^*_j \mathbb{W}_j\right)_{(d)}\right)\right] \\
&- 2\mathrm{E}\left[h_3\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j B^*_j) + 2\mathrm{E}\left[h_3\left(\chi^2_{l+2}(\Delta)\right)\right]\prod_{j=1}^{d+1}\mathrm{trace}(\mathbb{W}_j \Upsilon^*_j) \\
&- 2\mathrm{E}\left[h_3\left(\chi^2_{l+4}(\Delta)\right)\right]\mathrm{trace}\left(\mu^{**\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B^*_j \mathbb{W}_j\right)_{(d)}\right) \\
&+ 2\mathrm{E}\left[h_3\left(\chi^2_{l+4}(\Delta)\right)\right]\mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}^{1/2}_j\right)^{\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}^{1/2}_j\right)_{(d)}\right) \\
&- \mathrm{E}\left[h_3^2(\chi^2_{l+2}(\Delta))\right]\prod_{j=1}^{d+1}\mathrm{trace}\left(\mathbb{W}_j \Upsilon^*_j\right) \\
&- \mathrm{E}\left[h_3^2(\chi^2_{l+4}(\Delta))\right]\mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}^{1/2}_j\right)^{\prime}_{(d)}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}^{1/2}_j\right)_{(d)}\right)
\end{aligned}$$

This completes the proof. □

Proof of Corollary 3.3.2. From Lemma 3.3.2, we have

$$\mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right) - \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right) = c_4 - (2c_2 - c_3). \tag{B.16}$$

Note that $c_4 = (\mathrm{Vec}(\mu^*_1))'\varpi(\mathrm{Vec}(\mu^*_1))$. Since ϖ is positive definite then by Theorem 2.4.7 in Mathai and Provost (1992), we have

$$\mathrm{Ch}_{\min}(\varpi) \leqslant \frac{c_4}{\Delta} \leqslant \mathrm{Ch}_{\max}(\varpi),$$

and so we get $c_4 - (2c_2 - c_3) \leqslant \Delta\,\mathrm{Ch}_{\max}(\varpi) - (2c_2 - c_3)$. Therefore, from (B.16), we have $\mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right) - \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right) \leqslant 0$ if $\Delta\,\mathrm{Ch}_{\max}(\varpi) - (2c_2 - c_3) \leqslant 0$. Thus, we get $\mathrm{ADR}^1\left(\tilde{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right) \leqslant \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}^0, \mathbb{W}\right)$ if $\Delta \leqslant \frac{2c_2 - c_3}{\mathrm{Ch}_{\max}(\varpi)}$. Using similar steps, the proof of the second part of the corollary follows by using the inequality $\Delta\,\mathrm{Ch}_{\min}(\varpi) - (2c_2 - c_3) \leqslant c_4 - (2c_2 - c_3)$. □

Proof of Corollary 3.3.3. Using Corollary 3.3.1 and Lemma 3.3.2, we have

$$\begin{aligned}
&\mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^s, \boldsymbol{\delta}, \mathbb{W}\right) = \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}, \hat{m}), \boldsymbol{\delta}, \mathbb{W}\right) - 2\prod_{j=1}^{d+1} \mathrm{trace}(\mathbb{W}_j B_j^*) \\
+\ & \prod_{j=1}^{d+1} \mathrm{trace}(\mathbb{W}_j \Upsilon_j^*) + \mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}' \left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right) \\
-\ & 2\mathrm{E}\left[\left(1 - (l-2)\chi_{l+2}^{-2}(\Delta)\right) \mathrm{trace}\left(\mu_{(d)}^{**\prime}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right)\right] \\
+\ & 2\mathrm{E}\left[\left(1 - (l-2)\chi_{l+2}^{-2}(\Delta)\right)\right] \prod_{j=1}^{d+1} \mathrm{trace}(\mathbb{W}_j B_j^*) - 2\mathrm{E}\left[\left(1 - (l-2)\chi_{l+2}^{-2}(\Delta)\right)\right] \prod_{j=1}^{d+1} \mathrm{trace}(\mathbb{W}_j \Upsilon_j^*) \\
+\ & 2\mathrm{E}\left[\left(1 - (l-2)\chi_{l+4}^{-2}(\Delta)\right)\right] \mathrm{trace}\left(\mu_{(d)}^{**\prime}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right) \\
-\ & 2\mathrm{E}\left[\left(1 - (l-2)\chi_{l+4}^{-2}(\Delta)\right)\right] \mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}' \left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right) \\
+\ & \mathrm{E}\left[\left(1 - (l-2)\chi_{l+2}^{-2}(\Delta)\right)^2\right] \prod_{j=1}^{d+1} \mathrm{trace}\left(\mathbb{W}_j \Upsilon_j^*\right) \\
+\ & \mathrm{E}\left[(1 - (l-2)\chi_{l+4}^{-2}(\Delta))^2\right] \mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}' \left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right),
\end{aligned}$$

By simplifying and removing common terms we have,

$$\begin{aligned}
\mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^s, \boldsymbol{\delta}, \mathbb{W}\right) &= \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}, \hat{m}), \boldsymbol{\delta}, \mathbb{W}\right) \\
&+ 2\mathrm{E}\left[\left((l-2)\chi_{l+2}^{-2}(\Delta)\right) \mathrm{trace}\left(\mu_{(d)}^{**\prime}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right)\right] \\
&- 2\mathrm{E}\left[\left((l-2)\chi_{l+2}^{-2}(\Delta)\right)\right] \prod_{j=1}^{d+1} \mathrm{trace}(\mathbb{W}_j B_j^*) \\
&- 2\mathrm{E}\left[\left((l-2)\chi_{l+4}^{-2}(\Delta)\right)\right] \mathrm{trace}\left(\mu_{(d)}^{**\prime}\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j A_j B_j^* \mathbb{W}_j\right)_{(d)}\right) \\
&+ \mathrm{E}\left[((l-2)^2\chi_{l+2}^{-4}(\Delta))\right] \prod_{j=1}^{d+1} \mathrm{trace}\left(\mathbb{W}_j \Upsilon_j^*\right) \\
&+ \mathrm{E}\left[((l-2)^2\chi_{l+4}^{-4}(\Delta))\right] \mathrm{trace}\left(\left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}' \left(\mu^{**}(\bigtimes_{j=1}^{d+1})_j \mathbb{W}_j^{1/2}\right)_{(d)}\right).
\end{aligned}$$

Then, using the identity $\mathrm{E}\left(\chi_{l+4}^{-2}(\Delta)\right) = \mathrm{E}\left(\chi_{l+2}^{-2}(\Delta)\right) - 2\mathrm{E}\left(\chi_{l+4}^{-4}(\Delta)\right)$, we have

$$2\mathrm{E}\left[\left((l-2)\,\chi_{l+2}^{-2}\,(\Delta)\right)\right] c_1 - 2\mathrm{E}\left[\left((l-2)\,\chi_{l+4}^{-2}\,(\Delta)\right)\right] c_1 = 4(l-2)\chi_{l+4}^{-4}\,(\Delta)\,c_1,$$

and hence,

$$4(l-2)\chi_{l+4}^{-4}\,(\Delta)\,c_1 + \mathrm{E}\left[((l-2)^2\,\chi_{l+4}^{-4}\,(\Delta))\right] c_4 = (l-2)(4c_1 + (l-2)c_4)\mathrm{E}\left[\chi_{l+4}^{-4}\,(\Delta))\right].$$

As such, we get

$$\begin{aligned}\mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^s, \boldsymbol{\delta}, \mathbb{W}\right) = \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}, \mathbb{W}\right) + (l-2)(4c_1 + (l-2)c_4)\mathrm{E}\left[\chi_{l+4}^{-4}\,(\Delta))\right] \\ - 2\,(l-2)\,\mathrm{E}\left[\chi_{l+2}^{-2}\,(\Delta)\right] c_2 + (l-2)^2\,\mathrm{E}\left[\chi_{l+2}^{-4}\,(\Delta)\right] c_3.\end{aligned}$$

Using the identity $\mathrm{E}\left[\chi_{l+2}^{-2}(\Delta)\right] = (l-2)\mathrm{E}\left[\chi_{l+2}^{-4}(\Delta)\right] + 2\Delta\mathrm{E}\left[\chi_{l+4}^{-4}(\Delta)\right]$, we get

$$\begin{aligned}&\mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^s, \boldsymbol{\delta}, \mathbb{W}\right) - \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}, \mathbb{W}\right) \\ &= (l-2)(4c_1 + (l-2)c_4)\mathrm{E}\left[\chi_{l+4}^{-4}\,(\Delta))\right] + (l-2)^2\,\mathrm{E}\left[\chi_{l+2}^{-4}\,(\Delta)\right] c_3 \\ &- 2\,(l-2)\left((l-2)\mathrm{E}\left[\chi_{l+2}^{-4}(\Delta)\right] + 2\Delta\mathrm{E}\left[\chi_{l+4}^{-4}(\Delta)\right]\right) c_2 \\ &= (l-2)(4c_1 + (l-2)c_4)\mathrm{E}\left[\chi_{l+4}^{-4}\,(\Delta))\right] - 2\,(l-2)^2\,\mathrm{E}\left[\chi_{l+2}^{-4}(\Delta)\right] c_2 \\ &- 4\,(l-2)\,\Delta\mathrm{E}\left[\chi_{l+4}^{-4}(\Delta)\right] c_2 + (l-2)^2\,\mathrm{E}\left[\chi_{l+2}^{-4}\,(\Delta)\right] c_3 \\ &= -\,(l-2)^2\,(2c_2 - c_3)\mathrm{E}\left[\chi_{l+2}^{-4}(\Delta)\right] - (l-2)\,(4\Delta c_2 - 4c_1 - (l-2)c_4)\,\mathrm{E}\left[\chi_{l+4}^{-4}(\Delta)\right].\end{aligned}$$

Therefore, for $\mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}^s, \boldsymbol{\delta}, \mathbb{W}\right) - \mathrm{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}, \mathbb{W}\right) \le 0$ if both of the following are satisfied

1. $2c_2 - c_3 \ge 0$,

2. $4\Delta c_2 - 4c_1 - (l-2)c_4 \ge 0$.

If $4c_1 + (l-2)c_4 = 0$ then since $c_2 \ge 0$ by definition of trace, inequality 2 holds for any

$\Delta \geq 0$. Also,

$$\begin{aligned} c_1 &= (\mathrm{Vec}\,(\mu^{**}))' \left(\bigotimes_{j=d+1}^{1} A_j B_j^* \mathbb{W}_j \right) (\mathrm{Vec}\,(\mu^{**})) \\ &= (\mathrm{Vec}\,(\mu^{**}))' \left(\bigotimes_{j=d+1}^{1} A_j B_j^* \mathbb{W}_j \right) \left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j \Upsilon_j^* A_j \right) \right) \\ &= (\mathrm{Vec}\,(\mu^{**}))' \left(\bigotimes_{j=d+1}^{1} A_j B_j^* \mathbb{W}_j \Upsilon_j^* A_j \right) (\mathrm{Vec}\,(\mu^{**})), \end{aligned}$$

and

$$\begin{aligned} c_4 &= (\mathrm{Vec}\,(\mu^{**}))' \left(\bigotimes_{j=d+1}^{1} \mathbb{W}_j \right) (\mathrm{Vec}\,(\mu^{**})) \\ &= \left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j \Upsilon_j^* A_j \right) \right)' \left(\bigotimes_{j=d+1}^{1} \mathbb{W}_j \right) \left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j \Upsilon_j^* A_j \right) \right) \\ &= (\mathrm{Vec}\,(\mu^{**}))' \left(\bigotimes_{j=d+1}^{1} A_j \Upsilon_j^* \mathbb{W}_j \Upsilon_j^* A_j \right) (\mathrm{Vec}\,(\mu^{**})). \end{aligned}$$

Hence,

$$4c_1 + (l-2)c_4 = (\mathrm{Vec}\,(\mu^{**}))' \left(\bigotimes_{j=d+1}^{1} A_j^{1/2} \Pi^* \bigotimes_{j=d+1}^{1} A_j^{1/2} \right) (\mathrm{Vec}\,(\mu^{**})),$$

where $\Pi^* = \bigotimes_{j=d+1}^{1} A_j^{1/2} \left(4 \bigotimes_{j=d+1}^{1} B_j^* + (l-2) \bigotimes_{j=d+1}^{1} \Upsilon_j^* \right) \bigotimes_{j=d+1}^{1} \mathbb{W}_j \Upsilon_j^* A_j^{1/2}$. Then,

$$\frac{4c_1 + (l-2)c_4}{\Delta} = \frac{\left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j A_j^{1/2} \right) \right)' \Pi^* \left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j A_j^{1/2} \right) \right)}{\left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j A_j^{1/2} \right) \right)' \left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j A_j^{1/2} \right) \right)},$$

and by using the identity that $x'Bx = x'(\frac{B+B'}{2})x$, for some vector x, we have

$$\frac{4c_1 + (l-2)c_4}{\Delta} = \frac{\left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j A_j^{1/2} \right) \right)' \left(\frac{\Pi^* + \Pi^{*'}}{2} \right) \left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j A_j^{1/2} \right) \right)}{\left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j A_j^{1/2} \right) \right)' \left(\mathrm{Vec} \left(\mu^{**} (\bigtimes_{j=1}^{d+1})_j A_j^{1/2} \right) \right)}.$$

Therefore, by Courant's theorem, $\frac{4c_1+(l-2)c_4}{\Delta} \leq \mathrm{Ch}_{\max}\left(\left(\frac{\Pi^*+\Pi^{*'}}{2} \right) \right) = \mathrm{Ch}_{\max}(\Pi^{**})$. Hence, if $c_2 > 0$, then $4\Delta c_2 - 4c_1 - (l-2)c_4 \geq 0$ if $c_2 \geq \frac{\mathrm{Ch}_{\max}(\Pi^{**})}{4}$. Therefore, $\mathrm{ADR}^1 \left(\hat{\boldsymbol{\delta}}^s, \boldsymbol{\delta}, \mathbb{W}_j \right) \leq$

$\text{ADR}^1\left(\hat{\boldsymbol{\delta}}(\hat{\tau}), \boldsymbol{\delta}, \mathbb{W}\right)$ if $c_2 \geq \max\left\{\frac{c_3}{2}, \frac{\text{Ch}_{\max}(\Pi^{**})}{4}\right\}$. Similarly, it can be shown that $\text{ADR}^1\left(\hat{\boldsymbol{\delta}}^{s+}, \boldsymbol{\delta}, \mathbb{W}\right) \leq \text{ADR}^1\left(\hat{\boldsymbol{\delta}}^{s}, \boldsymbol{\delta}, \mathbb{W}\right)$. □

B.3 On the convergence of the estimators of the change-points

Proposition B.3.1. *Let* $\left\|\tilde{\mathbb{U}}\right\|_F^2$ *be the restricted* SSE *under* $\tilde{\tau}$, $\left\|\hat{\mathbb{U}}\right\|_F^2$ *be the unrestricted* SSE *under* $\hat{\tau}$ *and* $\left\|\tilde{\mathbb{U}}^0\right\|_F^2$ *be the restricted* SSE *under* $\tilde{\tau}^0 = \{\tilde{\tau}_1^0, \cdots, \tilde{\tau}_{m_0}^0\}$, *respectively. Then,*

$$\left\|\hat{\mathbb{U}}\right\|_F^2 \leqslant \left\|\tilde{\mathbb{U}}\right\|_F^2 \leqslant \left\|\tilde{\mathbb{U}}^0\right\|_F^2 \leqslant \|\mathbb{U}\|_F^2 .$$

Proof. Recall,

$$\hat{\tau} = \arg\min_{\tau} \text{SSE}_T^U(\tau), \tag{B.17}$$

and

$$\tilde{\tau} = \arg\min_{\boldsymbol{\tau}} \text{SSE}_T^R(\boldsymbol{\tau}), \tag{B.18}$$

with the restriction $\delta \times_1 R_1 \times_2 R_2 \times_3 \cdots \times_{d+1} R_{d+1} = r$. Also, recall that if $A \subseteq B$, then $\text{Inf}(A) \geq \text{Inf}(B)$, provided that $\text{Inf}(A)$ and $\text{Inf}(B)$ exist. Hence, $\|\hat{\mathbb{U}}\|_F^2 \leq \|\tilde{\mathbb{U}}\|_F^2$. Moreover, from the definition of the minimum, we have $\|\tilde{U}\|_F^2 \leq \min_{\tilde{\delta}} \text{SSE}_T^R(\tau^0) = \|\tilde{\mathbb{U}}^0\|_F^2 \leq \|\mathbb{U}\|_F^2$. This completes the proof. □

In deriving the consistency of the rate of change-points, we use the following result. To simplify some notations, let
$Q^0 = \bar{Z}^0 \otimes I_{q_d} \otimes \cdots \otimes I_{q_1}$.

Proposition B.3.2. *Under Assumption 2, and assuming that the shifts in the coefficients are of fixed magnitudes independent of* T, *then*

$$Q^0\text{Vec}(\boldsymbol{\delta}^0) = O_p(T^{1/2}) \quad \textit{and} \quad (\text{Vec}(\mathbb{U}))' \, Q^0\text{Vec}(\boldsymbol{\delta}^0) = O_p(T^{1/2}).$$

Proof. First, we define $||\mathbb{A}||_F = \sqrt{(\text{Vec}(\mathbb{A}))' \, \text{Vec}(\mathbb{A})}$, for a tensor $\mathbb{A}$. Then, $||\bar{Z}^0||_F = \sqrt{\left(\text{Vec}(\bar{Z}^0)\right)' \text{Vec}(\bar{Z}^0)} = O_p(T^{1/2})$. In addition,

$$|| \, (\text{Vec}(\mathbb{U}))' \, Q^0||_F = ||\mathbb{U} \times_{d+1} \bar{Z}^{0'}||_F = \left(\sum_{p=1}^{m_0+1} \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \left(T^{1/2} \sum_{i=1}^{L_p} X_{p,i,s_1,\cdots,s_{d+1}} \right)^2 \right)^{1/2}.$$

Thus,

$$\begin{aligned} \text{P}\left(|| \, (\text{Vec}(\mathbb{U}))' \, Q^0||_F > T^{1/2}\right) &= \text{P}\left(\left(\sum_{p=1}^{m_0+1} \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \left(T^{1/2} \sum_{i=1}^{L_p} X_{p,i,s_1,\cdots,s_{d+1}} \right)^2 \right)^{1/2} > T^{1/2} \right) \\ &= \text{P}\left(\left(\sum_{p=1}^{m_0+1} \sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \left(T^{1/2} \sum_{i=1}^{L_p} X_{p,i,s_1,\cdots,s_{d+1}} \right)^2 \right) > T \right). \end{aligned}$$

By Markov's Inequality, we get

$$\text{P}\left(|| \, ((\text{Vec}(\mathbb{U})))' \, Q^0|| > T^{1/2}\right) \leqslant \sum_{p=1}^{m_0+1} \text{E}\left(\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \left(T^{1/2} \sum_{i=1}^{L_p} X_{p,i,s_1,\cdots,s_{d+1}} \right)^2 \right) / T.$$

Then,

$$\text{P}\left(|| \, ((\text{Vec}(\mathbb{U})))' \, Q^0|| > T^{1/2}\right) \leqslant \sum_{p=1}^{m_0+1} \text{E}\left(\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \left(\sum_{i=1}^{L_p} X_{p,i,s_1,\cdots,s_{d+1}} \right)^2 \right).$$

Hence, by using Lemma 3.3, we have

$$\text{E}\left(\sum_{s_1=1}^{q_1} \cdots \sum_{s_{d+1}=1}^{q_{d+1}} \left(\sum_{i=1}^{L_p} X_{p,i,s_1,\cdots,s_{d+1}} \right)^2 \right) = O\left(\sum_{i=1}^{L_p} c_{p,i}^2 \right) = o(1).$$

This implies that $||\text{Vec}(\mathbb{U})' Q^0|| = O(T^{1/2})$. In addition, by Cauchy-Shwarz's inequality,

$$||Q^0 \text{Vec}(\boldsymbol{\delta}^0)|| \leqslant ||Q^0|| ||\text{Vec}(\boldsymbol{\delta}^0)|| = ||\bar{Z}^0 \otimes I_{q_1 \cdots q_d}|| ||\text{Vec}(\boldsymbol{\delta}^0)|| = ||\bar{Z}^0||(\sqrt{q_1 \cdots q_d})||\text{Vec}(\boldsymbol{\delta}^0)||,$$

and

$$|| \, (\text{Vec}(\mathbb{U}))' \, Q^0 \text{Vec}(\boldsymbol{\delta}^0)|| \leqslant || \, (\text{Vec}(\mathbb{U}))' \, Q^0|| ||\text{Vec}(\boldsymbol{\delta}^0)||.$$

Therefore, under Assumption 3.2.3 and assuming that the shifts in the coefficients that are of fixed magnitudes which are independent of T, we have $Q^0 \text{Vec}(\boldsymbol{\delta}^0) = O_p(T^{1/2})$ and $(\text{Vec}(\mathbb{U}))' Q^0 \text{Vec}(\boldsymbol{\delta}^0) = O_p(T^{1/2})$. □

Proposition B.3.3. *Let $\bar{Z}^*$ be the diagonal partition of $\{Z_1, \cdots, Z_{m_0+1}\}$ under $\{\hat{\tau}_1, \cdots, \hat{\tau}_{m_0}\}$ and $\hat{\boldsymbol{\delta}}$ be the related unrestricted estimation of $\boldsymbol{\delta}$. If there exists a break rate (i.e. λ_j^0) which cannot be consistently estimated, then*

$$\left\|\boldsymbol{\delta}^0 \times_{d+1} \bar{Z}^0 - \hat{\boldsymbol{\delta}} \times_{d+1} \bar{Z}^*\right\|_F^2 \geqslant TC\|\boldsymbol{\delta}^0{}_{j_{(d+1)}} - \boldsymbol{\delta}^0{}_{j+1_{(d+1)}}\|_F^2$$

for some $C > 0$ with probability no less than $\epsilon > 0$.

