Preliminary Test Estimation in ULAN Models

Davy Paindaveine[1,2], Joséa Rasoafaraniaina[1] and Thomas Verdebout[1]

[1] ECARES and Mathematics Department, Université libre de Bruxelles (Belgium),

[2]Toulouse School of Economics, Université Toulouse Capitole (France)

Abstract

Preliminary test estimation is a methodology that combines goodness-of-fit testing and estimation. It is a natural procedure when it is suspected a priori that the parameter to be estimated satisfies some prespecified constraints, is a classical topic in estimation theory. In the present paper, we establish general results on the asymptotic behavior of preliminary test estimators. More precisely, we show that, in uniformly locally asymptotically normal (ULAN) models, a general asymptotic theory can be derived for preliminary test estimators based on estimators admitting generic Bahadur-type representations. This allows for a detailed comparison between classical estimators and preliminary test estimators in ULAN models. Our results, that, in standard linear regression models, are shown to reduce to some classical results, are also illustrated in more modern and involved setups, such as the multisample one where m covariance matrices $\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_m$ are to be estimated when it is suspected that these matrices might be equal, might be proportional, or might share a common "scale". Simulation results confirm our theoretical findings and an illustration on a real data example is provided.

Key words and phrases: ULAN models, Le Cam's asymptotic theory, Multisample covariance matrix estimation, Preliminary test estimation.

1 Introduction

Preliminary test estimation is a widely studied topic in Statistics and Econometrics, that can be traced back to the seminal paper by Bancroft (1944). Preliminary test estimators are typically useful when one has to perform statistical inference under "uncertain prior information". More formally, assume that one is interested in estimating a parameter $\boldsymbol{\theta}$ that belongs to some parameter space $\mathbf{\Theta} \subset \mathbb{R}^p$, under the uncertain prior information that $\boldsymbol{\theta}$ belongs to a given subset $\mathbf{\Theta}_0$ of $\mathbf{\Theta}$ (throughout, we assume that $\mathbf{\Theta}$ is an open subset of $\mathbb{R}^p$). Then, roughly speaking, the statistician may hesitate between (i) an *unconstrained* estimator $\hat{\boldsymbol{\theta}}_{\mathrm{U}}$ with values in $\boldsymbol{\Theta}$ or (ii) a *constrained* estimator $\hat{\boldsymbol{\theta}}_{\mathrm{C}}$ with values in $\boldsymbol{\Theta}_0$ only. The idea underpinning preliminary test estimation is relatively simple: if a suitable test ϕ_n for $\mathcal{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ against $\mathcal{H}_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$ does not reject the null hypothesis, then $\hat{\boldsymbol{\theta}}_{\mathrm{C}}$ should be used; on the contrary, if ϕ_n provided evidence against $\mathcal{H}_0$, then the unconstrained estimator $\hat{\boldsymbol{\theta}}_{\mathrm{U}}$ should be favoured. In other words, a preliminary test estimator based on the test ϕ_n and on the estimators $\hat{\boldsymbol{\theta}}_{\mathrm{U}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{C}}$ is

$$\hat{\boldsymbol{\theta}}_{\mathrm{PTE}} := \mathbb{I}[\phi_n = 0]\hat{\boldsymbol{\theta}}_{\mathrm{C}} + \mathbb{I}[\phi_n = 1]\hat{\boldsymbol{\theta}}_{\mathrm{U}}, \tag{1.1}$$

where $\mathbb{I}[A]$ stands for the indicator function associated with A and where $\phi_n = 1$ (resp., $\phi_n = 0$) indicates rejection (resp., non-rejection) of $\mathcal{H}_0$ by ϕ_n.

Since Bancroft (1944), preliminary test estimation has been an active research topic. Sen and Saleh (1979), Sen and Saleh (2006), Wan, Zou and Ohtani (2006) and Kibria and Saleh (2014) considered preliminary test estimation in regression models. Giles, Lieberman and Giles (1992) tackled the problem of selecting the size of the test ϕ_n when conducting preliminary test estimation in a misspecified regression model. Ohtani and Toyoda