Proof. For $\|\boldsymbol{\delta}^0 \times_{(d+1)} \bar{Z}^0 - \hat{\boldsymbol{\delta}} \times_{(d+1)} \bar{Z}^*\|_F^2$, if there exists a break rate which cannot be consistently estimated, then with some positive probability $\epsilon_0 > 0$, there exists an $\eta > 0$ such that this non-estimated break falls in the interval $\left[\tau_j^0 - T\eta, \tau_j^0 + T\eta\right]$. Suppose this interval satisfies $\hat{\tau}_{k-1} \leq \tau_j^0 - T\eta < \tau_j^0 + T\eta \leq \hat{\tau}_k$. Then, we have

$\|\delta^0 \times_{(d+1)} \bar{Z}^0 - \hat{\boldsymbol{\delta}} \times_{(d+1)} \bar{Z}^*\|_F^2 \geq \sum_{t=\tau_j^0-T\eta+1}^{\tau_j^0} ((\hat{\delta}'_{k_{(d+1)}} - \delta^{0'}_{j_{(d+1)}})z_t)'((\hat{\delta}'_{k_{(d+1)}} - \delta^{0'}_{j_{(d+1)}})z_t) + \sum_{t=\tau_j^0+1}^{\tau_j^0+T\eta} ((\hat{\delta}'_{k_{(d+1)}} - \delta^{0'}_{j+1_{(d+1)}})z_t)'((\hat{\delta}'_{k_{(d+1)}} - \delta^{0'}_{j+1_{(d+1)}})z_t)$.

Then,

$\|\delta^0 \times_{(d+1)} \bar{Z}^0 - \hat{\boldsymbol{\delta}} \times_{(d+1)} \bar{Z}^*\|_F^2 \geq \sum_{t=\tau_j^0-T\eta+1}^{\tau_j^0} z_t'(\hat{\delta}'_{k_{(d+1)}} - \delta^{0'}{}_{j_{(d+1)}})'(\hat{\delta}'_{k_{(d+1)}} - \delta^{0'}{}_{j_{(d+1)}})z_t + \sum_{t=\tau_j^0+1}^{\tau_j^0+T\eta} z_t'(\hat{\delta}'_{k_{(d+1)}} - \delta^{0'}{}_{j+1_{(d+1)}})'(\hat{\delta}'_{k_{(d+1)}} - \delta^{0'}{}_{j+1_{(d+1)}})z_t$.

This gives

$\|\delta^0\times_{(d+1)}\bar{Z}^0-\hat{\boldsymbol{\delta}}\times_{(d+1)}\bar{Z}^*\|_F^2 \geq \sum_{t=\tau_j^0-T\eta+1}^{\tau_j^0} \mathrm{trace}[z_t'(\hat{\delta}'_{k_{(d+1)}}-\delta^{0'}{}_{j_{(d+1)}})'(\hat{\delta}'_{k_{(d+1)}}-\delta^{0'}{}_{j_{(d+1)}})z_t]+\sum_{t=\tau_j^0+1}^{\tau_j^0+T\eta} \mathrm{trace}[z_t'(\hat{\delta}'_{k_{(d+1)}}-\delta^{0'}{}_{j+1_{(d+1)}})'(\hat{\delta}'_{k_{(d+1)}} - \delta^{0'}{}_{j+1_{(d+1)}})z_t]$, and then,

$\|\delta^0\times_{(d+1)}\bar{Z}^0-\hat{\boldsymbol{\delta}}\times_{(d+1)}\bar{Z}^*\|_F^2 \geq \sum_{t=\tau_j^0-T\eta+1}^{\tau_j^0} \mathrm{trace}[z_tz_t'(\hat{\delta}'_{k_{(d+1)}}-\delta^{0'}{}_{j_{(d+1)}})'(\hat{\delta}'_{k_{(d+1)}}-\delta^{0'}{}_{j_{(d+1)}})]+\sum_{t=\tau_j^0+1}^{\tau_j^0+T\eta} \mathrm{trace}[z_tz_t'(\hat{\delta}'_{k_{(d+1)}}-\delta^{0'}{}_{j+1_{(d+1)}})'(\hat{\delta}'_{k_{(d+1)}} - \delta^{0'}{}_{j+1_{(d+1)}})]$.

Hence, we have

$$\|\delta^0 \times_{(d+1)} \bar{Z}^0 - \hat{\delta} \times_{(d+1)} \bar{Z}^*\|_F^2 \geq \text{trace}\left[(\hat{\delta}_{k_{(d+1)}} - \delta^0{}_{j_{(d+1)}})' \left(\sum_{t=\tau_j^0 - T\eta+1}^{\tau_j^0} z_t z_t' \right) (\hat{\delta}_{k_{(d+1)}} - \delta^0{}_{j_{(d+1)}}) \right]$$

$$+ \quad \text{trace}\left[(\hat{\delta}_{k_{(d+1)}} - \delta^0{}_{j+1_{(d+1)}})' \left(\sum_{t=\tau_j^0+1}^{\tau_j^0+T\eta} z_t z_t' \right) (\hat{\delta}_{k_{(d+1)}} - \delta^0{}_{j+1_{(d+1)}})' \right]. \tag{B.19}$$

Let γ_t and γ_t^* be the smallest eigenvalues of $\sum_{t=\tau_j^0-T\eta+1}^{\tau_j^0} z_t z_t'$ and $\sum_{t=\tau_j^0+1}^{\tau_j^0+T\eta} z_t z_t'$ in (B.19), respectively. Then

$$\begin{aligned} \|\delta^0 \times_{(d+1)} \bar{Z}^0 &- \hat{\delta} \times_{(d+1)} \bar{Z}^*\|_F^2 \geq \gamma_t \|\hat{\delta}_{k_{(d+1)}} - \delta^0{}_{j_{(d+1)}}\|^2 + \gamma_t^* \|\hat{\delta}_{k_{(d+1)}} - \delta^0{}_{j+1_{(d+1)}}\|^2 \\ &\geq \min(\gamma_t, \gamma_t^*) \left(\|\hat{\delta}_{k_{(d+1)}} - \delta^0{}_{j_{(d+1)}}\|^2 + \|\hat{\delta}_{k_{(d+1)}} - \delta^0{}_{j+1_{(d+1)}}\|^2 \right). \end{aligned} \tag{B.20}$$

In addition, using the convexity of quadratic functions, we have

$$2\|a\|^2 + 2\|b\|^2 \geq \|a+b\|^2, \tag{B.21}$$

and combining relations (B.19),(B.20), and (B.21), we get

$$\|\delta^0 \times_{(d+1)} \bar{Z}^0 - \hat{\delta} \times_{(d+1)} \bar{Z}^*\|_F^2 \geq (1/2) \min(\gamma_t, \gamma_t^*) (\|\delta^0_{j_{(d+1)}} - \delta^0{}_{j+1_{(d+1)}}\|^2).$$

Further, let

$$\sum_{t=\tau_j^0-T\eta+1}^{\tau_j^0} z_t z_t' = \frac{T\eta}{T\eta} \sum_{t=\tau_j^0-T\eta+1}^{\tau_j^0} z_t z_t' = T\eta A_T,$$

where $A_t = \frac{1}{T\eta} \sum_{t=\tau_j^0-T\eta+1}^{\tau_j^0} z_t z_t'$. Under condition ($\mathscr{C}_2$), the smallest eigenvalue of A_T is bounded away from zero. Thus, the smallest eigenvalue of $T\eta A_T, \gamma_t$, is of order $T\eta$. Similarly, γ_t^* is also of order $T\eta$. Therefore

$$\|\delta^0 \times_{(d+1)} \bar{Z}^0 - \hat{\delta} \times_{(d+1)} \bar{Z}^*\|_F^2 \geq TC \|\delta^0_{j_{(d+1)}} - \delta^0_{j+1_{(d+1)}}\|_F^2$$

for some $C > 0$ with probability no less than $\epsilon > 0$. □

Lemma B.3.1. *If Assumption 3.2.3 holds, then,*

$$\sup_{1\leqslant i\leqslant m_0} \|(\tau_{i+1}-\tau_i)^{-1/2}\sum_{t=\tau_i+1}^{\tau_{i+1}} z_t(\mathrm{Vec}(\mathbb{U}_t))'\|_F = O_p(T^{\alpha^*}), \textit{for } \tfrac{1-\alpha}{2} < \alpha^* < 1/2.$$

Proof. Under ($\mathscr{C}_5$) and ($\mathscr{C}_6$), with $1/2-\alpha/2<\alpha^*<1/2$, we have

$$\begin{aligned}
&\mathrm{P}\left(\sup_{\epsilon T\leq(\tau_{i+1}-\tau_i)\leq T}\|(\tau_{i+1}-\tau_i)^{-1/2}\sum_{t=\tau_i+1}^{\tau_{i+1}} z_t \mathrm{Vec}(U_t)'\|_F > T^{\alpha^*}\right)\\
&= \mathrm{P}\left(\sup_{\epsilon T\leq L_{i+1}\leq T}\|(\tau_{i+1}-\tau_i)^{-1/2}T^{1/2}\sum_{t=1}^{L_{i+1}}\mathrm{Vec}(X_{i+1,t})\|_F > T^{\alpha^*}\right)\\
&\leq \mathrm{P}\left(\sup_{\epsilon T\leq L_{i+1}\leq T}\|(\tau_{i+1}-\tau_i)^{-1/2}\sum_{t=1}^{L_{i+1}}\mathrm{Vec}(X_{i+1,t})\|_F > T^{-1+\alpha^*}\right)\\
&= \sum_{L_{i+1}=[\epsilon T]}^{T}\mathrm{P}\left(L_{i+1}^{-2}\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\left(\sum_{t=1}^{L_{i+1}}X_{i+1,t,s_1,\cdots,s_{d+1}}\right)^2 > T^{-2+2\alpha^*}\right)\\
&\leq \sum_{L_{i+1}=[\epsilon T]}^{T}L_{i+1}^{-2}\mathrm{E}\left(\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\left(\sum_{t=1}^{L_{i+1}}X_{i+1,t,s_1,\cdots,s_{d+1}}\right)^2\right)/T^{-2+2\alpha^*}.
\end{aligned}$$

Since $\mathrm{E}\left(\sum_{s_1=1}^{q_1}\cdots\sum_{s_{d+1}=1}^{q_{d+1}}\left(\sum_{t=1}^{L_{i+1}}X_{i+1,t,s_1,\cdots,s_{d+1}}\right)^2\right) = O\left(\sum_{t=1}^{L_{i+1}}c_{i+1,t}^2\right) = O(T^{-\alpha})$ and $\sum_{L_{i+1}=[\epsilon T]}^{T}L_{i+1}^{-2}\leq 2\epsilon[\epsilon T]^{-1}$,

$$\mathrm{P}\left(\sup_{\epsilon T\leq(\tau_{i+1}-\tau_i)\leq T}\|(\tau_{i+1}-\tau_i)^{-1/2}\sum_{t=\tau_i+1}^{\tau_{i+1}} z_t \mathrm{Vec}(U_t)'\|_F > T^{\alpha^*}\right) = O(T^{1-\alpha-2\alpha^*}) = o(1).$$

This implies

$$\|(\tau_{i+1}-\tau_i)^{-1/2}\sum_{t=\tau_i+1}^{\tau_{i+1}} z_t \mathrm{Vec}(U_t)'\|_F = O_p(T^{\alpha^*}), \tag{B.22}$$

this completes the proof. □

Lemma B.3.2. *Let $\bar{Z}$ be the partitioned matrix of regressors based on $\{\tau_1,\cdots,\tau_{m_0}\}$. Then, under Assumption 2,* $\sup_{\tau_1,\cdots,\tau_{m_0}}(\|\mathbb{U}\times_{d+1}(\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}')\|_F) = O_p(T^{\alpha^*})$*, for some* $\frac{1-\alpha}{2}<\alpha^*<1/2$.

Proof. Note that

$$\|\mathbb{U}\times_{(d+1)}(\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}')\|_F^2 = \left(\mathrm{Vec}(\mathbb{U}\times_{(d+1)}(\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'))\right)'\mathrm{Vec}(\mathbb{U}\times_{(d+1)}(\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}')). \tag{B.23}$$

Then, we have

$$\begin{aligned}&\mathrm{Vec}(\mathbb{U}\times_{(d+1)}(\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'))'\mathrm{Vec}(\mathbb{U}\times_{(d+1)}(\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'))\\&=\sum_{i=1}^{m}\left(\sum_{t=\tau_i+1}^{\tau_{i+1}}z_t\mathrm{Vec}(U_t)'\right)'\left(\sum_{t=\tau_i+1}^{\tau_{i+1}}z_tz_t'\right)^{-1}\left(\sum_{t=\tau_i+1}^{\tau_{i+1}}z_t\mathrm{Vec}(U_t)'\right).\end{aligned}\tag{B.24}$$

Also, one can see that the summation $\sum_{t=\tau_i+1}^{\tau_{i+1}} z_t z_t'$ may contain either data from only one regime or may contain data from multiple regimes. In either case, it always contains at least $[\epsilon T/2]$ data points from one particular regime. Without loss of generality, we assume that these point are from regime k. That is, $\tau_i \le \tau_k^0 \le \tau_k^0 + [\epsilon T/2] \le \tau_{i+1}$. Then,

$$\sum_{t=\tau_i+1}^{\tau_{i+1}} z_t z_t' = \sum_{t=\tau_i+1}^{\tau_k^0} z_t z_t' + \sum_{t=\tau_k^0+1}^{\tau_k^0+[\epsilon T/2]} z_t z_t' + \sum_{t=\tau_k^0+[\epsilon T/2]+1}^{\tau_{i+1}} z_t z_t'.$$

Since the difference between the two matrices is positive definite, we have

$$\left(\sum_{t=\tau_i+1}^{\tau_{i+1}} z_t z_t'\right)^{-1} \le \left(\sum_{t=\tau_k^0+1}^{\tau_k^0+[\epsilon T/2]} z_t z_t'\right)^{-1},$$

where the notation $A \le B$ means that $B - A$ is nonnegative definite. Therefore, under Condition $(\mathscr{C}_2)$,

$$\left\|\left((\tau_{i+1}-\tau_i)^{-1}\sum_{t=\tau_i+1}^{\tau_{i+1}} z_t z_t'\right)^{-1}\right\|_F \le \left\|\left((\tau_{i+1}-\tau_i)^{-1}\sum_{t=\tau_k^0+1}^{\tau_k^0+[\epsilon T/2]} z_t z_t'\right)^{-1}\right\|_F = \frac{1}{\epsilon}O_p(1).$$

Combining the relation $(B.24)$ and Lemma B.3.1, we get

$$\|\mathbb{U}\times_{(d+1)}(\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}')\|_F^2 = \sum_{i=1}^{m}\left(\sum_{t=\tau_i+1}^{\tau_{i+1}}z_t\mathrm{Vec}(U_t)'\right)'\left(\sum_{t=\tau_i+1}^{\tau_{i+1}}z_tz_t'\right)^{-1}\left(\sum_{t=\tau_i+1}^{\tau_{i+1}}z_t\mathrm{Vec}(U_t)'\right) = O_p(T^{2\alpha^*}).$$

Therefore,

$$\sup_{\tau_1,\cdots,\tau_m}\left(\|\mathbb{U}\times_{(d+1)}(\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}')\|_F\right) = O_p(T^{\alpha^*}).$$

□

In deriving the consistency of the rate of change-points, we use the following result. To simplify some notations, let $Q^0 = \bar{Z}^0 \otimes I_{q_d} \otimes \cdots \otimes I_{q_1}$, let $\{\tau_1, \cdots, \tau_{m_0}\}$ be a partition in the set

$$V_{\epsilon,i}(C) = \{(\tau_1, \cdots, \tau_{m_0}) : C < |\tau_i - \tau_i^0| < \epsilon T, \text{ for some } i\}, \tag{B.25}$$

with associated unrestricted sum of squared residuals SSE_1^U, regressor matrix $\bar{Z}$ and the UE $\hat{\boldsymbol{\delta}}_1$. Similarly, let $\{\tau_1, \cdots, \tau_i^0, \cdots, \tau_{m_0}\}$ be a partition with associated unrestricted sum of squared residuals SSE_0^U, regressor matrix $\bar{Z}_0$ and the UE $\hat{\boldsymbol{\delta}}_0$. Let $\hat{\boldsymbol{\delta}}_{r_0} = \hat{\boldsymbol{\delta}}_1 - \hat{\boldsymbol{\delta}}_0$.

Proposition B.3.4. *If Assumption 2 and (3.7) hold, then on the set $V_{\epsilon,i}(C)$ for C large enough, 1.* $\hat{\boldsymbol{\delta}}_{r_0}(\bigtimes_{j=1}^{d+1})_j R_j = O_p(|\tau_i - \tau_i^0|T^{-1})$; *2.* $r - \hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_j R_j = |\tau_i - \tau_i^0| O_p(T^{-1})$.

Proof. By Tobing and McGilchrist (1992) on the set $V_{\epsilon,i}(C)$ for C large enough, we have

$$(\bar{Z}_1'\bar{Z}_1)^{-1} = (\bar{Z}_0'\bar{Z}_0)^{-1} + O_p\left(\frac{\tau_i - \tau_i^0}{T^2}\right).$$

Since $\|(\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}^0 + \mathbb{U}) \times_{(d+1)} \bar{Z}_1'\|_F \leq \|\bar{Z}_1'\|_F \|\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}^0\|_F + \|\mathbb{U} \times_{(d+1)} \bar{Z}_1'\|_F = O_p(T)$, then,

$$\begin{aligned}
\hat{\boldsymbol{\delta}}_{r_0} &= (\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}_1 + \mathbb{U}) \times_{(d+1)} (\bar{Z}_1'\bar{Z}_1)^{-1}\bar{Z}_1' - (\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}_0 + \mathbb{U}) \times_{(d+1)} (\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0' \\
&= (\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}_1 + \mathbb{U}) \times_{(d+1)} \left((\bar{Z}_0'\bar{Z}_0)^{-1} + O_p\left(\tfrac{\tau_i - \tau_i^0}{T^2}\right)\right)\bar{Z}_1' - (\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}_0 + \mathbb{U}) \times_{(d+1)} (\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0' \\
&= (\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}_0) \times_{(d+1)} (\bar{Z}_0'\bar{Z}_0)^{-1}(\bar{Z}_1' - \bar{Z}_0') + \mathbb{U} \times_{(d+1)} (\bar{Z}_0'\bar{Z}_0)^{-1}(\bar{Z}_1' - \bar{Z}_0') + |\tau_i - \tau_i^0| O_p(T^{-1}). \\
&= \boldsymbol{\delta}_0 \times_{(d+1)} (\bar{Z}_0'\bar{Z}_0)^{-1}(\bar{Z}_1' - \bar{Z}_0')\bar{Z}_0 + (\mathbb{U} \times_{(d+1)} (\bar{Z}_1' - \bar{Z}_0')) \times_{(d+1)} (\bar{Z}_0'\bar{Z}_0)^{-1} + |\tau_i - \tau_i^0| O_p(T^{-1}).
\end{aligned}$$

By the definitions of $\bar{Z}_0$ and $\bar{Z}_1$, $(\bar{Z}_1' - \bar{Z}_0')\bar{Z}^0$ has at most $|\tau_i - \tau_i^0|$ terms. As such, under condition $(\mathscr{C}_2)$, $(\bar{Z}_1' - \bar{Z}_0')\bar{Z}^0 = O_p(|\tau_i - \tau_i^0|)$. This gives

$$(\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}_0) \times_{(d+1)} (\bar{Z}_0'\bar{Z}_0)^{-1}(\bar{Z}_1' - \bar{Z}_0') = O_p(|\tau_i - \tau_i^0|) O_p(T^{-1}).$$

Also,

$$\mathbb{U} \times_{(d+1)} (\bar{Z}_1' - \bar{Z}_0') = O_p(|\tau_i - \tau_i^0|).$$

Therefore, $\hat{\boldsymbol{\delta}}_{r_0} = O_p(|\tau_i - \tau_i^0|T^{-1})$ and

$$\hat{\delta}_{r_0}(\bigtimes_{j=1}^{d+1})_j R_j = O_p(|\tau_i - \tau_i^0|T^{-1}). \tag{B.26}$$

In addition, since $\hat{\boldsymbol{\delta}}_1 = \hat{\boldsymbol{\delta}}_{r_0} + \hat{\boldsymbol{\delta}}_0$, and

$$\boldsymbol{r} - \hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_j R_j = (\boldsymbol{\delta}_0 - \hat{\boldsymbol{\delta}}_0)(\bigtimes_{j=1}^{d+1})_j R_j.$$

This becomes

$$\begin{aligned}
(\boldsymbol{\delta}_0 - \hat{\boldsymbol{\delta}}_0)(\bigtimes_{i=1}^{d+1})_i R_i &= \Big(\boldsymbol{\delta}_0 - Y \times_{(d+1)} (\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0'\Big)(\bigtimes_{j=1}^{d+1})_j R_j \\
&= \Big(\Big(\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}_0\Big) \times_{(d+1)} (\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0' - \Big(\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}^0 + \mathbb{U}\Big) \times_{(d+1)} (\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0'\Big)(\bigtimes_{j=1}^{d+1})_j R_j \\
&= \Big(\boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}_0 - \boldsymbol{\delta}_0 \times_{(d+1)} \bar{Z}^0 - \mathbb{U}\Big) \bigtimes_{j=1}^{d+1} R_j \Big((\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0'\Big).
\end{aligned}$$

$(\bar{Z}_0 - \bar{Z}^0)$ has at most $2\epsilon T$ non-zero terms in each column, $\|\bar{Z}_0 - \bar{Z}^0\|_F = (2\epsilon)^{1/2}O_p(T^{1/2})$ and $\bar{Z}_0'(\bar{Z}_0 - \bar{Z}^0) = 2\epsilon O_p(T)$. Choosing ϵ small enough, $r - \hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_j R_j$ can be made arbitrarily small. Therefore,

$$r - \hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_j R_j = r - \hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_j R_j - \hat{\boldsymbol{\delta}}_{r_0}(\bigtimes_{j=1}^{d+1})_j R_j = |\tau_i - \tau_i^0|O_p(T^{-1}).$$

This completes the proof. □

To simplify some mathematical expressions, let

$$\begin{aligned}
f(\tau_i, \tau_i^0) &= \mathrm{SSE}_T^R(\tau_1, \cdots, \tau_i, \cdots, \tau_{m_0}) - \mathrm{SSE}_T^R(\tau_1, \cdots, \tau_i^0, \cdots, \tau_{m_0}) \\
\boldsymbol{Q}_0^* &= R_{d+1}'[R_{d+1}(\bar{Z}_0'\bar{Z}_0)^{-1}R_{d+1}']^{-1}R_{d+1}(\bar{Z}_0'\bar{Z}_0)^{-1} \otimes \bigotimes_{i=d}^{1} \mathbb{J}_i
\end{aligned} \tag{B.27}$$

$$\boldsymbol{Q}_1^* = R_{d+1}'[R_{d+1}(\bar{Z}_0'\bar{Z}_0)^{-1}R_{d+1}']^{-1}R_{d+1} \otimes \bigotimes_{i=d}^{1} \mathbb{J}_i, \quad \mathbb{Z}_0 = (\bar{Z}_0'\bar{Z}_0)^{-1}\bar{Z}_0' \otimes \bigotimes_{i=d}^{1} I_{q_i},$$

and, for a full rank matrix $\boldsymbol{x}$, let

$$\mathcal{R}(\boldsymbol{x}) = [R_{d+1}(\boldsymbol{x}'\boldsymbol{x})^{-1}R_{d+1}']^{-1} \otimes \bigotimes_{i=d}^{1}(R_iR_i')^{-1}, \quad \mathcal{P}(\boldsymbol{x}) = R_{d+1}'[R_{d+1}(\boldsymbol{x}'\boldsymbol{x})^{-1}R_{d+1}']^{-1} \otimes \bigotimes_{j=d}^{1} \mathbb{G}_j.$$

Proposition B.3.5. *Under the conditions of Proposition B.3.4, we have*

$$f(\tau_i,\tau_i^0)=\mathrm{SSE}_1^U-\mathrm{SSE}_0^U+\left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)'\boldsymbol{Q}_1^*\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})+2\left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)'\boldsymbol{Q}_1^*\mathbb{Z}_0\mathrm{Vec}(\mathbb{U})$$
$$+2\left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)'\boldsymbol{Q}_1^*\mathbb{Z}_0\mathrm{Vec}(\boldsymbol{\delta}^0\times_{d+1}(\bar{Z}^0-\bar{Z}_0))+|\tau_i-\tau_i^0|^3O_p(T^{-2}).$$

Proof. It can be shown that

$$\mathrm{SSE}_T^R(\tau_1,\cdots,\tau_i,\cdots,\tau_{m_0})=\mathrm{SSE}_1^U+\left(\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_jR_j\right)\right)'\mathcal{R}(\bar{Z}_1)\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_jR_j\right).$$

Similarly, we have

$$\mathrm{SSE}_T^R(\tau_1,\cdots,\tau_i^0,\cdots,\tau_{m_0})=\mathrm{SSE}_0^U+\left(\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_jR_j\right)\right)'\mathcal{R}(\bar{Z}_0)\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_jR_j\right).$$

Then, we have

$$f(\tau_i,\tau_i^0)=\mathrm{SSE}_1^U-\mathrm{SSE}_0^U+\left(\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_jR_j\right)\right)'\mathcal{R}(\bar{Z}_1)\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_jR_j\right)$$
$$-\left(\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_jR_j\right)\right)'\mathcal{R}(\bar{Z}_0)\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_jR_j\right).$$

By Proposition B.3.4, $\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_jR_j\right)=|\tau_i-\tau_i^0|O_p(T^{-1})$, and by Tobing and McGilchrist (1992), on the set $V_{\epsilon,i}(C)$ for a large enough C, we have

$$[R_{d+1}(\bar{Z}_1'\bar{Z}_1)^{-1}R_{d+1}']^{-1}=[R_{d+1}(\bar{Z}_0'\bar{Z}_0)^{-1}R_{d+1}']^{-1}+O_p(\tau_i-\tau_i^0).$$

Thus, we have

$$f(\tau_i,\tau_i^0)=\mathrm{SSE}_1^U-\mathrm{SSE}_0^U+\left(\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_jR_j\right)\right)'\mathcal{R}(\bar{Z}_0)\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_jR_j\right)$$
$$-\left(\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_jR_j\right)\right)'\mathcal{R}(\bar{Z}_0)\mathrm{Vec}\left(r-\hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_jR_j\right)+|\tau_i-\tau_i^0|^3O_p(T^{-2}).$$

This gives

$$f(\tau_i,\tau_i^0)=\mathrm{SSE}_1^U-\mathrm{SSE}_0^U+\left(\mathrm{Vec}\left((\hat{\boldsymbol{\delta}}_0+\hat{\boldsymbol{\delta}}_{r_0})(\bigtimes_{j=1}^{d+1})_jR_j\right)\right)'\mathcal{R}(\bar{Z}_0)\mathrm{Vec}\left((\hat{\boldsymbol{\delta}}_0+\hat{\boldsymbol{\delta}}_{r_0})(\bigtimes_{j=1}^{d+1})_jR_j\right)$$
$$-\left(\mathrm{Vec}\left(\hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_jR_j\right)\right)'\mathcal{R}(\bar{Z}_0)\mathrm{Vec}\left(\hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_jR_j\right)-2\left(\mathrm{Vec}\left(\hat{\boldsymbol{\delta}}_{r_0}(\bigtimes_{j=1}^{d+1})_jR_j\right)\right)'\mathcal{R}(\bar{Z}_0)\mathrm{Vec}(r)$$
$$+|\tau_i-\tau_i^0|^3O_p(T^{-2}).$$

This becomes

$$f(\tau_i,\tau_i^0) = \mathrm{SSE}_1^U - \mathrm{SSE}_0^U + \left(\mathrm{Vec}\left(r - \hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_j R_j\right)\right)' \mathcal{R}(\bar{Z}_0)\mathrm{Vec}\left(r - \hat{\boldsymbol{\delta}}_1(\bigtimes_{j=1}^{d+1})_j R_j\right)$$
$$-\left(\mathrm{Vec}\left(r - \hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_j R_j\right)\right)' \mathcal{R}(\bar{Z}_0)\mathrm{Vec}\left(r - \hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_j R_j\right) + |\tau_i - \tau_i^0|^2 O_p(T^{-2}).$$

Then,

$$f(\tau_i,\tau_i^0) = \mathrm{SSE}_1^U - \mathrm{SSE}_0^U + \left(\mathrm{Vec}\left((\hat{\boldsymbol{\delta}}_0 + \hat{\boldsymbol{\delta}}_{r_0})(\bigtimes_{j=1}^{d+1})_j R_j\right)\right)' \mathcal{R}(\bar{Z}_0)\mathrm{Vec}\left((\hat{\boldsymbol{\delta}}_0 + \hat{\boldsymbol{\delta}}_{r_0})(\bigtimes_{j=1}^{d+1})_j R_j\right)$$
$$-\left(\mathrm{Vec}\left(\hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_j R_j\right)\right)' \mathcal{R}(\bar{Z}_0)\mathrm{Vec}\left(\hat{\boldsymbol{\delta}}_0(\bigtimes_{j=1}^{d+1})_j R_j\right) - 2\left(\mathrm{Vec}\left(\hat{\boldsymbol{\delta}}_{r_0}(\bigtimes_{j=1}^{d+1})_j R_j\right)\right)' \mathcal{R}(\bar{Z}_0)\mathrm{Vec}(r)$$
$$+|\tau_i - \tau_i^0|^3 O_p(T^{-2}).$$

By simplifying, we have

$$f(\tau_i,\tau_i^0) = \mathrm{SSE}_1^U - \mathrm{SSE}_0^U + \left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_0 + \hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathrm{Vec}(\hat{\boldsymbol{\delta}}_0 + \hat{\boldsymbol{\delta}}_{r_0}) - \left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_0)\right)' \boldsymbol{Q}_1^* \mathrm{Vec}(\hat{\boldsymbol{\delta}}_0)$$
$$-2\left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \mathcal{P}(\bar{Z}_0)\mathrm{Vec}(r) + |\tau_i - \tau_i^0|^3 O_p(T^{-2}).$$

This results in

$$f(\tau_i,\tau_i^0) = \mathrm{SSE}_1^U - \mathrm{SSE}_0^U + 2\left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \mathcal{P}(\bar{Z}_0)(\bigotimes_{j=d+1}^{1} R_j)\mathbb{Z}_0 \mathrm{Vec}\left(\boldsymbol{\delta}^0 \times_{d+1} \bar{Z}^0 + \mathbb{U}\right)$$
$$-\mathrm{Vec}(r) + \left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0}) + |\tau_i - \tau_i^0|^3 O_p(T^{-2}).$$

Hence,

$$f(\tau_i,\tau_i^0) = \mathrm{SSE}_1^U - \mathrm{SSE}_0^U + \left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0}) + 2\left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathbb{Z}_0 \mathrm{Vec}(\mathbb{U})$$
$$+2\left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathbb{Z}_0 \mathrm{Vec}(\boldsymbol{\delta}^0 \times_{d+1} (\bar{Z}^0 - \bar{Z}_0)) + |\tau_i - \tau_i^0|^3 O_p(T^{-2}),$$

this completes the proof. □

We also derive the following proposition which is useful in establishing the convergence of $\hat{\lambda}$ and $\tilde{\lambda}$. Let $\mathcal{J}(\boldsymbol{x}) = R'_{d+1}[R_{d+1}(\boldsymbol{x}'\boldsymbol{x})^{-1}R'_{d+1}]^{-1}R_{d+1} \otimes \bigotimes_{i=d}^{1} \mathbb{J}_i$ for a full rank matrix $\boldsymbol{x}$.

Proposition B.3.6. *Let* $(\tau_1, \cdots, \tau_{m_0}) \in V_{\epsilon,i}(C)$ *where* $V_{\epsilon,i}(C)$ *is as in* (B.25). *Then,*

1. $\dfrac{\left(\text{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathbb{Z}_0 \text{Vec}(\mathbb{U})}{|\tau_i - \tau_i^0|} = o_p(1),$ *2.* $\dfrac{\left(\text{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \mathcal{J}(\bar{Z}_0) \text{Vec}(\hat{\boldsymbol{\delta}}_{r_0})}{|\tau_i - \tau_i^0|} = O_p(1),$

3. $\dfrac{\left(\text{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathbb{Z}_0 \text{Vec}(\boldsymbol{\delta}^0 \times_{d+1} (\bar{Z}^0 - \bar{Z}_0))}{|\tau_i - \tau_i^0|} = O_p(1).$

Proof. We have,

$$\left\| \left(\text{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathbb{Z}_0 \text{Vec}(\mathbb{U}) \right\|_F \leqslant \left\| \left(\text{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \right\|_F \left\| R'_{d+1} [R_{d+1} (\bar{Z}_0' \bar{Z}_0)^{-1} R'_{d+1}]^{-1} R_{d+1} \right\|_F$$
$$\times \left\| \left((\bar{Z}_0' \bar{Z}_0)^{-1} \bar{Z}_0' \otimes \bigotimes_{i=d}^{1} \mathbb{J}_i \right) \text{Vec}(\mathbb{U}) \right\|_F .$$

One can also verify that $\left\| \left(\text{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \right\|_F = O_p(|\tau_i - \tau_i^0| T^{-1})$ and

$$\left\| \left((\bar{Z}_0' \bar{Z}_0)^{-1} \bar{Z}_0' \otimes \bigotimes_{i=d}^{1} \mathbb{J}_i \right) \text{Vec}(\mathbb{U}) \right\|_F = O_p(T^{-1} T^{1/2+\alpha^*}) = o_p(1), \tag{B.28}$$

for some $1/2 - \alpha/2 < \alpha^* < 1/2$. Hence, we have

$$\frac{\left(\text{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathbb{Z}_0 \text{Vec}(\mathbb{U})}{|\tau_i - \tau_i^0|} = O_P\left(\frac{|\tau_i - \tau_i^0| T^{-1} T T^{-1/2+\alpha^*}}{|\tau_i - \tau_i^0|} \right) = o_p(1), \tag{B.29}$$

this proves the first claim. To derive the second claim, we use similar techniques to prove that

$$\left(\text{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \mathcal{J}(\bar{Z}_0) \text{Vec}(\hat{\boldsymbol{\delta}}_{r_0}) / |\tau_i - \tau_i^0| = |\tau_i - \tau_i^0| O_p(T^{-1}) = \epsilon O_p(1),$$

which can be made arbitrarily small by choosing a small enough ϵ. To prove the third claim, note that $(\bar{Z}^0 - \bar{Z}_0)$ has at most $2\epsilon T$ non-zero terms in each column. Then, $(\bar{Z}^0 - \bar{Z}_0) = (2\epsilon)^{1/2} O_p(T^{1/2})$ and $\bar{Z}_0'(\bar{Z}^0 - \bar{Z}_0) = 2\epsilon O_p(T)$. Therefore,

$$\left(\text{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathbb{Z}_0 \text{Vec}(\boldsymbol{\delta}^0 \times_{d+1} (\bar{Z}^0 - \bar{Z}_0)) / |\tau_i - \tau_i^0| = (2\epsilon) O_p(1),$$

which can be also made arbitrarily small by choosing ϵ small enough, this completes the proof. □

Proposition B.3.7. *Under Assumption 3.2.3, and if the shifts in the coefficients are of fixed magnitudes independent of T, then,*

1. $\max_{1\leqslant j\leqslant m_0}\left|\hat{\lambda}_j-\lambda_j^0\right| \xrightarrow[T\to\infty]{\mathrm{P}} 0,$ *and* $\max_{1\leqslant j\leqslant m_0}\left|\tilde{\lambda}_j-\lambda_j^0\right| \xrightarrow[T\to\infty]{\mathrm{P}} 0.$

2. For every $\epsilon>0$*, there exists a* $C<\infty$*, such that for large enough T,*

$\max_{1\leqslant j\leqslant m_0}\Big(\mathrm{P}(|\hat{\tau}_j-\tau_j^0|>C)\Big)<\epsilon,$ *and* $\max_{1\leqslant j\leqslant m_0}\Big(\mathrm{P}(|\tilde{\tau}_j-\tau_j^0|>C)\Big)<\epsilon.$

Proof. 1. From Proposition B.3.1,

$$\|\hat{\mathbb{U}}\|_F^2 \leqslant \|\tilde{\mathbb{U}}\|_F^2 \leqslant \|\tilde{\mathbb{U}}^0\|_F^2 \leqslant \|\mathbb{U}\|_F^2.$$

Further, let $\bar{Z}^*$ be the diagonal partition of $\{Z_1,\cdots,Z_{m_0}\}$ under $\hat{\tau}$ and let $\hat{\boldsymbol{\delta}}$ be the corresponding UE of $\boldsymbol{\delta}$. Then, $\left\|\hat{\mathbb{U}}\right\|_F^2=\left\|\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0-\hat{\boldsymbol{\delta}}\times_{d+1}\bar{Z}^*+\mathbb{U}\right\|_F^2$. Then,

$$\left\|\hat{\mathbb{U}}'\right\|_F^2=\left\|\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0-\hat{\boldsymbol{\delta}}\times_{d+1}\bar{Z}^*\right\|_F^2+2\,(\mathrm{Vec}(\mathbb{U}))'\,\mathrm{Vec}(\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0-\hat{\boldsymbol{\delta}}\times_{d+1}\bar{Z}^*)+\|U\|_F^2\,.$$

Since $\hat{\boldsymbol{\delta}}=\mathbb{Y}\times_{d+1}(\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime}$ where $\mathbb{Y}=\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0+\mathbb{U}$, then

$$\begin{aligned}(\mathrm{Vec}(\mathbb{U}))'\,\mathrm{Vec}(\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0-\hat{\boldsymbol{\delta}}\times_{d+1}\bar{Z}^*)=&-\Big(\mathrm{Vec}(\mathbb{U}\times_{d+1}\bar{Z}^*(\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime})\Big)'\,\mathrm{Vec}(\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0)\\&-\left\|\mathbb{U}\times_{d+1}\bar{Z}^*(\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime}\right\|_F^2+(\mathrm{Vec}(\mathbb{U}))'\,\mathrm{Vec}(\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0).\end{aligned}$$

From Assumption 3.2.2 and Proposition B.3.2, $\|\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0\|_F=O_p(T^{1/2})$, and by Lemma B.3.2, $\Big(\mathrm{Vec}(\mathbb{U}\times_{d+1}\bar{Z}^*(\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime})\Big)'=O_p(T^{\alpha^*})$. Then, $\left\|\mathbb{U}\times_{d+1}\bar{Z}^*(\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime})\right\|_F^2=O_p(T^{2\alpha^*})$. Therefore, under Assumption 2,

$$\begin{aligned}&\left\|\Big(\mathrm{Vec}(\mathbb{U}\times_{d+1}\bar{Z}^*(\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime})\Big)'\,\mathrm{Vec}(\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0)\right\|_F\\\leqslant&\left\|\Big(\mathrm{Vec}(\mathbb{U}\times_{d+1}\bar{Z}^*(\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime})\Big)'\right\|_F\left\|\mathrm{Vec}(\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0)\right\|_F=O_p(T^{1/2+\alpha^*})\end{aligned}$$

uniformly over all partitions. Hence, by Proposition B.3.2, we have

$$(\mathrm{Vec}(\mathbb{U}))'\,\mathrm{Vec}(\boldsymbol{\delta}^0\times_{d+1}\bar{Z}^0-\hat{\boldsymbol{\delta}}\times_{d+1}\bar{Z}^*)=O_p(T^{1/2+\alpha^*}).\tag{B.30}$$

Further, by combining Proposition B.3.3 and (B.30), we have that if at least one break is not consistently estimated, then

$$T^{-1}\left\|\tilde{\mathbb{U}}\right\|_F^2 \geq T^{-1}\left\|\hat{\mathbb{U}}\right\|_F^2 \geq T^{-1}\|\mathbb{U}\|_F^2 + C\left\|\boldsymbol{\delta}^0_{j_{(d+1)}} - \boldsymbol{\delta}^0_{j+1_{(d+1)}}\right\|_F^2 + o_p(1) > T^{-1}\|\mathbb{U}\|_F^2$$

holds with some positive probability. But, this contradicts $T^{-1}\left\|\tilde{\mathbb{U}}\right\|_F^2 \leqslant T^{-1}\|\mathbb{U}\|_F^2$, this proves the first claim of Part 1. The second claim of Part 1 is established in a similar way.

2. Let $(\tau_1, \cdots, \tau_{m_0}) \in V_{\epsilon,i}(C)$. Then, by Proposition B.3.5 along with the notations of Proposition B.3.4,

$$\begin{aligned} f(\tau_i, \tau_i^0) \;\; = \;\; & \mathrm{SSE}_1^U - \mathrm{SSE}_0^U + 2\left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathbb{Z}_0 \mathrm{Vec}(\mathbb{U}) + \left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \mathcal{J}(\bar{Z}_0)\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0}) \\ & +2\left(\mathrm{Vec}(\hat{\boldsymbol{\delta}}_{r_0})\right)' \boldsymbol{Q}_1^* \mathbb{Z}_0 \mathrm{Vec}(\boldsymbol{\delta}^0 \times_{d+1} (\bar{Z}^0 - \bar{Z}_0)) + |\tau_i - \tau_i^0|^3 O_p(T^{-2}). \qquad \text{(B.31)} \end{aligned}$$

Further, as in Bai and Perron (1998), we have

$$\begin{aligned} \frac{\mathrm{SSE}_1^U - \mathrm{SSE}_0^U}{|\tau_i - \tau_i^0|} \;\; \geqslant \;\; & 2^{-1}\left(\mathrm{Vec}(\boldsymbol{\delta}^0_{i+1} - \boldsymbol{\delta}^0_i)\right)' \tfrac{(\bar{Z}_1 - \bar{Z}_0)(\bar{Z}_1 - \bar{Z}_0)' \otimes I_{q_1 \cdots q_d}}{|\tau_i - \tau_i^0|} \left(\mathrm{Vec}(\boldsymbol{\delta}^0_{i+1} - \boldsymbol{\delta}^0_i)\right) \\ & -\epsilon O_p(1) - \rho O_p(1), \qquad \text{(B.32)} \end{aligned}$$

where ϵ and ρ can be made arbitrarily small by choosing a small ϵ and a large T. Moreover, $(\bar{Z}_1 - \bar{Z}_0)(\bar{Z}_1 - \bar{Z}_0)' \otimes I_{q_1 \cdots q_d}$ has at most $2|\tau_i - \tau_i^0|$ terms and under $(\mathscr{C}_2)$, it has a minimum eigenvalue bounded away from zero. Thus, the term

$$2^{-1}\left(\mathrm{Vec}(\boldsymbol{\delta}^0_{i+1} - \boldsymbol{\delta}^0_i)\right)' \left((\bar{Z}_1 - \bar{Z}_0)(\bar{Z}_1 - \bar{Z}_0)' \otimes I_{q_1 \cdots q_d}\right)\left(\mathrm{Vec}(\boldsymbol{\delta}^0_{i+1} - \boldsymbol{\delta}^0_i)\right)/|\tau_i - \tau_i^0|$$

is positive and dominates the other two terms by choosing a small enough ϵ and ρ.

Hence, by combining Proposition B.3.6 with (B.31) and (B.32), we conclude first that $(\mathrm{SSE}_1^U - \mathrm{SSE}_0^U)/(|\tau_i - \tau_i^0|)$ dominates all others and is positive with probability one for large T. However, $\mathrm{SSE}_1^U = \mathrm{SSE}_T^U(\tau_1, \tau_2, \cdots, \tau_{m_0})$ is the minimum among all possible values of τ and we thus have $\mathrm{SSE}_1^U \leq \mathrm{SSE}_T^U(\tau_1, \tau_2, \cdots, \tau_{m_0})$, which is a contradiction, this proves the first claim of Part 2. Further, by combining Proposition B.3.6 with (B.31) and

(B.32), we also conclude that $f(\tau_i, \tau_i^0)/|\tau_i - \tau_i^0| > 0$. But, since $\mathrm{SSE}_T^R(\tau_1, \cdots, \tau_i, \cdots, \tau_{m_0})$ is the minimization among all possible values of $\{\tau_1, \cdots, \tau_{m_0}\}$, we get $f(\tau_i, \tau_i^0) \leqslant 0$, with probability one. This is a contradiction, which shows that $|\tilde{\tau}_i - \tau_i^0| < C$ when T is large, this completes the proof. □

Proposition B.3.8. *Suppose that Assumption 3.2.3 holds and* $r_{0i} = \boldsymbol{\delta}^0_{T,i+1} - \boldsymbol{\delta}^0_{T,i} = \nu_T r_{0i}$, *where* r_{0i} *is independent of* T *and* $\nu_T > 0$ *with* $\nu_T \to 0$ *and as* $T \to \infty$, $T^{-1/2-\eta\nu} \to \infty$, *for some* $\eta \in (0, 1/2)$. *Then, 1.* $\max_{1 \leqslant j \leqslant m_0} \left|\hat{\lambda}_j - \lambda_j^0\right| \xrightarrow[T\to\infty]{\mathrm{P}} 0$ *and* $\max_{1 \leqslant j \leqslant m_0} \left|\tilde{\lambda}_j - \lambda_j^0\right| \xrightarrow[T\to\infty]{\mathrm{P}} 0.$
2. For every $\epsilon > 0$, *there exists a* $C < \infty$, *such that for large* T,

$$\max_{1 \leqslant j \leqslant m_0} \left(\mathrm{P}(\nu_T^2 |\hat{\tau}_j - \tau_j^0| > C)\right) < \epsilon \qquad \textit{and} \qquad \max_{1 \leqslant j \leqslant m_0} \left(\mathrm{P}(\nu_T^2 |\tilde{\tau}_j - \tau_j^0| > C)\right) < \epsilon.$$

Proof. We have $\boldsymbol{\delta}^0_{T,i+1} - \boldsymbol{\delta}^0_{T,i} = O(\nu_T)$, which implies that $\boldsymbol{\delta}^0_{T,i} - \boldsymbol{\delta}^0_{T,j} = O(\nu_T)$ for all i and j. Further, $\boldsymbol{\delta}^0 \times_{d+1} (\bar{Z}^0 - \bar{Z}^*)$ depends on changes in the parameters (i.e. $\boldsymbol{\delta}_i^0 - \boldsymbol{\delta}_j^0$ for some i and j). To see this, without loss of generality, consider the case where $m = 1$ and assume that $\tau_1 < \tau_1^0$. Then,

$$\mathrm{Vec}(\boldsymbol{\delta}^0 \times_{d+1} (\bar{Z}^0 - \bar{Z}^*)) = (\boldsymbol{\delta}^0_{1_{(d+1)}} - \boldsymbol{\delta}^0_{2_{(d+1)}})'(0, \cdots, 0, z'_{T_1+1}, \cdots, z'_{T_1^0}, 0, \cdots, 0))'.$$

This implies that $\boldsymbol{\delta}^0 \times_{d+1} (\bar{Z}^0 - \bar{Z}^*)$ is at most $O_p(T^{1/2}\nu_T)$. Further, in similar ways as in proof of Proposition B.1, one proves that

$$\begin{aligned}
\left\|\left(\mathrm{Vec}(\mathbb{U} \times_{d+1} \bar{Z}^*(\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime})\right)' \mathrm{Vec}(\boldsymbol{\delta}^0 \times_{d+1} (\bar{Z}^0 - \bar{Z}^*))\right\|_F &= O_p(T^{1/2+\alpha^*}\nu_T), \\
(\mathrm{Vec}(\mathbb{U}))' \,\mathrm{Vec}(\boldsymbol{\delta}^0 \times_{d+1} (\bar{Z}^* - \bar{Z}^0)) &= O_p(T^{1/2}\nu_T). \\
(\mathrm{Vec}(\mathbb{U}))' \,\mathrm{Vec}(\hat{\boldsymbol{\delta}} \times_{d+1} \bar{Z}^* - \boldsymbol{\delta}^0 \times_{d+1} \bar{Z}^0) &= O_p\left(T^{1/2+\alpha^*}\nu_T\right), \\
\left\|\boldsymbol{\delta}^0 \times_{d+1} \bar{Z}^0 - \hat{\boldsymbol{\delta}} \times_{d+1} \bar{Z}^*\right\|_F^2 \geqslant TC \left\|\boldsymbol{\delta}^0_{j_{(d+1)}} - \boldsymbol{\delta}^0_{j+1_{(d+1)}}\right\|_F &\geqslant TC\nu_T^2,
\end{aligned}$$

in the case where at least one break is not consistently estimated. Hence, together with some algebraic computations, we get

$$T^{-1}\left\|\tilde{\mathbb{U}}\right\|_F^2 \geq T^{-1}\left\|\hat{\mathbb{U}}\right\|_F^2 \geqslant T^{-1}\left\|\mathbb{U}\right\|_F^2 + C\nu_T^2 + o_p(1) > T^{-1}\left\|\mathbb{U}\right\|_F^2$$

with some positive probability. This is a contradiction with Proposition 2.1, this proves the first claim of Part **(1)**. The other statements are established in a similar way, this completes the proof. □

Proof of Theorem 3.4.1. **(1).** We have

$$\mathrm{IC}(m_0) = \sum_{j=1}^{m_0+1} \sum_{i=\hat{\tau}_{i-1}}^{\hat{\tau}_j} \left(\mathrm{Vec}(Y_i) - \hat{\boldsymbol{\delta}}'_{j_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(Y_i) - \hat{\boldsymbol{\delta}}'_{j_{(d+1)}} z_i\right) + (m+1)(q_{d+1}+1)\log(T),$$

where $\hat{\tau}$ is obtained via (2.3). Moreover, define

$$\mathrm{IC}^0(m_0) = \sum_{j=1}^{m_0+1} \sum_{i=\tau^0_{i-1}}^{\tau^0_j} \left(\mathrm{Vec}(Y_i) - \hat{\boldsymbol{\delta}}^{0\prime}_{j_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(Y_i) - \hat{\boldsymbol{\delta}}^{0\prime}_{j_{(d+1)}} z_i\right) + (m+1)(q_{d+1}+1)\log(T),$$

where $\hat{\boldsymbol{\delta}}^0_j = \bar{Y}_j \times_{d+1} \left(Z^{0\prime}_j Z^0_j\right)^{-1} Z^{0\prime}_j$; $Z^0_j = (z_{\tau_{j-1}+1}, \cdots, z_{\tau_j})'$; $\bar{Y}_j = Y_{\tau_{j-1}+1} \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} Y_{\tau_j}$. Since $\hat{\tau}$ is obtained by minimizing SSR^U_T, we get $\mathrm{IC}(m_0) \leqslant \mathrm{IC}^0(m_0)$ with probability 1. Thus, it remains to show that $\mathrm{IC}(m) > \mathrm{IC}^0(m_0)$, for all $m < m_0$ with probability 1. For some positive integer m^* such that $0 < m^* < m_0$, suppose that the corresponding estimated change-points are $\hat{\tau}^* = (\hat{\tau}^*_1, \hat{\tau}^*_2, \cdots, \hat{\tau}^*_{m^*})$, and the corresponding UE is $\hat{\boldsymbol{\delta}}^*_j = \bar{Y}^*_j \times_{d+1} \left(\bar{Z}^{*\prime}_1 \bar{Z}^*_1\right)^{-1} \bar{Z}^{*\prime}_1$, $Z^*_j = (z_{\tau^*_{j-1}+1}, \cdots, z_{\tau^*_j})'$ and $\bar{Y}^*_j = Y_{\tau^*_{j-1}+1} \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} Y^*_{\tau^*_j}$. Then, for $m < m_0$, we have

$$\mathrm{IC}(m) - \mathrm{IC}^0(m_0) = \sum_{j=1}^{m_0+1} \sum_{i=\hat{\tau}_{i-1}}^{\hat{\tau}_j} \left(\mathrm{Vec}(Y_i) - \hat{\boldsymbol{\delta}}^{*\prime}_{j_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(Y_i) - \hat{\boldsymbol{\delta}}^{*\prime}_{j_{(d+1)}} z_i\right)$$
$$- \sum_{j=1}^{m_0+1} \sum_{i=\tau^0_{i-1}}^{\tau^0_j} \left(\mathrm{Vec}(Y_i) - \hat{\boldsymbol{\delta}}^{0\prime}_{j_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(Y_i) - \hat{\boldsymbol{\delta}}^{0\prime}_{j_{(d+1)}} z_i\right) + (m^* - m_0)(q_{d+1}+1)\log(T).$$

Since $m^* < m_0$, then there exists at least one change-point that cannot be consistently estimated. Without loss of generality, let that change-point be τ^0_j. Using similar techniques as in Proposition B.3.3, we get

$$1/T\left(\mathrm{IC}(m) - \mathrm{IC}^0(m_0)\right) \geqslant C^* \|\delta^0_{j_{(d+1)}} - \delta^0_{j+1_{(d+1)}}\|^2 + o_p(1), \tag{B.33}$$

for all $m < m_0$, for some $C^* > 0$ with probability 1. Therefore, for large T, $\mathrm{IC}(m) > \mathrm{IC}^0(m_0)$ with probability 1, for all $m < m^0$, and then, $\lim_{T\to\infty} \mathrm{P}\left(\mathrm{IC}(m_0) < \mathrm{IC}(m)\right) = 1$, for all $m < m^0$.

(2). For $m* > m_0$, let the estimated change points be $\hat{\tau}^* = (\hat{\tau}^*_1, \cdots, \hat{\tau}^*_{m^*})$. Then,

$$\mathrm{IC}(m^*) - \mathrm{IC}^0(m_0) = \sum_{i=0}^{T} \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{*\prime}_{j_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{*\prime}_{j_{(d+1)}} z_i\right)$$
$$- \sum_{i=0}^{T} \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{(0)\prime}_{i_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{(0)\prime}_{i_{(d+1)}} z_i\right) + (m^* - m_0)(q_{d+1} + 1)\log(T).$$

where $\hat{\boldsymbol{\delta}}^{(*)}_i = \sum_{j=1}^{m^*+1} \hat{\boldsymbol{\delta}}^*_j \mathbb{I}(\hat{\tau}^*_{j-1} < t_i \leqslant \hat{\tau}^*_j)$ and $\hat{\boldsymbol{\delta}}^{(0)}_i = \sum_{j=1}^{m_0+1} \hat{\boldsymbol{\delta}}^0_j \mathbb{I}(\tau^0_{j-1} < t_i \leqslant \tau^0_j)$. It should be noted that when $m^* > m_0$, there are $m^* - m_0$ estimated change points that divide $[0, T]$ into $m^* - m_0 + 1$ different regimes such that within each regime the number of estimated change points is equal to the number of the actual change points. Hence, denote these $m^* - m_0$ change-points as $\tilde{\tau}^*_j = (\tau^*_1, \cdots, \tau^*_{m^*-m_0})$. Let $\tau^*_0 = 0$ and $\tau^*_{m^*-m_0+1} = T$. Hence,

$$\mathrm{IC}(m^*) - \mathrm{IC}^0(m_0) = \sum_{j=1}^{m^*-m_0+1} \sum_{i=\tilde{\tau}^*_{j-1}}^{\tilde{\tau}^*_j} \Big[\left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{*\prime}_{i_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{*\prime}_{i_{(d+1)}} z_i\right)$$
$$- \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{(0)\prime}_{i_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{(0)\prime}_{i_{(d+1)}} z_i\right) + \tfrac{(m^*-m_0)}{m^*-m_0+1}(q_{d+1} + 1)\log(T)\Big].$$

Define $\delta^{(0)}_i = \sum_{j=1}^{m^*+1} \delta^0_j \mathbb{I}(\tau^0_{j-1} < t_i \leqslant \tau^0_j)$. This becomes

$$\mathrm{IC}(m^*) - \mathrm{IC}^0(m_0) = \sum_{j=1}^{m^*-m_0+1} \sum_{i=\tilde{\tau}^*_{j-1}}^{\tilde{\tau}^*_j} z'_i \left(\delta^{(0)\prime}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(*)\prime}_{i_{(d+1)}}\right)' \left(\delta^{(0)\prime}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(*)\prime}_{i_{(d+1)}}\right) z_i$$
$$- \sum_{j=1}^{m^*-m_0+1} \sum_{i=\tilde{\tau}^*_{j-1}}^{\tilde{\tau}^*_j} z'_i \left(\delta^{(0)\prime}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(0)\prime}_{i_{(d+1)}}\right)' \left(\delta^{(0)\prime}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(0)\prime}_{i_{(d+1)}}\right) z_i \qquad \text{(B.34)}$$
$$+ 2 \sum_{j=1}^{m^*-m_0+1} \sum_{i=\tilde{\tau}^*_{j-1}}^{\tilde{\tau}^*_j} (\mathrm{Vec}(\mathbb{U}_i))' \left(\hat{\boldsymbol{\delta}}^{(*)\prime}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(0)\prime}_{i_{(d+1)}}\right) z_i + \sum_{j=1}^{m^*-m_0+1} \sum_{i=\tilde{\tau}^*_{j-1}}^{\tilde{\tau}^*_j} \tfrac{(m^*-m_0)}{m^*-m_0+1}(q_{d+1} + 1)\log(T).$$

Hence, it suffices to show that for each $(\tilde{\tau}^*_{j-1}, \tilde{\tau}^*_j]$, $j = 1, \cdots, m^* - m_0 + 1$,

$$\sum_{i=\tilde{\tau}^*_{j-1}}^{\tilde{\tau}^*_j} z'_i \left(\delta^{(0)\prime}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(*)\prime}_{i_{(d+1)}}\right)' \left(\delta^{(0)\prime}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(*)\prime}_{i_{(d+1)}}\right) z_i - \sum_{i=\tilde{\tau}^*_{j-1}}^{\tilde{\tau}^*_j} z'_i \left(\delta^{(0)\prime}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(0)\prime}_{i_{(d+1)}}\right)' \left(\delta^{(0)\prime}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(0)\prime}_{i_{(d+1)}}\right) z_i$$
$$+ 2 \sum_{i=\tilde{\tau}^*_{j-1}}^{\tilde{\tau}^*_j} (\mathrm{Vec}(\mathbb{U}_i))' \left(\hat{\boldsymbol{\delta}}^{(*)\prime}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(0)\prime}_{i_{(d+1)}}\right) z_i + \sum_{i=\tilde{\tau}^*_{j-1}}^{\tilde{\tau}^*_j} \tfrac{(m^*-m_0)}{m^*-m_0+1}(q_{d+1} + 1)\log(T) \qquad \text{(B.35)}$$

is positive with probability one, whenever T is large. We first consider the case where there are no change-points within $(\tilde{\tau}^*_{j-1}, \tilde{\tau}^*_j]$. Then, we have $\tau^0_{k^*-1} < \tilde{\tau}^*_{j-1} < \tilde{\tau}^*_j < \tau^0_{k^*}$, for some

$k^* > 0$. Then, we have

$$\hat{\boldsymbol{\delta}}_i^{(*)} = \delta_i^{(0)} + \bar{\mathbb{U}}_i^* \times_{d+1} \left(\bar{Z}_j^{*\prime}\bar{Z}_j^*\right)^{-1} \bar{Z}_j^{*\prime} \text{ and } \hat{\boldsymbol{\delta}}_i^{(*)} = \delta_i^{(0)} + \bar{\mathbb{U}}_i^0 \times_{d+1} \left(\bar{Z}_j^{0\prime}\bar{Z}_j^0\right)^{-1} \bar{Z}_j^{0\prime}.$$

Using these expressions, (B.35) becomes

$$\sum_{i=\tilde{\tau}_{j-1}^*}^{\tilde{\tau}_j^*} \left[z_i' \left(\delta_{i_{(d+1)}}^{(0)\prime} - \hat{\boldsymbol{\delta}}_{i_{(d+1)}}^{(*)\prime}\right)' \left(\delta_{i_{(d+1)}}^{(0)\prime} - \hat{\boldsymbol{\delta}}_{i_{(d+1)}}^{(*)\prime}\right) z_i - z_i' \left(\delta_{i_{(d+1)}}^{(0)\prime} - \hat{\boldsymbol{\delta}}_{i_{(d+1)}}^{(0)\prime}\right)' \left(\delta_{i_{(d+1)}}^{(0)\prime} - \hat{\boldsymbol{\delta}}_{i_{(d+1)}}^{(0)\prime}\right) z_i \right]$$
$$+2 \sum_{i=\tilde{\tau}_{j-1}^*}^{\tilde{\tau}_j^*} (\mathrm{Vec}(\mathbb{U}_i))' \left(\hat{\boldsymbol{\delta}}_{i_{(d+1)}}^{(*)\prime} - \hat{\boldsymbol{\delta}}_{i_{(d+1)}}^{(0)\prime}\right) z_i + \sum_{i=\tilde{\tau}_{j-1}^*}^{\tilde{\tau}_j^*} \tfrac{(m^*-m_0)}{m^*-m_0+1}(q_{d+1}+1)\log(T).$$

Since $\hat{\boldsymbol{\delta}}_i^{(*)} = \delta_i^{(0)} + \bar{\mathbb{U}}_j^* \times_{d+1} \left(\bar{Z}_j^{*\prime}\bar{Z}_j^*\right)^{-1} \bar{Z}_j^{*\prime}$ and $\hat{\boldsymbol{\delta}}_i^{(0)} = \delta_i^{(0)} + \bar{\mathbb{U}}_j^0 \times_{d+1} \left(\bar{Z}_j^{0\prime}\bar{Z}_j^0\right)^{-1} \bar{Z}_j^{0\prime}$, where $\bar{\mathbb{U}}^* = \mathbb{U}_{\tilde{\tau}_{j-1}^*+1} \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \mathbb{U}_{\tilde{\tau}_j^*}$ and $\bar{\mathbb{U}}^* = \mathbb{U}_{\tau_{k-1}^0+1} \boxplus_{(d+1)} \cdots \boxplus_{(d+1)} \mathbb{U}_{\tilde{\tau}_k^0}$, (B.35) becomes

$$\sum_{i=\tilde{\tau}_{j-1}^*}^{\tilde{\tau}_j^*} z_i' \left(\left(\bar{\mathbb{U}}_j^* \times_{d+1} \bar{Z}_j^{*\prime}\right)_{(d+1)}' \left(\sum_{i=\tilde{\tau}_{j-1}^*+1}^{\tilde{\tau}_j^*} z_i z_i' \right)^{-1} \right)' \left(\left(\bar{\mathbb{U}}_j^* \times_{d+1} \bar{Z}_j^{*\prime}\right)_{(d+1)}' \left(\sum_{i=\tilde{\tau}_{j-1}^*+1}^{\tilde{\tau}_j^*} z_i z_i' \right)^{-1} \right) z_i$$
$$- \sum_{t_i=\tilde{\tau}_{j-1}^0}^{\tilde{\tau}_j^0} z_i' \left(\left(\bar{\mathbb{U}}_k^0 \times_{d+1} \bar{Z}_k^{0\prime}\right)_{(d+1)}' \left(\sum_{i=\tau_{k-1}^0+1}^{\tau_k^0} z_i z_i' \right)^{-1} \right)' \left(\left(\bar{\mathbb{U}}_k^0 \times_{d+1} \bar{Z}_k^{0\prime}\right)_{(d+1)}' \left(\sum_{i=\tau_{k-1}^0+1}^{\tau_k^0} z_i z_i' \right)^{-1} \right) z_i$$
$$+ \sum_{i=\tilde{\tau}_{j-1}^*}^{\tilde{\tau}_j^*} 2\,(\mathrm{Vec}(\mathbb{U}_i))' \left(\left(\bar{\mathbb{U}}_j^* \times_{d+1} \bar{Z}_j^{*\prime}\right)_{(d+1)}' \left(\sum_{i=\tilde{\tau}_{j-1}^*+1}^{\tilde{\tau}_j^*} z_i z_i' \right)^{-1} - \left(\bar{\mathbb{U}}_k^0 \times_{d+1} \bar{Z}_k^{0\prime}\right)_{(d+1)}' \left(\sum_{i=\tau_{k-1}^0+1}^{\tau_k^0} z_i z_i' \right)^{-1} \right) z_i$$
$$+ \sum_{t_i=\tilde{\tau}_{j-1}^*}^{\tilde{\tau}_j^*} \tfrac{(m^*-m_0)}{m^*-m_0+1}(q_{d+1}+1)\log(T).$$

Then, the first term is bounded by

$$\left\| \left(T^{-1/2}\bar{\mathbb{U}}_j^* \times_{d+1} \bar{Z}_j^{*\prime}\right)_{(d+1)} \right\|_{\mathrm{F}}^2 \left\| \left(1/T \sum_{i=\tilde{\tau}_{j-1}^*+1}^{\tilde{\tau}_j^*} z_i z_i' \right)^{-1} \right\|_{\mathrm{F}}.$$

Thus, by similar arguments as in Proposition B.3.3 and Lemma B.3.1, for $1 - \alpha/2 < \alpha^* < 1/2$, we get

$$\sum_{t_i=\tilde{\tau}_{j-1}^*}^{\tilde{\tau}_j^*} z_i' \left(\left(\bar{\mathbb{U}}_j^* \times_{d+1} \bar{Z}_j^{*\prime}\right)_{(d+1)}' \left(\sum_{i=\tilde{\tau}_{j-1}^*+1}^{\tilde{\tau}_j^*} z_i z_i' \right)^{-1} \right)' \left(\left(\bar{\mathbb{U}}_j^* \times_{d+1} \bar{Z}_j^{*\prime}\right)_{(d+1)}' \left(\sum_{i=\tilde{\tau}_{j-1}^*+1}^{\tilde{\tau}_j^*} z_i z_i' \right)^{-1} \right) z_i = O_p\left(\log(T)^{2\alpha^*}\right).$$

Similar results hold for the second and third terms. Therefore, for large T, (B.34) is dominated by $\frac{(m^*-m_0)}{m^*-m_0+1}(q_{d+1}+1)\log(T)$. Since this term is positive, we conclude that, for large

T, the term in (B.34) is positive with probability 1. Next, we consider the case where there exist m_j exact change points in $(\tilde{\tau}^*_{j-1}, \tilde{\tau}^*_j]$, where $0 < m_j < m_0$. Since

$$\sum_{t_i=\tilde{\tau}^*_{j-1}}^{\tilde{\tau}^*_j} \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{(0)'}_{i_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{(0)'}_{i_{(d+1)}} z_i\right) \leqslant \sum_{t_i=0}^{T} \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{(0)'}_{i_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{(0)'}_{i_{(d+1)}} z_i\right)$$

$$= \mathrm{trace}\left(\left(\bar{\mathbb{U}} \times_{d+1} \bar{Z}'\right)_{(d+1)} \left(\bar{\mathbb{U}} \times_{d+1} \bar{Z}'\right)'_{(d+1)} \left(\sum_{i=\tau^0_{j-1}+1}^{\tau^0_j} z_i z_i' \right)^{-1} \right) = O_p(\log T^{2\alpha^*}).$$

In addition, using similar arguments as in the proof of Proposition B.1, we have that

$$\sum_{t_i \in (\tilde{\tau}^*_{j-1}, \tilde{\tau}^*_j]} 2\,(\mathrm{Vec}(\mathbb{U}_i))' \left(\hat{\boldsymbol{\delta}}^{(*)'}_{i_{(d+1)}} - \hat{\boldsymbol{\delta}}^{(0)'}_{i_{(d+1)}}\right) z_i = O_p(\log T^{2\alpha^*}).$$

Since for large T, $(\log T)^{2\alpha^*} < \frac{(m^*-m_0)}{m^*-m_0+1}(q_{d+1}+1)\log(T)$, we have, for large T, (B.34) is dominated by either

$$\frac{(m^* - m_0)}{m^* - m_0 + 1}(q_{d+1} + 1)\log(T) \quad \text{or} \quad \sum_{t_i \in (\tilde{\tau}^*_{j-1}, \tilde{\tau}^*_j]} \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{*'}_{i_{(d+1)}} z_i\right)' \left(\mathrm{Vec}(\mathbb{Y}_i) - \hat{\boldsymbol{\delta}}^{*'}_{i_{(d+1)}} z_i\right),$$

which are both positive. This implies that, for large T, the term in (B.34) is positive with probability 1. Therefore, $\lim_{T\to\infty} \mathrm{P}\left(\mathrm{IC}(m) > \mathrm{IC}^0(m_0)\right) = 1$, for $m > m_0$, this proves Part **(2)**. Part **(3)** follows directly from Parts **(1)** and **(2)**, and Part **(4)** follows directly from Parts **(1)**-**(3)**, this completes the proof. □

B.4 Algorithm for estimating location of change-points

In this section, we outline the dynamic programming algorithm used in estimating the change-point locations. Let $H^U_1(r, T_r) = \min_{\boldsymbol{\tau}} \mathrm{SSR}^U_{[0,T_r]}(\boldsymbol{\tau})$ and $H^R_1(r, T_r) = \min_{\boldsymbol{\tau}} \mathrm{SSR}^R_{[0,T_r]}(\boldsymbol{\tau})$ to be the unrestricted and restricted residual sum of squares, respectively, computed based on the optimal partition of the time interval $[0, T_r]$ that contains r change-points. Also, let $H^U_2(a, b) = \min_{\boldsymbol{\tau}} \mathrm{SSR}^U_{[a,b]}(\boldsymbol{\tau})$ and $H^R_2(a, b) = \min_{\boldsymbol{\tau}} \mathrm{SSR}^R_{[a,b]}(\boldsymbol{\tau})$ be the unrestricted and restricted residual sum of squares based on a time regime $(a, b]$. Let h be the minimal permissible length of a time regime. Here, h is designated by the user.

B.4.1 Case 1: known number of change-points

In this subsection, we outline the dynamic programming algorithm used in estimating the change-point locations in the case where the number of change-points, m_0, is known. For known m_0, we compute (3.3) using the following steps.

Step 1. Compute and save $H_2^U(a, b)$ and $H_2^R(a, b)$ for time periods $(a, b]$ that satisfy $b-a \geq h$.

Step 2. Compute and save $H_1^U(1, T_1)$ and $H_1^R(1, T_1)$ by solving

$$\begin{cases} H_1^U(1, T_1) = \min_{a\in[h,T_1-h]} \left[H_2^U(0, a) + H_2^U(a, T_1)\right] \\ H_1^R(1, T_1) = \min_{a\in[h,T_1-h]} \left[H_2^R(0, a) + H_2^R(a, T_1)\right] \end{cases}$$

for all $T_1 \in [2h, T - (m_0 - 1)h]$.

Step 3. Sequentially compute and save

$$\begin{cases} H_1^U(r, T_r) = \min_{a\in[rh,T_r-h]} \left[H_1^U(r-1, a) + H_2^U(a, T_r)\right] \\ H_1^R(r, T_r) = \min_{a\in[rh,T_r-h]} \left[H_1^R(r-1, a) + H_2^R(a, T_r)\right], \end{cases}$$

for $r = 2, \ldots, m_0 - 1$ and $T_r \in [(r + 1)h, T - (m_0 - r)h]$.

Step 4. The estimated change points can then be obtained by solving

$$\begin{cases} H_1^U(m_0, T) = \min_{a\in[m_0h,T-h]} \left[H_1^U(m_0-1, a) + H_2^U(a, T)\right] \\ H_1^R(m_0, T) = \min_{a\in[m_0h,T-h]} \left[H_1^R(m_0-1, a) + H_2^R(a, T)\right], \end{cases}$$

and $H_1^U(m_0 - 1, a) = H_2^U(0, a)$, $H_1^R(m_0 - 1, a) = H_2^R(0, a)$, if $m_0 = 1$.

B.4.2 Case 2: unknown number of change-points

In this subsection, we outline the dynamic programming algorithm used in estimating the change-point locations in the case where the number of change-points, m_0, is also unknown. For change-point number m, we compute the estimates in (3.3) by using the following steps.

Step 1. Compute and save $H_2^U(a,b)$ and $H_2^R(a,b)$ for time periods $(a,b]$ that satisfy $b-a \geq h$.

Step 2. Compute and save $H_1^U(1,T_1)$ and $H_1^R(1,T_1)$ by solving

$$\begin{cases} H_1^U(1,T_1) = \min_{a\in[h,T_1-h]}\left[H_2^U(0,a) + H_2^U(a,T_1)\right] \\ H_1^R(1,T_1) = \min_{a\in[h,T_1-h]}\left[H_2^R(0,a) + H_2^R(a,T_1)\right] \end{cases}$$

for all $T_1 \in [2h, T-(m-1)h]$.

Step 3. Sequentially compute and save

$$\begin{cases} H_1^U(r,T_r) = \min_{a\in[rh,T_r-h]}\left[H_1^U(r-1,a) + H_2^U(a,T_r)\right] \\ H_1^R(r,T_r) = \min_{a\in[rh,T_r-h]}\left[H_1^R(r-1,a) + H_2^R(a,T_r)\right], \end{cases}$$

for $r = 2,\ldots,m-1$ and $T_r \in [(r+1)h, T-(m-r)h]$.

Step 4. The estimated change points can then be obtained by solving

$$\begin{cases} H_1^U(m,T) = \min_{a\in[mh,T-h]}\left[H_1^U(m-1,a) + H_2^U(a,T)\right] \\ H_1^R(m,T) = \min_{a\in[mh,T-h]}\left[H_1^R(m-1,a) + H_2^R(a,T)\right], \end{cases}$$

and $H_1^U(m-1,a) = H_2^U(0,a)$, $H_1^R(m-1,a) = H_2^R(0,a)$, if $m = 1$.

Step 5. Follow steps 1-4 to search for the optimal locations of the m estimated change-points and store the value of IC(m).

Step 6. Repeat the above steps 1-5 for $m = 0,1,\cdots,m_{\max}$.

The estimated number of change-points, $\hat{m}$, can then be obtained by taking the m with smallest IC(m) value. The user can set $m_{\max}$ such that $0 \leqslant m_{\max} \leqslant \lceil T/h \rceil$. $m_{\max}$ can also be determined by observing and analyzing the data or from available literature.

Appendix C

Generalized Tensor Regression

C.1 Definitions

Below, we define $\alpha-$mixing (strong-mixing), near-epoch dependence, Hausdorff and connected spaces and Stone topology.

Definition C.1.1 (α-mixing)**.** *Let* $\{\mathscr{F}_{n,s}^{t}, -\infty \leqslant s \leqslant t \leqslant \infty,\ n \geq 1\}$ *be a family of sigma subfields of* $\mathscr{F}$ *and let* $G \in \mathscr{F}_{n,-\infty}^{t}$ *and* $H \in \mathscr{F}_{n,t+m}^{\infty}$ *be events. Define* $\alpha_m = \sup |\mathrm{P}(G \cap H) - \mathrm{P}(G)\mathrm{P}(H)|$ *with the supremum being taken over events G and H and over t and n. Then, the family is said to be* $\alpha-$*mixing/strong mixing if* $\alpha_m \to 0$ *as* $m \to \infty$*. Moreover, the family of sigma subfields is said to be* α*/strong-mixing of size* $-\lambda_0$ *if* $\alpha_m = O(m^{-\lambda})$ *for* $\lambda > \lambda_0$*.*

Definition C.1.2 (near-epoch dependence)**.** $\{\boldsymbol{U}_{n,t}\}$ *is said to be near-epoch dependent in* $\mathcal{L}^p$ *norm of size* $-\lambda_0$ *if and only if*

$$||\boldsymbol{U}_{n,t} - \mathrm{E}(\boldsymbol{U}_{n,t}|\mathscr{F}_{n,t-j}^{t+j})||_p \leqslant d_{nt}\nu_j, \tag{C.1}$$

where the sigma subfields $\mathscr{F}_{n,t-j}^{t+j}$ *are defined as above,* $\{d_{nt}\}$ *is an array of positive constants and* $\nu_j = O(j^{-\lambda})$ *for* $\lambda > \lambda_0 > 0$*.*

As explained in Davidson (1992), there is no loss of generality in assuming $v_j \leqslant 1$.

Definition C.1.3 (Hausdorff space)**.** *A Hausdorff space is a topological space with a separation property. In other words, any two distinct points can be separated by disjoint open sets. That is, whenever x and y are distinct points of a set $\mathcal{X}$, there exist disjoint open sets U_x and U_y such that U_x contains x and U_y contains y.*

Definition C.1.4 (Connected space)**.** *A connected space is a topological space which cannot be written as a union of two-empty disjoint open sets. An example of a connected space is the set of real numbers. Conversely, a disconnected space is a topological space which can be written as a union of two empty disjoint open sets. An example of a disconnected space is the set of rational numbers and any discrete space.*

Definition C.1.5 (Stone topology/space)**.** *A Stone topology/space is a compact, totally disconnected Hausdorff space. Examples of a Stone space include finite discrete spaces and the Cantor set and any product of finite discrete spaces is also a Stone space.*

Definition C.1.6 (Completely monotonic function)**.** *A function f on $(0, +\infty)$ is completely monotonic if it the derivatives $f^{(n)}(x)$ exist for all $n = 0, 1, 2, \cdots$ and if $(-1)^n f^{(n)}(x) \geqslant 0$, for all $x > 0$.*

C.2 Some results and proofs

In this section, for the convenience of the reader we present some alternate proofs of the theorems in Chapter 4. Note that the results of Chapter 4 follow from the results of Chapter 3 by setting $m_0 = 0$, $T = n$, $s_1 = \prod_{i=1}^{d} p_i$, and $s_2 = s_3 = \cdots = s_{d+1} = 1$. Set $\boldsymbol{D}_{i,k} = \boldsymbol{Z}_{n,i} - \mathrm{E}[\boldsymbol{Z}_{n,i} | \mathscr{F}_{n,-\infty}^{i+k}]$ and set $D_{i,k,s}$ be the s^{th} element.

The following Corollary C.2.1 follows immediately from Corollary 3.2.3 by taking $m_0 = 0$,

$T = n$, $s_1 = \prod_{i=1}^{d} p_i$, and $s_2 = s_3 = \cdots = s_{d+1} = 1$. We provide an alternate prove of Corollary C.2.1 below.

Corollary C.2.1. *Suppose that conditions of Assumption 4.2.2 hold. Then, for* $s = 1, \cdots \prod_{i=1}^{d} p_i$, *we have* **1.** $\mathrm{E}\left(\sum_{i=1}^{n}(D_{i,k-1,s} - D_{i,k,s})^2\right) = \sum_{i=1}^{n} \mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})) - \sum_{i=1}^{n} \mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1}))$;
2. $\sum_{i=1}^{n}\sum_{j=1}^{i-1} \mathrm{E}\left[(D_{i,k-1,s} - D_{i,k,s})(D_{j,k-1,s} - D_{j,k,s})\right] = 0$;
3. $\sum_{i=1}^{n}\left[\mathrm{E}(D_{i,k-1,s}^2) - \mathrm{E}(D_{i,k,s}^2)\right] = \sum_{i=1}^{n}\left[\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})) - \mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1}))\right]$.

Proof. The proof is similar to that of Corollary 3.2.3. Below, we provide an alternative proof with more details. **1.** Expanding the left side, we have

$$
\begin{aligned}
&\mathrm{E}\left(\sum_{i=1}^{n}(D_{i,k-1,s} - D_{i,k,s})^2\right) \\
&= \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k}) + \mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1})] - 2\sum_{i=1}^{n} \mathrm{E}(\mathrm{E}Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})\mathrm{E}(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1})) \\
&= \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})] + \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1})] - 2\sum_{i=1}^{n} \mathrm{E}(\mathrm{E}(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})\mathrm{E}(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1})),
\end{aligned}
$$

then,

$$
\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{n}(D_{i,k-1,s} - D_{i,k,s})^2\right) &= \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})] + \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1})] \\
&- 2\sum_{i=1}^{n} \mathrm{E}[\mathrm{E}(\mathrm{E}Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})|\mathscr{F}_{n,-\infty}^{i+k-1})\mathrm{E}(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1})].
\end{aligned}
$$

Then, since $\mathrm{E}(\mathrm{E}(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})|\mathscr{F}_{n,-\infty}^{i+k-1}) = \mathrm{E}(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1})$ a.s., we have,

$$
\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{n}(D_{i,k-1,s} - D_{i,k,s})^2\right) &= \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})] + \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1})] \\
&\qquad - 2\sum_{i=1}^{n} \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1})]
\end{aligned}
$$

and then,

$$
\mathrm{E}\left(\sum_{i=1}^{n}(D_{i,k-1,s} - D_{i,k,s})^2\right) = \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})] - \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1})].
$$

2. Using the fact that $\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k}_{n,-\infty}) - \mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k-1}_{n,-\infty})$ is $\mathscr{F}_{j+k}$-measurable for all $j = 1, \cdots, i-1$ and $i = 1, \cdots, n$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{i-1} \mathrm{E}[((D_{i,k-1,s} - D_{i,k,s})(D_{j,k-1,s} - D_{j,k,s}))]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{i-1} \mathrm{E}[\mathrm{E}((D_{i,k-1,s} - D_{i,k,s})|\mathscr{F}_{j+k})(D_{j,k-1,s} - D_{j,k,s})].$$

The right side of this term above becomes
$\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{i-1} \mathrm{E}[(\mathrm{E}(Z_{n,i,k-1,s}|\mathscr{F}_{j+k}) - \mathrm{E}(Z_{n,i,k-1,s}|\mathscr{F}_{j+k}))(D_{j,k-1,s} - D_{j,k,s}))] = 0.$

3.

$$\begin{aligned}\sum_{i=1}^{n}[\mathrm{E}(D^2_{i,k-1,s}) - \mathrm{E}(D^2_{i,k,s})] &= \sum_{i=1}^{n}[\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}^{i+k-1}_{n,-\infty}) - 2\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k-1}_{n,-\infty})Z_{n,i,s} + Z^2_{n,i,s}) \\ &\quad - \mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}^{i+k}_{n,-\infty}) - 2\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k}_{n,-\infty})Z_{n,i,s} + Z^2_{n,i,s})] \\ &= \sum_{i=1}^{n}[\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}^{i+k-1}_{n,-\infty})) - 2\mathrm{E}(\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k-1}_{n,-\infty})Z_{n,i,s}) \\ &\quad - \mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}^{i+k}_{n,-\infty})) + 2\mathrm{E}(\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k}_{n,-\infty})Z_{n,i,s})].\end{aligned}$$

Note that, since $\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k}_{n,-\infty})$ is $\mathscr{F}^{i+k}_{n,-\infty}$-measurable,

$$\mathrm{E}(\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k}_{n,-\infty})Z_{n,i,s}) = \mathrm{E}[\mathrm{E}(\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k}_{n,-\infty})Z_{n,i,s})|\mathscr{F}^{i+k}_{n,-\infty}] = \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}^{i+k}_{n,-\infty})].$$

Similarly, since $\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k-1}_{n,-\infty})$ is $\mathscr{F}^{i+k-1}_{n,-\infty}$ -measurable, we have

$$\mathrm{E}(\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k-1}_{n,-\infty})Z_{n,i,s}) = \mathrm{E}[\mathrm{E}(\mathrm{E}(Z_{n,i,s}|\mathscr{F}^{i+k-1}_{n,-\infty})Z_{n,i,s})|\mathscr{F}^{i+k-1}_{n,-\infty}] = \mathrm{E}[\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}^{i+k-1}_{n,-\infty})].$$

Therefore,

$$\sum_{i=1}^{n}\left[\mathrm{E}(D^2_{i,k-1,s}) - \mathrm{E}(D^2_{i,k,s})\right] = \sum_{i=1}^{n}\left[\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}^{i+k}_{n,-\infty})) - \mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}^{i+k-1}_{n,-\infty}))\right],$$

this completes the proof. □

Using Corollary C.2.1, we establish the following result which follows immediately from Corollary 3.2.4 by taking $m_0 = 0$, $T = n$, $s_1 = \prod\limits_{i=1}^{d} p_i$, and $s_2 = s_3 = \cdots = s_{d+1} = 1$. We also provide an alternate proof of Corollary C.2.2 below.

Corollary C.2.2. *Let* $\{a_k\}_{-\infty}^{\infty}$ *be as in Lemma 3.2.2, Then, for* $s = 1, \cdots, \prod_{i=1}^{d} p_i$,

1. $\sum_{k=1}^{\infty}\sum_{i=1}^{n}(a_k^{-1} - a_{k-1}^{-1})\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k})) < \infty$; **2.** $\sum_{k=1}^{\infty}\sum_{i=1}^{n}(a_{k+1}^{-1} - a_k^{-1})\mathrm{E}(D_{i,k,s}^2) < \infty$;

3. $\sum_{k=1}^{\infty}\sum_{i=1}^{n}[a_k^{-1}\mathrm{E}(D_{i,k-1,s}^2) - a_{k+1}^{-1}\mathrm{E}(D_{i,k,s}^2)] = \sum_{i=1}^{n} a_1^{-1}\mathrm{E}(D_{i,0,s}^2) < \infty$;

4. $\sum_{k=1}^{\infty}\sum_{i=1}^{n}[a_{k-1}^{-1}\mathrm{E}(\mathrm{E}^2(Z_{ni,s}|\mathscr{F}_{n,-\infty}^{i-k})) - a_k^{-1}\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k-1}))] = \sum_{i=1}^{n} a_0^{-1}\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-1})) < \infty$.

Proof. Define $\boldsymbol{\kappa}^*(j) = \boldsymbol{\kappa}(j)\mathbb{I}\{\xi_j^{(2)} = O(j^{-1/2-\varsigma}\boldsymbol{\kappa}^{-1}(\boldsymbol{j}))\} + j^{\delta_0}\mathbb{I}\{\xi_j^{(2)} = O(j^{-1-\delta_0})\}$, for $\delta_0 > 0$. Then, $\boldsymbol{\kappa}^*$ satisfies the conditions of Lemma 3.2.2.

1. Since $a_j^{-1} - a_{j-1}^{-1} = O(\boldsymbol{\kappa}^*(j))$, we have $a_m^{-1} - a_0^{-1} = \sum_{j=1}^{m}\left(a_j^{-1} - a_{j-1}^{-1}\right)$, then $a_m^{-1} - a_0^{-1} \leqslant B_0\sum_{j=1}^{m}\boldsymbol{\kappa}^*(j)$, for some $0 < B_0 < \infty$. Then, we have $a_m^{-1}\zeta_{2,m}^2 \leqslant a_0^{-1}\zeta_{2,m}^2 + B_0\zeta_{2,m}^2\sum_{j=1}^{m}\boldsymbol{\kappa}^*(j)$. Since $\boldsymbol{\kappa}^*(j)$ is increasing, we get $a_m^{-1}\zeta_{2,m}^2 \leqslant a_0^{-1}\zeta_{2,m}^2 + B_0\zeta_{2,m}^2\sum_{k=1}^{m}\boldsymbol{\kappa}^*(m) = a_0^{-1}\zeta_{2,m}^2 + B_0\zeta_{2,m}^2 m\boldsymbol{\kappa}^*(m)$. Since $\zeta_{2,m}^2 = O\left(\frac{1}{m\boldsymbol{\kappa}^{*2}(m)}\right)$, we get $a_m^{-1}\zeta_{2,m}^2 \leqslant a_0^{-1}\zeta_{2,m}^2 + B_1\frac{1}{m\boldsymbol{\kappa}^{*2}(m)}m\boldsymbol{\kappa}^*(m) = a_0^{-1}\zeta_{2,m}^2 + B_1\frac{1}{\boldsymbol{\kappa}^*(m)}$, for some $0 < B_1 < \infty$. Hence, since $a_{m+1}^{-1}\zeta_{2,m+1}^2 \leqslant a_0^{-1}\zeta_{2,m+1}^2 + B_1\frac{1}{\boldsymbol{\kappa}^*(m+1)}$, then

$$0 \leqslant \lim_{m\to\infty} a_{m+1}^{-1}\zeta_{2,m+1}^2 \leqslant a_0^{-1}\lim_{m\to\infty}\zeta_{2,m+1}^2 + B_1\lim_{m\to\infty}\frac{1}{\boldsymbol{\kappa}^*(m+1)} = 0,$$

this proves Part **1**.

2. We have $\sum_{k=1}^{\infty}\sum_{i=1}^{n}\left|a_k^{-1} - a_{k-1}^{-1}\right|\mathrm{E}\left(\mathrm{E}^2\left(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k}\right)\right) \leqslant \sum_{i=1}^{n} c_{n,i}^2\sum_{k=1}^{\infty}\left|a_k^{-1} - a_{k-1}^{-1}\right|\zeta_{2,k}^2$. Thus, by taking $b_k = \zeta_{2,k}^2$ and by applying Lemma 3.2.2, we get the statement of Part **2**.

3. We have $\sum_{k=1}^{\infty}\sum_{i=1}^{n}\left|a_{k+1}^{-1} - a_k^{-1}\right|\mathrm{E}\left(D_{i+k,s}^2\right) \leqslant \sum_{i=1}^{n} c_{ni}^2\sum_{k=1}^{\infty}\left|a_{k+1}^{-1} - a_k^{-1}\right|\zeta_{2,k+1}^2$. Hence, by Lemma 3.2.2, $\sum_{k=1}^{\infty}\sum_{i=1}^{n}\left(a_{k+1}^{-1} - a_k^{-1}\right)\mathrm{E}\left(D_{i,k,s}^2\right) \leqslant \sum_{i=1}^{n} c_{ni}^2\sum_{k=2}^{\infty}\left(a_k^{-1} - a_{k-1}^{-1}\right)\zeta_{2,k}^2 < \infty$, this proves Part **3**.

4. We have $\sum_{k=1}^{\infty}\sum_{i=1}^{n}\left[a_k^{-1}\mathrm{E}\left(D_{i,k-1,s}^2\right) - a_{k+1}^{-1}\mathrm{E}\left(D_{i,k,s}^2\right)\right] = \sum_{i=1}^{n} a_1^{-1}\mathrm{E}\left(D_{i,0,s}^2\right) - \lim_{m\to\infty} a_{m+1}^{-1}\sum_{i=1}^{n}\mathrm{E}\left(D_{i,m,s}^2\right)$, and from Part **1**, $\lim_{m\to\infty} a_{m+1}^{-1}\sum_{i=1}^{n}\mathrm{E}\left(D_{i,m,s}^2\right) \leqslant \lim_{m\to\infty}\sum_{i=1}^{n} c_{n,i}^2 a_{m+1}^{-1}\zeta_{2,m+1}^2 = 0$. This proves Part **4**.

5. Note that

$$\sum_{k=1}^{\infty}\sum_{i=1}^{n}\left[a_{k-1}^{-1}\mathrm{E}\left(\mathrm{E}^2\left(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k}\right)\right) - a_k^{-1}\mathrm{E}\left(\mathrm{E}^2\left(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k-1}\right)\right)\right]$$
$$= \sum_{i=1}^{n} a_0^{-1}\mathrm{E}\left(\mathrm{E}^2\left(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-1}\right)\right) - \lim_{m\to\infty}\sum_{i=1}^{n} a_m^{-1}\mathrm{E}\left(\mathrm{E}^2\left(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-m-1}\right)\right).$$

By Part **1**, we get $\lim_{m\to\infty}\sum_{i=1}^{n} a_m^{-1}\mathrm{E}\left(\mathrm{E}^2\left(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-m-1}\right)\right) \leqslant \lim_{m\to\infty}\sum_{i=1}^{n} c_{n,i}^2 a_m^{-1}\zeta_{2,m+1}^2 = 0$. Therefore, $\sum_{k=1}^{\infty}\sum_{i=1}^{n}\left[a_{k-1}^{-1}\mathrm{E}\left(\mathrm{E}^2\left(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-1}\right)\right) - a_k^{-1}\mathrm{E}\left(\mathrm{E}^2\left(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k-1}\right)\right)\right] = \sum_{i=1}^{n} a_0^{-1}\mathrm{E}\left(\mathrm{E}^2\left(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-1}\right)\right) < \infty$, this completes the proof. □

Corollary C.2.3. *Suppose that the conditions of Corollary C.2.2 hold and let* $V_{n,k}(l) = \sum_{i=l+1}^{l+j}\left[\mathrm{E}(\boldsymbol{Z}_{n,i}|\mathscr{F}_{n,-\infty}^{i+k} - \mathrm{E}(\boldsymbol{Z}_{n,i}|\mathscr{F}_{n,-\infty}^{i+k-1})\right]$, $j = 1,\cdots,n, k = 1,2,\cdots$. *Then, for* $s = 1,\cdots,\prod_{i=1}^{d} p_i$,

1. $\sum_{k=1}^{\infty} a_k^{-1}\mathrm{E}(V_{n,k,s}^2(l)) = \sum_{i=l+1}^{l+n}\left(\sum_{k=1}^{\infty}(a_{k+1}^{-1} - a_k^{-1})\mathrm{E}(D_{i,k,s}^2) + a_1^{-1}\mathrm{E}(D_{i,0,s}^2)\right)$;

2. $\sum_{k=-\infty}^{-1} a_k^{-1}\mathrm{E}(V_{n,k,s}^2(l)) = \sum_{i=l+1}^{l+n}\left(\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\mathrm{E}(\mathrm{E}^2(Z_{n,i,}|\mathscr{F}_{n,-\infty}^{i-k})) + a_0^{-1}\mathrm{E}(\mathrm{E}(Z_{n,i,s}^2|\mathscr{F}_{n,-\infty}^{i-1}))\right)$.

Proof. This result follows directly from Corollary 3.2.5. Below, we provide an alternative proof with more details. Put $D_{i,k} = \boldsymbol{Z}_{n,i} - \mathrm{E}(\boldsymbol{Z}_{n,i}|\mathscr{F}_{n,-\infty}^{i+k})$ for $k \geq 0$ and note that for each $s = 1,\cdots,\prod_{i=1}^{d} p_i$,
$\mathrm{E}(V_{n,k,s}^2) = \mathrm{E}\left[\left(\sum_{i=1}^{n}(D_{i,k-1,s} - D_{i,k,s})\right)^2\right]$. This gives

$$\mathrm{E}(V_{n,k,s}^2) = \mathrm{E}\left[\sum_{i=1}^{n}(D_{i,k-1,s} - D_{i,k,s})^2\right] + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1}\mathrm{E}\left[(D_{i,k-1,s} - D_{i,k,s})(D_{j,k-1,s} - D_{j,k,s})\right].$$

Hence, it follows from Proposition C.2.1 that

$$\mathrm{E}(V_{n,k,s}^2) = \sum_{i=1}^{n}\left[\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k})) - \mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1}))\right] = \sum_{i=1}^{n}\left[\mathrm{E}(D_{i,k-1,s}^2) - \mathrm{E}(D_{i,k,s}^2)\right].$$

Using this, we have

$$a_0^{-1}\mathrm{E}(V_{n,0,s}^2) = \sum_{i=1}^{n} a_0^{-1}\left[\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i})) - \mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-1}))\right]$$

and

$$\sum_{k=1}^{\infty} a_k^{-1}\mathrm{E}(V_{n,k,s}^1) = \sum_{k=1}^{\infty} a_k^{-1}\left(\sum_{i=1}^{n}\left[\mathrm{E}(D_{i,k-1,s}^2) - \mathrm{E}(D_{i,k,s}^2)\right]\right).$$

Moreover, using the second and third statements in Proposition C.2.2, we have

$$\sum_{k=1}^{\infty}\left(\sum_{i=1}^{n}(a_{k+1}^{-1}-a_k^{-1})\mathrm{E}(D_{i,k,s}^2)\right)+\sum_{k=1}^{\infty}\sum_{i=1}^{n}(a_k^{-1}\mathrm{E}(D_{i,k-1,s}^2)-a_{k+1}^{-1}\mathrm{E}(D_{i,k,s}^2))$$
$$=\sum_{i=1}^{n}a_1^{-1}\mathrm{E}(D_{i,0,s}^2)+\sum_{k=1}^{\infty}a_k^{-1}\left(\sum_{i=1}^{n}\left[\mathrm{E}(D_{i,k-1,s}^2)-\mathrm{E}(D_{i,k,s}^2)\right]\right)<\infty.$$

Therefore,

$$\sum_{k=1}^{\infty}a_k^{-1}\mathrm{E}(V_{n,k,s}^2)=\sum_{i=1}^{n}\left(\sum_{k=1}^{\infty}(a_{k+1}^{-1}-a_k^{-1})\mathrm{E}(D_{i,k,s}^2)+a_1^{-1}\mathrm{E}(D_{i,0,s}^2)\right),$$

this proves the first statement. We prove the second statement by following similar steps and using the assumption that $a_k = a_{-k}$. Namely, we have

$$\sum_{k=-\infty}^{-1}a_k^{-1}\mathrm{E}(V_{n,k,s}^2)=\sum_{k=-\infty}^{-1}a_k^{-1}\sum_{i=1}^{n}\left[\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k}))-\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1}))\right]$$
$$=\sum_{k=1}^{\infty}\sum_{i=1}^{n}a_k^{-1}\left[\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k}))-\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i+k-1}))\right].$$

Now, using the first and fourth statements in Proposition C.2.2, we have

$$\sum_{k=1}^{\infty}\sum_{i=1}^{n}(a_k^{-1}-a_{k-1}^{-1})\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k}))+\sum_{k=1}^{\infty}\sum_{i=1}^{n}\left(a_{k-1}^{-1}\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k}))-a_k^{-1}\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k-1}))\right)$$
$$=\sum_{k=1}^{\infty}\sum_{i=1}^{n}a_k^{-1}\left(\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k}))-\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k-1}))\right)<\infty.$$

Hence,

$$\sum_{k=-\infty}^{-1}a_k^{-1}\mathrm{E}(V_{n,k,s}^2)=\sum_{i=1}^{n}\left(\sum_{k=1}^{\infty}(a_k^{-1}-a_{k-1}^{-1})\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-k}))+a_0^{-1}\mathrm{E}(\mathrm{E}^2(Z_{n,i,s}|\mathscr{F}_{n,-\infty}^{i-1}))\right),$$

which completes the proof. □

Once again this result follows from Corollary 3.2.5 by taking $m_0 = 0$, $T = n$, $s_1 = \prod_{i=1}^{d} p_i$, and $s_2 = s_3 = \cdots = s_{d+1} = 1$. An alternate proof of Corollary C.2.3 is also provided for convenience.

Proposition C.2.1. *Suppose $\{a_k\}_{k=-\infty}^{\infty}$ are as in Lemma 3.2.2 and $V_{n,k}$ is as in Corollary C.2.3. Then,*

$$\sum_{s=1}^{\prod_{i=1}^{d} p_i} \sum_{k=-\infty}^{\infty} a_k^{-1} \mathrm{E}\left(V_{n,k,s}^2(l)\right) \leqslant \left(\sum_{i=l+1}^{l+n} c_{n,i}^2\right)\left\{a_0^{-1}\left(\zeta_{2,0}^2 + \zeta_{2,1}^2\right) + 2\sum_{k=1}^{\infty}(a_k^{-1} - a_{k-1}^{-1})\zeta_{2,k}^2\right\} < \infty.$$

Proof. The proof is similar to that of Proposition 3.2.7. □

Lemma C.2.1. *Under Assumption 4.2.2, for $L = 1, 2, 3, \ldots$; $l = 0, 1, 2, \ldots$*

$$\sum_{s=1}^{\prod_{i=1}^{d} p_i} \mathrm{E}\left(\max_{j\leqslant L} (S_{l+j,s} - S_{l,s})^2\right) \leqslant K \sum_{i=l+1}^{l+L} c_{n,i}^2, \textit{for some } K > 0.$$

Proof. The proof is similar to that of Lemma 3.2.3. □

For the next proposition, let $\boldsymbol{Z}_{n,i}$ be as in (4.11) and define

$Z_{n,i,s}^a = Z_{n,i,s}\mathbb{I}[|Z_{n,i,s}| \leqslant ac_{n,i}]$, $\mathrm{E}^{i+m}Z_{n,i,s}^a = \mathrm{E}(Z_{n,i,s}^a|\mathscr{F}_{n,-\infty}^{i+m})$, $U_{1,i,s} = \mathrm{E}^{i+m}Z_{n,i,s}^a - \mathrm{E}^{i-m}Z_{n,i,s}^a$,

$U_{2,i,s} = Z_{n,i,s} - \mathrm{E}^{i+m}Z_{n,i,s} + \mathrm{E}^{i-m}Z_{n,i,s}$, $U_{3,i,s} = \mathrm{E}^{i+m}(Z_{n,i,s} - Z_{n,i,s}^a) - \mathrm{E}^{i-m}(Z_{n,i,s} - Z_{n,i,s}^a)$.

Also, let $v_j^2 = \sum_{i=1}^{j} c_{n,i}^2$, $\tilde{v}_j^2(k) = \sum_{i=k+1}^{k+j} c_{n,i}^2$, $k = 0, 1, \ldots, j$, $j = 1, 2, \ldots$, $\bar{U}_{t,j,s}(l) = \sum_{i=l+1}^{l+j} U_{t,i,s}$,

$t = 1, 2, 3$, $A(a, m) = \sum_{s=1}^{\prod_{i=1}^{d} p_i} \max_{j\leqslant L} \frac{\bar{U}_{1,j,s}^2}{\tilde{v}_L^2(l)}$, $B(a, m) = \sum_{s=1}^{\prod_{i=1}^{d} p_i} \max_{j\leqslant L} \frac{\bar{U}_{2,j,s}^2}{\tilde{v}_L^2(l)}$, $C(a, m) = \sum_{s=1}^{\prod_{i=1}^{d} p_i} \max_{j\leqslant L} \frac{\bar{U}_{3,j,s}^2}{\tilde{v}_L^2(l)}$.

Proposition C.2.2. *Suppose that the conditions of Lemma C.2.1 hold. Then,*

1. *For fixed (m, a) and for any $\epsilon > 0$, one can choose a, b such that*

$\mathcal{J}_1(a, b, m) = \mathrm{E}[A(a, m)\mathbb{I}(A(a, m) > b/9)] < \epsilon$.

2. *For any $\epsilon > 0$, one can choose m such that $\mathcal{J}_2(m) = \mathrm{E}(B(m)) < \epsilon$.*

3. *For a fixed m, for any $\epsilon > 0$, one can choose a such that $\mathcal{J}_3(a, m) = \mathrm{E}(C(a, m)) < \epsilon$.*

Proof. The proof is similar to that of Proposition 3.2.8. □

Proposition C.2.3. *Let $\mathscr{F}_i^*$ be the $\sigma - field$ generated by $\{U_{ib_n}, U_{ib_n-1}, \cdots\}$ with U_i a random variable defined on $(\Omega, \mathscr{F}, P)$ such that $\mathscr{F}_{i-1}^* \subseteq \mathscr{F}_{n,-\infty}^{i-j}$. Then, under Assumption 4.2.2,*

$$\sum_{i=1}^{r_n} \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_{i-1}^*) \xrightarrow[n\to\infty]{\mathrm{P}} 0 \textit{ and } \sum_{i=1}^{r_n}(\boldsymbol{V}_{n,i} - \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_{i-1}^*)) \xrightarrow[n\to\infty]{\mathrm{P}} 0.$$

Proof. The proof is similar to that of Proposition 3.2.11. □

Proposition C.2.4. *Suppose that the conditions of Proposition C.2.3 hold and let*

$\bar{\zeta}_{1,i,n} = (\boldsymbol{V}_i - \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_i^*) + \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_{i-1}^*))(\boldsymbol{V}_{n,i} + \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_i^*) - \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_{i-1}^*))'$,

$\bar{\zeta}_{2,i,n} = \boldsymbol{V}_{n,i}(\mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_i^*))' - \boldsymbol{V}_{n,i}(\mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_{i-1}^*))' - (\mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_i^*))\boldsymbol{V}_{n,i}' + (\mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_{i-1}^*))\boldsymbol{V}_{n,i}'$. *Then,*

$$\sum_{i=1}^{r_n} \|\bar{\zeta}_{1,i,n}\|_1 = o(1) \text{ and } \sum_{i=1}^{r_n} \|\bar{\zeta}_{2,i,n}\|_1 = o(1). \tag{C.2}$$

Proof. The proof follows from Proposition 3.2.13. □

Proposition C.2.5. *Suppose that the conditions of Proposition C.2.3 hold. Then,*

$$\sum_{i=1}^{r_n} [\boldsymbol{V}_{n,i}\boldsymbol{V}_{n,i}' - \boldsymbol{W}_{n,i}\boldsymbol{W}_{n,i}'] \xrightarrow[n\to\infty]{\mathrm{P}} 0; \tag{C.3}$$

$$\sum_{s=1}^{\prod_{i=1}^{d} p_i} \mathrm{E}\left[(W_{n,i,s})^2 \mathbb{I}\left(\|\boldsymbol{W}_{n,i}\|_{\mathrm{F}}^2 \geqslant \epsilon\right)\right] \xrightarrow[n\to\infty]{\mathrm{P}} 0 \text{ for all } \epsilon > 0; \text{ and } \sum_{i=1}^{r_n} \boldsymbol{W}_{n,i}\boldsymbol{W}_{n,i}' \xrightarrow[n\to\infty]{\mathrm{P}} \boldsymbol{\Phi}^*. \tag{C.4}$$

Proof. We have $\left\|\sum_{i=1}^{r_n} [\boldsymbol{V}_{n,i}\boldsymbol{V}_{n,i}' - \boldsymbol{W}_{n,i}\boldsymbol{W}_{n,i}']\right\|_1 \leqslant \sum_{i=1}^{r_n} \left\|[\boldsymbol{V}_{n,i}\boldsymbol{V}_{n,i}' - \boldsymbol{W}_{n,i}\boldsymbol{W}_{n,i}']\right\|_1$. Using the same techniques as in Proposition C.2.4, we have

$$\begin{aligned}
&\|\boldsymbol{V}_{n,i}\boldsymbol{V}_{n,i}' - \boldsymbol{W}_{n,i}\boldsymbol{W}_{n,i}'\|_1 \\
&\leqslant \|(\boldsymbol{V}_{n,i} - \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_i^*) + \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_{i-1}^*))(\boldsymbol{V}_{n,i} + \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_i^*) - \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_{i-1}^*))' - \|\bar{\zeta}_{2,i,n}\|_1.
\end{aligned}$$

So, we have $\|\boldsymbol{V}_{n,i}\boldsymbol{V}_{n,i}' - \boldsymbol{W}_{n,i}\boldsymbol{W}_{n,i}'\|_1 \leqslant \|\bar{\zeta}_{1,i,n}\|_1 + \|\bar{\zeta}_{2,i,n}\|_1$. Then, the proof of (C.3) follows from Proposition C.2.4. To prove the first statement of (C.4), note that $\{\mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_i^*) - \mathrm{E}(\boldsymbol{V}_{n,i}|\mathscr{F}_{i-1}^*), \mathscr{F}_i^*\}$ is an $\mathcal{L}^2-$ mixingale array of size $-1/2$ with mixingale magnitude indices $2c_{n,i}$. Also, let $\tilde{v}_i^2 = \sum_{t=(i-1)b_n+l_n+1}^{ib_n} c_{n,t}^2$, then,

$$\sum_{i=1}^{r_n}\sum_{s=1}^{\prod_{i=1}^d p_i} \mathrm{E}\left(W_{n,i,s}^2 \mathbb{I}\left(\sum_{s=1}^{\prod_{i=1}^d p_i} W_{n,i,s}^2 > \epsilon\right)\right) = \sum_{i=1}^{r_n}\sum_{s=1}^{\prod_{i=1}^d p_i} \mathrm{E}\left(W_{n,i,s}^2/\tilde{v}_i^2 \mathbb{I}\left(\|\boldsymbol{W}_{n,i}\|_{\mathrm{F}}^2/\tilde{v}_i^2 > \epsilon/\tilde{v}_i^2\right)\right)\tilde{v}_i^2$$

$$\leqslant \sum_{i=1}^{r_n}\sum_{s=1}^{\prod_{i=1}^d p_i} \mathrm{E}\left(W_{n,i,s}^2/\tilde{v}_i^2 \mathbb{I}\left(\|\boldsymbol{W}_{n,i}\|_{\mathrm{F}}^2/\tilde{v}_i^2 > \frac{\epsilon}{\max_{1\leqslant i\leqslant r_n}\tilde{v}_i^2}\right)\right)\tilde{v}_i^2$$

$$\leqslant \sum_{i=1}^{r_n}\sum_{s=1}^{\prod_{i=1}^d p_i} \mathrm{E}\left(W_{n,i,s}^2/\tilde{v}_i^2 \mathbb{I}\left(\max_{1\leqslant i\leqslant r_n}\|\boldsymbol{W}_{n,i}\|_{\mathrm{F}}^2/\tilde{v}_i^2 > \frac{\epsilon}{\max_{1\leqslant i\leqslant r_n}\tilde{v}_i^2}\right)\right)\tilde{v}_i^2.$$

So we have,

$$\sum_{i=1}^{r_n}\sum_{s=1}^{\prod_{i=1}^d p_i} \mathrm{E}\left(W_{n,i,s}^2 \mathbb{I}\left(\|\boldsymbol{W}_{n,i}\|_{\mathrm{F}}^2 > \epsilon\right)\right) \leqslant \max_{1\leqslant i\leqslant r_n}\left[\sum_{s=1}^{\prod_{i=1}^d p_i} \mathrm{E}\left(W_{n,i,s}^2/\tilde{v}_i^2 \mathbb{I}\left(\max_{1\leqslant i\leqslant r_n}\|\boldsymbol{W}_{n,i}\|_{\mathrm{F}}^2/\tilde{v}_i^2 > \frac{\epsilon}{\max_{1\leqslant i\leqslant r_n}\tilde{v}_i^2}\right)\right)\right]\sum_{i=1}^{r_n}\tilde{v}_i^2.$$

From Assumption 4.2.2, we have $\sum_{i=1}^{r_n}\tilde{v}_i^2 = \sum_{i=1}^{r_n}\sum_{t=(i-1)b_n+l_n+1}^{ib_n} c_{n,t}^2 \leqslant \sum_{i=1}^{r_n} b_n\left(\max_{(i-1)b_n+1\leqslant t\leqslant ib_n} c_{n,t}\right)^2 = O(n^{-\alpha})$. Then $\sum_{i=1}^{r_n}\tilde{v}_i^2 = o(1)$. Hence,

$$\sum_{i=1}^{r_n}\sum_{s=1}^{\prod_{i=1}^d p_i} \mathrm{E}\left(W_{n,i,s}^2 \mathbb{I}\left(\sum_{s=1}^{\prod_{i=1}^d p_i} W_{n,i,s}^2 > \epsilon\right)\right)$$

$$\leqslant \max_{1\leqslant i\leqslant r_n}\left[\sum_{s=1}^{\prod_{i=1}^d p_i} \mathrm{E}\left(W_{n,i,s}^2/\tilde{v}_i^2 \mathbb{I}\left(\max_{1\leqslant i\leqslant r_n}\|\boldsymbol{W}_{n,i}\|_{\mathrm{F}}^2/\tilde{v}_i^2 > \frac{\epsilon}{\max_{1\leqslant i\leqslant r_n}\tilde{v}_i^2}\right)\right)\right]\sum_{i=1}^{r_n}\left(\max_{(i-1)b_n+1\leqslant t\leqslant ib_n} c_{n,t}\right)^2 b_n$$

$$= O\left(\max_{1\leqslant i\leqslant r_n}\left[\sum_{s=1}^{\prod_{i=1}^d p_i} \mathrm{E}\left(W_{n,i,s}^2/\tilde{v}_i^2 \mathbb{I}\left(\max_{1\leqslant i\leqslant r_n}\|\boldsymbol{W}_{n,i}\|_{\mathrm{F}}^2/\tilde{v}_i^2 > \frac{\epsilon}{\max_{1\leqslant i\leqslant r_n}\tilde{v}_i^2}\right)\right)\right]\right).$$ It follows from Lemma 4.2.2 that $\sum_{s=1}^{\prod_{i=1}^d p_i} W_{n,i,s}^2/\tilde{v}_i^2$ is uniformly integrable and $\lim_{n\to\infty}\max_{1\leqslant i\leqslant r_n}\tilde{v}_i = 0$ which implies that the last term above converges to 0 for any $\epsilon > 0$. This completes the proof of the first claim of (C.4). To prove the second statement of (C.4), we get $\sum_{i=1}^{r_n}\boldsymbol{W}_{n,i}\boldsymbol{W}_{n,i}' = \sum_{i=1}^{r_n}\boldsymbol{V}_{n,i}\boldsymbol{V}_{n,i}' -$

$\sum_{i=1}^{r_n}\left[\boldsymbol{V}_{n,i}\boldsymbol{V}'_{n,i}-\boldsymbol{W}_{n,i}\boldsymbol{W}'_{n,i}\right]$. Hence, by combining Assumption 4.2.2, Proposition C.2.5 and Slutsky's theorem, we have $\sum_{i=1}^{r_n}\boldsymbol{W}_{n,i}\boldsymbol{W}'_{n,i}\xrightarrow[n\to\infty]{\mathrm{P}}\boldsymbol{\Phi}^*$. □

C.3 Derivation of ADR[1] for the elliptically contoured distribution

In this subsection, we outline the derivation of the asymptotic distributional risk function of the proposed class of shrinkage estimators under quadratic loss function. Let $\hat{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\theta}}$ be estimators given in (4.48) and let $\hat{\boldsymbol{\vartheta}}\left(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}\right)$ be as in (3.13). Let

$$\hat{\boldsymbol{\theta}}^{s}=\tilde{\boldsymbol{\theta}}+\left[1-\left(\prod_{j=1}^{d}l_i-2\right)\Big/\left(n\|\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}\|^2_{\{\hat{\Xi}_i,i=1,\cdots,d\}}\right)\right]\left(\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}\right), \tag{C.5}$$

$$\hat{\boldsymbol{\theta}}^{sp}=\tilde{\boldsymbol{\theta}}+\max\left\{\left[1-\left(\prod_{j=1}^{d}l_i-2\right)\Big/\left(n\|\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}\|^2_{\{\hat{\Xi}_i,i=1,\cdots,d\}}\right)\right],0\right\}\left(\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}\right),$$

where $\hat{\Xi}_i$, $i=1,2,\ldots,d$ are consistent estimators for $\Xi_i=\boldsymbol{H}'_i(\boldsymbol{H}_i\Lambda_i\boldsymbol{H}'_i)^{-1}\boldsymbol{H}_i$ with $\Lambda_1,\Lambda_2,\ldots,\Lambda_d$ be positive definite matrices, $\boldsymbol{H}_i$ a full rank $l_i\times p_i$-matrix, $l_i\leqslant p_i$, $i=1,2,\ldots,d$. Let $\bar{\Lambda}_1,\bar{\Lambda}_2,\ldots,\bar{\Lambda}_d$ be positive definite matrices, let $\boldsymbol{J}_i=\bar{\Lambda}_i\boldsymbol{H}'_i(\boldsymbol{H}_i\bar{\Lambda}_i\boldsymbol{H}'_i)^{-1}$, $i=1,2,\ldots,d$, $\boldsymbol{J}_0=\bigotimes_{j=d}^{1}\boldsymbol{J}_i\boldsymbol{H}_i$. Let $\mathrm{Vec}(\boldsymbol{\vartheta}_2)=(\boldsymbol{I}_{p_1p_2\cdots p_d}-\boldsymbol{J}_0)\mathrm{Vec}(\boldsymbol{\vartheta}_1)+\mathrm{Vec}(\delta)$ and $\mathrm{Vec}(\boldsymbol{\vartheta}_3)=\boldsymbol{J}_0\mathrm{Vec}(\boldsymbol{\vartheta}_1)-\mathrm{Vec}(\delta)$, $\boldsymbol{\delta}^*=\boldsymbol{\delta}(\bigtimes_{i=1}^{d})_i\boldsymbol{W}_i^{1/2}$, $\Delta=\mathrm{trace}\left(\left(\boldsymbol{\delta}(\bigtimes_{i=1}^{d})_i\Xi_i^{1/2}\right)'_{(d)}\left(\boldsymbol{\delta}(\bigtimes_{i=1}^{d})_i\Xi_i^{1/2}\right)_{(d)}\right)$.

As in Section 4.3, suppose that

$$\sqrt{n}\left(\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\boxplus_{(d+1)}\left(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\boxplus_{(d+1)}\left(\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}\right)\right)\xrightarrow[n\to\infty]{d}\boldsymbol{\vartheta}_1\boxplus_{(d+1)}\boldsymbol{\vartheta}_2\boxplus_{(d+1)}\boldsymbol{\vartheta}_3$$

where $\boldsymbol{\vartheta}_1$ satisfies the following condition.

Assumption C.3.1. *We assume* $\boldsymbol{\vartheta}_1\sim\mathscr{E}_{p_1\times\cdots\times p_d}\left(0,\bigotimes_{i=d}^{1}\Lambda_i;g\right)$ *where* Λ_i, $i=1,2,\ldots,d$ *are positive definite matrices.*

Under Assumption C.3.1, we have

$$\boldsymbol{\vartheta}_1 \boxplus_{(d+1)} \boldsymbol{\vartheta}_2 \boxplus_{(d+1)} \boldsymbol{\vartheta}_3 \sim \mathscr{E}^{l_1\times\cdots\times l_d}_{p_1\times\cdots\times 3p_d}\left(0 \boxplus_{(d+1)} \boldsymbol{\delta} \boxplus_{(d+1)} -\boldsymbol{\delta},\ \boldsymbol{\Sigma}^*; g\right) \text{ with } \boldsymbol{\Sigma}^* = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{pmatrix},$$

with $\boldsymbol{\Sigma}_{11} = \bigotimes_{i=d}^{1} \boldsymbol{\Lambda}_i$, $\boldsymbol{\Sigma}_{12} = \bigotimes_{i=d}^{1} \boldsymbol{\Lambda}_j - \bigotimes_{i=d}^{1} \boldsymbol{\Lambda}_i \boldsymbol{H}_i' \boldsymbol{J}_i'$, $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}'$, $\boldsymbol{\Sigma}_{31}' = \boldsymbol{\Sigma}_{13} = \bigotimes_{i=d}^{1} \boldsymbol{\Lambda}_i \boldsymbol{H}_i' \boldsymbol{J}_i'$, $\boldsymbol{\Sigma}_{23} = \bigotimes_{i=d}^{1} \boldsymbol{\Lambda}_i \boldsymbol{H}_i' \boldsymbol{J}_i' - \bigotimes_{i=d}^{1} \boldsymbol{J}_i \boldsymbol{H}_i \boldsymbol{\Lambda}_i \boldsymbol{H}_i' \boldsymbol{J}_i'$, $\boldsymbol{\Sigma}_{33} = \bigotimes_{i=d}^{1} \boldsymbol{J}_i \boldsymbol{H}_i \boldsymbol{\Lambda}_i \boldsymbol{H}_i' \boldsymbol{J}_i'$, $\boldsymbol{\Sigma}_{32} = \boldsymbol{\Sigma}_{23}'$, $\boldsymbol{\Sigma}_{22} = \bigotimes_{i=d}^{1} \boldsymbol{\Lambda}_i - \bigotimes_{i=d}^{1} \boldsymbol{J}_i \boldsymbol{H}_i \boldsymbol{\Lambda}_i - \bigotimes_{i=d}^{1} \boldsymbol{\Lambda}_i \boldsymbol{H}_i' \boldsymbol{J}_i' + \bigotimes_{i=d}^{1} \boldsymbol{J}_i \boldsymbol{H}_i \boldsymbol{\Lambda}_i \boldsymbol{H}_i' \boldsymbol{J}_i'$.

Below, we derive the ADR of $\hat{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}^s$, $\hat{\boldsymbol{\theta}}^{sp}$ under quadratic loss. To this end, let $\chi^2_n(\lambda)$ denote a chi-square random variable with n degrees of freedom and non-centrality parameter λ, let $\omega(t)$, $t \geqslant 0$ be the weight function associated to the elliptically contoured distribution in Assumption C.3.1, let

$$\psi^{(1)}_{i,n}(x) = \int_0^\infty \mathrm{E}[h^i(t^{-1}\chi^2_n(tx))]\omega(t)dt, \quad \psi^{(2)}_{i,n}(x) = \int_0^\infty t^{-1}\mathrm{E}[h^i(t^{-1}\chi^2_n(tx))]\omega(t)dt,\ x \geq 0, \tag{C.6}$$

$c = \psi^{(2)}_{0,1}(x) = \psi^{(2)}_{0,n}(x)$, $x \geq 0$, and let

$$\boldsymbol{\vartheta}(h) = \boldsymbol{\vartheta}_2 + h\left(\|\boldsymbol{\vartheta}_2 - \boldsymbol{\vartheta}_1\|^2_{\Xi_i, i=1,2,\ldots,d}\right)(\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_2), \quad \text{and} \quad \boldsymbol{\vartheta}^*(h) = \boldsymbol{\vartheta}(h)\Big(\bigtimes_{i=1}^{d}\Big)_i \boldsymbol{W}_i^{1/2}, \tag{C.7}$$

where $\boldsymbol{W}_i$, $i = 1, 2, \ldots, d$ are nonnegative definite matrices. Below, we establish a proposition which gives the ADR of the UE and the RE.

Proposition C.3.1. *Under Assumption C.3.1, we have*

$$\mathrm{ADR}^1\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right) = c \prod_{i=1}^{d} \mathrm{trace}(\boldsymbol{W}_i \boldsymbol{\Lambda}_i); \tag{C.8}$$

$$\begin{aligned}\mathrm{ADR}^1\left(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right) = c \prod_{i=1}^{d} \mathrm{trace}(\boldsymbol{W}_i \boldsymbol{\Lambda}_i) - 2c \prod_{i=1}^{d} \mathrm{trace}(\boldsymbol{W}_i \boldsymbol{\Upsilon}_i^*) + c \prod_{i=1}^{d} \mathrm{trace}(\mathbb{W}_i \boldsymbol{\Upsilon}_i^* \boldsymbol{H}_i' \boldsymbol{J}_i') \\ + \mathrm{trace}\left(\boldsymbol{\delta}^{*\prime}_{(d)} \boldsymbol{\delta}^*_{(d)}\right).\end{aligned}$$

Proof. Let $\boldsymbol{\vartheta}_1^* = \boldsymbol{\vartheta}_1(\bigtimes_{i=1}^{d})_i \boldsymbol{W}_i^{1/2}$ and $\boldsymbol{\vartheta}_2^* = \boldsymbol{\vartheta}_2(\bigtimes_{i=1}^{d})_i \boldsymbol{W}_i^{1/2}$. From Assumption C.3.1, we have $\mathrm{E}\left[\mathrm{trace}\left(\boldsymbol{\vartheta}^{*\prime}_{1_{(d)}} \boldsymbol{\vartheta}^*_{1_{(d)}}\right)\right] = c \prod_{i=1}^{d} \mathrm{trace}(\boldsymbol{W}_i \boldsymbol{\Lambda}_i)$; with $c = \psi^{(2)}_{0,1}(x) = \psi^{(2)}_{0,n}(x)$, $x \geq 0$, this proves the first statement of the proposition. Further, let $\mathbb{N}$ denote a $q_1 \times \cdots \times q_d$ such that $\mathbb{N} \sim \mathcal{N}_{q_1\times\cdots\times q_d}\left(\boldsymbol{\delta}, z^{-1}\left(\bigotimes_{i=d}^{1} \boldsymbol{\Lambda}_i - \bigotimes_{i=d}^{1} \boldsymbol{J}_i\boldsymbol{H}_i\boldsymbol{\Lambda}_i - \bigotimes_{i=d}^{1} \boldsymbol{\Lambda}_i\boldsymbol{H}_i'\boldsymbol{J}_i' + \bigotimes_{i=d}^{1} \boldsymbol{J}_i\boldsymbol{H}_i\boldsymbol{\Lambda}_i\boldsymbol{H}_i'\boldsymbol{J}_i'\right)\right)$ and let $\mathbb{N}^* = \mathbb{N}(\bigtimes_{i=1}^{d})_i \boldsymbol{W}_i^{1/2}$. We have

$$\mathrm{E}\left[\operatorname{trace}\left(\boldsymbol{\vartheta}^{*\prime}_{2_{(d)}}\boldsymbol{\vartheta}^{*}_{2_{(d)}}\right)\right] = \int_0^\infty \mathrm{E}_z\left[\operatorname{trace}\left(\mathbb{N}^{*\prime}_{(d)}\mathbb{N}^{*}_{(d)}\right)\right]\omega(z)dz, \tag{C.9}$$

where

$$\mathrm{E}_z\left[\operatorname{trace}\left(\mathbb{N}^{*\prime}_{(d)}\mathbb{N}^{*}_{(d)}\right)\right] = \operatorname{trace}\left(\left(\bigotimes_{i=d}^{1}\boldsymbol{W}_i\right)z^{-1}\left(\bigotimes_{i=d}^{1}\boldsymbol{\Lambda}_i - \bigotimes_{i=d}^{1}\boldsymbol{J}_i\boldsymbol{H}_i\boldsymbol{\Lambda}_i - \bigotimes_{i=d}^{1}\boldsymbol{\Lambda}_i\boldsymbol{H}_i'\boldsymbol{J}_i'\right)\right)$$
$$+\operatorname{trace}\left(\left(\bigotimes_{i=d}^{1}\boldsymbol{W}_i\right)z^{-1}\left(\bigotimes_{i=d}^{1}\boldsymbol{J}_i\boldsymbol{H}_i\boldsymbol{\Lambda}_i\boldsymbol{H}_i'\boldsymbol{J}_i'\right)\right) + \operatorname{trace}\left((-\boldsymbol{\delta}^{*\prime}_{(d)})(-\boldsymbol{\delta}^{*}_{(d)})\right).$$

Then,

$$\mathrm{E}_z\left[\operatorname{trace}\left(\mathbb{N}^{*\prime}_{(d)}\mathbb{N}^{*}_{(d)}\right)\right] = z^{-1}\left(\prod_{i=1}^{d}\operatorname{trace}(\boldsymbol{W}_i\boldsymbol{\Lambda}_i) - 2\prod_{i=1}^{d}\operatorname{trace}(\boldsymbol{W}_i\boldsymbol{\Upsilon}_i^*)\right)$$
$$+z^{-1}\prod_{i=1}^{d}\operatorname{trace}(\mathbb{W}_i\boldsymbol{\Upsilon}_i^*\boldsymbol{H}_i'\boldsymbol{J}_i') + \operatorname{trace}\left(\boldsymbol{\delta}^{*\prime}_{(d)}\boldsymbol{\delta}^{*}_{(d)}\right).$$

Therefore, together with (C.9) we get the result stated. □

More generally, the following theorem gives $\mathrm{ADR}^1\left(\hat{\boldsymbol{\vartheta}}\left(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}\right),\boldsymbol{\theta};\boldsymbol{W}\right)$.

Theorem C.3.1. *Under Assumption C.3.1,*

$$\mathrm{ADR}^1\left(\hat{\boldsymbol{\vartheta}}\left(h,\hat{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}\right),\boldsymbol{\theta};\boldsymbol{W}\right) = \mathrm{ADR}^1\left(\tilde{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W}\right) + \psi^{(2)}_{2,l+2}(\Delta)\prod_{i=1}^{d}\operatorname{trace}(\boldsymbol{W}_i\boldsymbol{\Upsilon}_i^*\boldsymbol{H}_i'\boldsymbol{J}_i')$$
$$-\,2\,\psi^{(1)}_{1,l+2}(\Delta)\operatorname{trace}\left(\boldsymbol{\delta}'_{(d)}\left(\boldsymbol{\delta}(\bigtimes_{j=1}^{d})_j\,\boldsymbol{\Xi}_j\boldsymbol{\Upsilon}_j^*\boldsymbol{W}_j\right)_{(d)}\right) + 2\,\psi^{(2)}_{1,l+2}(\Delta)\prod_{j=1}^{d}\operatorname{trace}(\boldsymbol{W}_j\boldsymbol{\Upsilon}_j^*)$$
$$-\,2\,\psi^{(1)}_{1,l+4}(\Delta)\operatorname{trace}\left(\boldsymbol{\delta}^{*\prime}_{(d)}\boldsymbol{\delta}^{*}_{(d)}\right) - 2\,\psi^{(2)}_{1,l+2}(\Delta)\prod_{j=1}^{d}\operatorname{trace}(\boldsymbol{W}_j\boldsymbol{\Upsilon}_i^*\boldsymbol{H}_i'\boldsymbol{J}_i')$$
$$+\,2\,\psi^{(1)}_{1,l+4}(\Delta)\operatorname{trace}\left(\boldsymbol{\delta}'_{(d)}\left(\boldsymbol{\delta}(\bigtimes_{j=1}^{d})_j\,\boldsymbol{\Xi}_j\boldsymbol{\Upsilon}_j^*\boldsymbol{W}_j\right)_{(d)}\right) + \psi^{(1)}_{2,l+4}(\Delta)\operatorname{trace}(\boldsymbol{\delta}^{*\prime}_{(d)}\boldsymbol{\delta}^{*}_{(d)}).$$

Proof. Let $\boldsymbol{\vartheta}_3^* = \boldsymbol{\vartheta}_3(\bigotimes_{i=1}^{d})_i \boldsymbol{W}_{\boldsymbol{i}}^{1/2}$. We have

$$\mathrm{E}\left[\mathrm{trace}\left(\boldsymbol{\vartheta}^{*'}(h)_{(d)}\boldsymbol{\vartheta}^{*}(h)_{(d)}\right)\right] = \mathrm{E}\left[\mathrm{trace}\left(\boldsymbol{\vartheta}^{*'}_{2_{(d)}}\boldsymbol{\vartheta}^{*}_{2_{(d)}}\right)\right] + 2\mathrm{E}\left[h(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi_i, i=1,\cdots,d\}})\mathrm{trace}(\boldsymbol{\vartheta}^{*'}_{3_{(d)}}\boldsymbol{\vartheta}^{*}_{2_{(d)}})\right] + \mathrm{E}\left[h^2(\|\boldsymbol{\vartheta}_3\|^2_{\{\Xi_i, i=1,\cdots,d\}})\mathrm{trace}(\boldsymbol{\vartheta}^{*'}_{3_{(d)}}\boldsymbol{\vartheta}^{*}_{3_{(d)}})\right].$$

As established in Proposition C.3.1, we have $\mathrm{E}\left[\mathrm{trace}\left(\boldsymbol{\vartheta}^{*'}_{2_{(d)}}\boldsymbol{\vartheta}^{*}_{2_{(d)}}\right)\right] = \mathrm{ADR}^1\left(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right)$. Further, by using Theorem 2.2.2 and Theorem 2.2.3 of Chapter 2, along with some algebraic computations, we get

$$\begin{aligned}
&\mathrm{E}\left[h\left(\|\boldsymbol{\xi}_3\|^2_{\{\Xi_i, i=1,\cdots,d\}}\right)\mathrm{trace}\left(\boldsymbol{\xi}^{*'}_{3_{(d)}}\boldsymbol{\xi}^{*}_{2_{(d)}}\right)\right] = \psi^{(1)}_{1,l+2}(\Delta)\,\mathrm{trace}\left(-\boldsymbol{\delta}^{*'}_{(d)}\boldsymbol{\delta}^{*}_{(d)}\right)\\
&+\psi^{(2)}_{1,l+2}(\Delta)\prod_{j=1}^{d}\mathrm{trace}\left(\boldsymbol{W}_j\boldsymbol{\Upsilon}^*_j\right) - \psi^{(1)}_{1,l+2}(\Delta)\,\mathrm{trace}\left(\boldsymbol{\delta}'_{(d)}\left(\boldsymbol{\delta}\Big(\bigotimes_{j=1}^{d}\Big)_j \Xi_j\boldsymbol{\Upsilon}^*_j\boldsymbol{W}_j\right)_{(d)}\right)\\
&+\psi^{(1)}_{1,l+2}(\Delta)\,\mathrm{trace}\left(\boldsymbol{\delta}^{*'}_{(d)}\boldsymbol{\delta}^{*}_{(d)}\right) - \psi^{(1)}_{1,l+4}(\Delta)\,\mathrm{trace}\left(\boldsymbol{\delta}^{*'}_{(d)}\boldsymbol{\delta}^{*}_{(d)}\right)\\
&+\psi^{(1)}_{1,l+4}(\Delta)\,\mathrm{trace}\left(\boldsymbol{\delta}'_{(d)}\left(\boldsymbol{\delta}\Big(\bigotimes_{j=1}^{d}\Big)_j \Xi_j\boldsymbol{\Upsilon}^*_j\boldsymbol{W}_j\right)_{(d)}\right) - \psi^{(2)}_{1,l+2}(\Delta)\prod_{j=1}^{d}\mathrm{trace}\left(\boldsymbol{W}_j\boldsymbol{\Upsilon}^*_{\boldsymbol{i}}\boldsymbol{H}'_{\boldsymbol{i}}\boldsymbol{J}'_{\boldsymbol{i}}\right),
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{E}\left[h^2\left(\|\boldsymbol{\xi}_3\|^2_{\{\Xi_i, i=1,\cdots,d\}}\right)\mathrm{trace}\left(\boldsymbol{\xi}^{*'}_{3_{(d)}}\boldsymbol{\xi}^{*}_{3_{(d)}}\right)\right]\\
&= \psi^{(1)}_{2,l+4}(\Delta)\mathrm{trace}\left(\boldsymbol{\delta}^{*'}_{(d)}\boldsymbol{\delta}^{*}_{(d)}\right) + \psi^{(2)}_{2,l+2}(\Delta)\prod_{i=1}^{d}\mathrm{trace}\left(\boldsymbol{W}_i\boldsymbol{\Upsilon}^*_{\boldsymbol{i}}\boldsymbol{H}'_{\boldsymbol{i}}\boldsymbol{J}'_{\boldsymbol{i}}\right),
\end{aligned}$$

this completes the proof. □

From Theorem C.3.1, one can obtain the results of Proposition C.3.1 by taking $h(x) = 1$ and $h(x) = 0$, respectively and by using the fact that, when $h(x) = 1$, $\psi^{(1)}_{1,l+4}(\Delta) = \psi^{(1)}_{2,l+4}(\Delta) = 1$ and $\psi^{(2)}_{1,l+2}(\Delta) = \psi^{(2)}_{2,l+2}(\Delta) = c$. From Theorem C.3.1, by using the fact that the distribution of $\boldsymbol{\epsilon}_1$ is a particular case of the one in Assumption C.3.1, one can deduce $\mathrm{ADR}^1\left(\hat{\mathbb{B}}(h), \mathbb{B}; \boldsymbol{W}\right)$. Further, by taking suitable function h, one can deduce $\mathrm{ADR}^1\left(\hat{\mathbb{B}}^s, \mathbb{B}; \boldsymbol{W}\right)$ and $\mathrm{ADR}^1\left(\hat{\mathbb{B}}^{sp}, \mathbb{B}; \boldsymbol{W}\right)$. Below, we establish a result which shows that for a suitable weighting matrix $\boldsymbol{W}$, SEs dominates the UE. As an intermediate step, we first derive the following proposition. To simplify some notations, let $c_1 = \operatorname{trace}\left(\boldsymbol{\delta}'_{(d)}\left(\boldsymbol{\delta}\Big(\bigtimes_{j=1}^{d}\Big)_j \boldsymbol{\Xi}_j \boldsymbol{\Upsilon}_j^* \boldsymbol{W}_j\right)_{(d)}\right)$, $c_2 = \prod_{j=1}^{d} \operatorname{trace}(\boldsymbol{W}_j \boldsymbol{\Upsilon}_j^*)$, $c_3 = \prod_{i=1}^{d} \operatorname{trace}(\boldsymbol{W}_i \boldsymbol{\Upsilon}_{\boldsymbol{i}}^* \boldsymbol{H}'_{\boldsymbol{i}} \boldsymbol{J}'_{\boldsymbol{i}})$, $c_4 = \operatorname{trace}\left(\boldsymbol{\delta}^{*\prime}_{(d)} \boldsymbol{\delta}^*_{(d)}\right)$.

Proposition C.3.2. *Suppose that Assumption C.3.1 holds and let*
$f_1(\Delta) = \mathrm{ADR}^1(\hat{\boldsymbol{\theta}}^s, \boldsymbol{\theta}; \boldsymbol{W}) - \mathrm{ADR}^1(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W})$. *Then, for all* $\Delta \geqslant 0$,

$$f_1(\Delta) = -2(l-2)^2\left(c_2 - \tfrac{c_3}{2}\right)\int_0^{\infty} \mathrm{E}\left[\chi_{l+2}^{-4}(t\Delta)\right]\omega(t)dt - \tfrac{2(l-2)}{2\Delta}\int_0^{\infty}(t-1)e^{-\Delta t/2}\omega(t)dt$$
$$-(l-2)\left(4\Delta c_2 - 4c_1 - (l-2)c_4\right)\int_0^{\infty} t\mathrm{E}\left[\chi_{l+4}^{-4}(t\Delta)\right]\omega(t)dt$$
$$-(l-2)^2\left(c_3 + \tfrac{4c_4}{\Delta}\right)\int_0^{\infty}(1-t)\mathrm{E}\left[\chi_{l+2}^{-4}(t\Delta)\right]\omega(t)dt.$$

Proof. We have $\mathrm{ADR}^1(\hat{\boldsymbol{\theta}}^s, \boldsymbol{\theta}; \boldsymbol{W}) = \int_0^{\infty} \mathrm{ADR}^1(\hat{\boldsymbol{\theta}}^s, \boldsymbol{\theta}; \boldsymbol{W} \mid t)\omega(t)dt$, with

$$\begin{aligned}
\mathrm{ADR}^1(\hat{\boldsymbol{\theta}}^s, \boldsymbol{\theta}; \boldsymbol{W} \mid t) &= t^{-1}\prod_{i=1}^{d}\operatorname{trace}(\boldsymbol{W}_i\boldsymbol{\Lambda}_i) - t^{-1}\prod_{i=1}^{d}\operatorname{trace}(\boldsymbol{W}_i\boldsymbol{\Upsilon}_i^*) + \operatorname{trace}(\boldsymbol{\delta}^{*\prime}_{(d)}\boldsymbol{\delta}^*_{(d)})\\
&-2\left(1 - t(l-2)\mathrm{E}\left[\chi_{l+2}^{-2}(t\,\Delta)\right]\right)\operatorname{trace}\left(\boldsymbol{\delta}^{*\prime}_{(d)}\boldsymbol{\delta}^*_{(d)}\right) - 2\left(t^{-1} - (l-2)\mathrm{E}\left[\chi_{l+2}^{-2}(t\,\Delta)\right]\right)c_3\\
&-2\left(1 - t\,(l-2)\mathrm{E}\left[\chi_{l+2}^{-2}(t\,\Delta)\right]\right)\operatorname{trace}\left(\boldsymbol{\delta}'_{(d)}\left(\boldsymbol{\delta}\Big(\bigtimes_{j=1}^{d}\Big)_j\boldsymbol{\Xi}_j\boldsymbol{\Upsilon}_j^*\boldsymbol{W}_j\right)_{(d)}\right)\\
&+2\left(t^{-1} - (l-2)\mathrm{E}\left[\chi_{l+2}^{-2}(t\,\Delta)\right]\right)\prod_{j=1}^{d}\operatorname{trace}(\boldsymbol{W}_j\boldsymbol{\Upsilon}_j^*)\\
&+2\left(1 - t\,(l-2)\mathrm{E}\left[\chi_{l+2}^{-2}(t\,\Delta)\right]\right)\operatorname{trace}\left(\boldsymbol{\delta}^{*\prime}_{(d)}\boldsymbol{\delta}^*_{(d)}\right)\\
&-2\left(1 - t\,(l-2)\mathrm{E}\left[\chi_{l+4}^{-2}(t\,\Delta)\right]\right)\operatorname{trace}\left(\boldsymbol{\delta}^{*\prime}_{(d)}\boldsymbol{\delta}^*_{(d)}\right)\\
&+2\left(1 - t\,(l-2)\mathrm{E}\left[\chi_{l+4}^{-2}(t\,\Delta)\right]\right)\operatorname{trace}\left(\boldsymbol{\delta}'_{(d)}\left(\boldsymbol{\delta}\Big(\bigtimes_{j=1}^{d}\Big)_j\boldsymbol{\Xi}_j\boldsymbol{\Upsilon}_j^*\boldsymbol{W}_j\right)_{(d)}\right)\\
&+\left(1 - 2t\,(l-2)\mathrm{E}\left[\chi_{l+4}^{-2}(\Delta)\right] + t^2(l-2)^2\mathrm{E}\left[\chi_{l+4}^{-4}(\Delta)\right]\right)\operatorname{trace}(\boldsymbol{\delta}^{*\prime}_{(d)}\boldsymbol{\delta}^*_{(d)})\\
&+\left(t^{-1} - 2\,(l-2)\mathrm{E}\left[\chi_{l+2}^{-2}(t\,\Delta)\right] + t(l-2)^2\mathrm{E}\left[\chi_{l+2}^{-4}(t\,\Delta)\right]\right)\prod_{i=1}^{d}\operatorname{trace}(\boldsymbol{W}_i\boldsymbol{\Upsilon}_{\boldsymbol{i}}^*\boldsymbol{H}'_{\boldsymbol{i}}\boldsymbol{J}'_{\boldsymbol{i}}).
\end{aligned}$$

The rest of the proof follows from some algebraic computations along with the identities

$$\mathrm{E}\left(\chi_{l+4}^{-2}(t\Delta)\right) = \mathrm{E}\left(\chi_{l+2}^{-2}(t\Delta)\right) - 2\mathrm{E}\left(\chi_{l+4}^{-4}(t\Delta)\right);\, t\,\mathrm{E}\left[\chi_{l+4}^{-4}(t\Delta)\right] = \frac{2}{\Delta}\mathrm{E}\left[\chi_{l+2}^{-4}(\Delta t)\right] - \frac{2}{\Delta p(l-2)}\,e^{-\Delta t/2};$$

$$\mathrm{E}_t\left[\chi_{l+2}^{-2}(t\Delta)\right] = (l-2)\mathrm{E}_t\left[\chi_{l+2}^{-4}(t\Delta)\right] + 2\Delta t\mathrm{E}_t\left[\chi_{l+4}^{-4}(t\Delta)\right].$$

□

From Proposition C.3.2 we derive below a result which shows that for a suitable choice of the weight matrices, $\boldsymbol{W}_i,\ i = 1, \cdots, d$, the SEs always dominate the UE.

To introduce some notation, let $\boldsymbol{\Pi}^{**} = \left(\boldsymbol{\Pi}^{*} + \boldsymbol{\Pi}^{*'}\right)/2$ where

$$\boldsymbol{\Pi}^{*} = \bigotimes_{i=d}^{1}\Xi_i^{1/2}\left(4\bigotimes_{i=d}^{1}\boldsymbol{\Upsilon}_i^{*} + (l-2)\bigotimes_{i=d}^{1}\boldsymbol{J}_i\boldsymbol{H}_i\boldsymbol{\Upsilon}_i^{*'}\right)\bigotimes_{i=d}^{1}\boldsymbol{W}_i\boldsymbol{\Upsilon}_i^{*}\boldsymbol{H}_i'\boldsymbol{J}_i'\Xi_i^{1/2}.$$

Define $\mathrm{Ch}_{\max}(\boldsymbol{A})$ to denote the maximum eigenvalue of a matrix $\boldsymbol{A}$.

Corollary C.3.1. *Suppose that Assumption C.3.1 holds where the weight mixture function* $\omega(.)$ *is such that* $\int_0^\infty(1-t)\mathrm{E}\left[\chi_{l+2}^{-4}(t\Delta)\right]\omega(t)dt \leqslant 0$, $\int_0^\infty(t-1)e^{-\Delta t/2}\omega(t)dt \leqslant 0$ *for all* $\Delta \geqslant 0$, *and suppose that*

$c_2 \geqslant \max\left\{\frac{c_3}{2}, \frac{\mathrm{Ch}_{\max}(\boldsymbol{\Pi}^{**})}{4}\right\}$. *Then,* $\mathrm{ADR}^1(\hat{\boldsymbol{\theta}}^{sp}, \boldsymbol{\theta}; \boldsymbol{W}) \leqslant \mathrm{ADR}^1(\hat{\boldsymbol{\theta}}^{s}, \boldsymbol{\theta}; \boldsymbol{W}) \leqslant \mathrm{ADR}^1(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W})$, *for all* $\Delta \geqslant 0$.

Proof. From Proposition C.3.2, $\mathrm{ADR}^1(\hat{\boldsymbol{\theta}}^{s}, \boldsymbol{\theta}; \boldsymbol{W}) - \mathrm{ADR}^1(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}) \leqslant 0$ provided that the following conditions hold: **(i)**. $2\,c_2 - c_3 \geq 0$, and **(ii)**. $4\,\Delta c_2 - 4\,c_1 - (l-2)\,c_4 \geqslant 0$. First, note that, if $4\,c_1 + (l-2)\,c_4 = 0$, then since $c_2 \geqslant 0$, the inequality in **(ii)** holds for any $\Delta \geqslant 0$. Second, note that

$$c_1 = (\mathrm{Vec}(\boldsymbol{\delta}))'\left(\bigotimes_{i=d}^{1}\Xi_i\boldsymbol{\Upsilon}_i^{*}\boldsymbol{W}_i\right)(\mathrm{Vec}(\boldsymbol{\delta})) = (\mathrm{Vec}(\boldsymbol{\delta}))'\left(\bigotimes_{i=d}^{1}\Xi_i\boldsymbol{\Upsilon}_i^{*}\boldsymbol{W}_i\right)\mathrm{Vec}\left(\boldsymbol{\delta}\Big(\bigtimes_{i=1}^{d}\Big)_i\boldsymbol{\Upsilon}_i^{*}\boldsymbol{H}_i'\boldsymbol{J}_i'\Xi_i\right).$$

Then, $c_1 = (\mathrm{Vec}(\boldsymbol{\delta}))'\left(\bigotimes_{i=d}^{1}\Xi_i\boldsymbol{\Upsilon}_i^{*}\boldsymbol{W}_i\boldsymbol{\Upsilon}_i^{*}\boldsymbol{H}_i'\boldsymbol{J}_i'\Xi_i\right)(\mathrm{Vec}\,(\boldsymbol{\delta}))$. We also have

$$c_4 = (\mathrm{Vec}(\boldsymbol{\delta}))'\left(\bigotimes_{i=d}^{1}\boldsymbol{W}_i\right)(\mathrm{Vec}(\boldsymbol{\delta}))$$
$$= \left(\mathrm{Vec}\left(\boldsymbol{\delta}\Big(\bigtimes_{i=1}^{d}\Big)_i\boldsymbol{\Upsilon}_i^{*}\boldsymbol{H}_i'\boldsymbol{J}_i'\Xi_i\right)\right)'\left(\bigotimes_{i=d}^{1}\boldsymbol{W}_i\right)\left(\mathrm{Vec}\left(\boldsymbol{\delta}\Big(\bigtimes_{i=1}^{d}\Big)_i\boldsymbol{\Upsilon}_i^{*}\boldsymbol{H}_i'\boldsymbol{J}_i'\Xi_i\right)\right).$$ Then,

$$c_4 = (\mathrm{Vec}(\boldsymbol{\delta}))'\left(\bigotimes_{i=d}^{1}\Xi_i\boldsymbol{J}_i\boldsymbol{H}_i\boldsymbol{\Upsilon}_i^{*'}\boldsymbol{W}_i\boldsymbol{\Upsilon}_i^{*}\boldsymbol{H}_i'\boldsymbol{J}_i'\Xi_i\right)(\mathrm{Vec}\,(\boldsymbol{\delta})).$$

Hence, $4\,c_1 + (l-2)\,c_4 = (\mathrm{Vec}(\boldsymbol{\delta}))'\left(\bigotimes_{i=d}^{1}\boldsymbol{\Xi}_i^{1/2}\boldsymbol{\Pi}^*\bigotimes_{i=d}^{1}\Xi_i^{1/2}\right)(\mathrm{Vec}(\boldsymbol{\delta}))$, where $\boldsymbol{\Pi}^* = \bigotimes_{i=d}^{1}\boldsymbol{\Xi}_i^{1/2}\left(4\bigotimes_{i=d}^{1}\boldsymbol{\Upsilon}_i^* + (l-2)\bigotimes_{i=d}^{1}\boldsymbol{J_i H_i \Upsilon_i^{*\prime}}\right)\bigotimes_{i=d}^{1}\boldsymbol{W}_i\boldsymbol{\Upsilon}_i^*\boldsymbol{H}_i'\boldsymbol{J}_i'\Xi_i^{1/2}$. Then,

$$\frac{4\,c_1 + (l-2)\,c_4}{\Delta} = \frac{\left(\mathrm{Vec}\left(\boldsymbol{\delta}\left(\bigtimes_{i=1}^{d}\right)_i\Xi_i^{1/2}\right)\right)'\boldsymbol{\Pi}^*\left(\mathrm{Vec}\left(\boldsymbol{\delta}\left(\bigtimes_{i=1}^{d}\right)_i\Xi_i^{1/2}\right)\right)}{\left(\mathrm{Vec}\left(\boldsymbol{\delta}\left(\bigtimes_{i=1}^{d}\right)_i\Xi_i^{1/2}\right)\right)'\left(\mathrm{Vec}\left(\boldsymbol{\delta}\left(\bigtimes_{i=1}^{d}\right)_i\Xi_i^{1/2}\right)\right)}.$$

This gives

$$\frac{4\,c_1 + (l-2)\,c_4}{\Delta} = \frac{\left(\mathrm{Vec}\left(\boldsymbol{\delta}\left(\bigtimes_{i=1}^{d}\right)_i\Xi_i^{1/2}\right)\right)'\left(\frac{\boldsymbol{\Pi}^* + \Pi^{*\prime}}{2}\right)\left(\mathrm{Vec}\left(\boldsymbol{\delta}\left(\bigtimes_{i=1}^{d}\right)_i\Xi_i^{1/2}\right)\right)}{\left(\mathrm{Vec}\left(\boldsymbol{\delta}\left(\bigtimes_{i=1}^{d}\right)_i\Xi_i^{1/2}\right)\right)'\left(\mathrm{Vec}\left(\boldsymbol{\delta}\left(\bigtimes_{i=1}^{d}\right)_i\Xi_i^{1/2}\right)\right)}.$$

Hence,

$$\frac{4\,c_1 + (l-2)\,c_4}{\Delta} \leqslant \mathrm{Ch}_{\max}\left(\frac{\boldsymbol{\Pi}^* + \Pi^{*\prime}}{2}\right) = \mathrm{Ch}_{\max}\left(\Pi^{**}\right).$$

Thus, if $c_2 > 0$, then $4\Delta\,c_2 - 4\,c_1 - (l-2)\,c_4 \geqslant 0$ if $c_2 \geqslant \max\left\{\frac{c_3}{2}, \frac{\mathrm{Ch}_{\max}(\Pi^{**})}{4}\right\}$. Similarly, one proves that $\mathrm{ADR}^1(\hat{\boldsymbol{\theta}}^{sp}, \boldsymbol{\theta}; \boldsymbol{W}) \leqslant \mathrm{ADR}^1(\hat{\boldsymbol{\theta}}^{s}, \boldsymbol{\theta}; \boldsymbol{W})$. □

Vita Auctoris

Mai Ghannam was born in 1992 in Tripoli, Libya. She graduated from Vincent Massey Secondary School in 2010. In 2014, she obtained a B.Sc. in Mathematics (Honours) with great distinction at the University of Windsor. She also completed her Master's in Statistics at the University of Windsor under the supervision of Dr. Sévérien Nkurunziza in May 2016. In Fall 2017, she returned to the University of Windsor to pursue her PhD studies in Statistics.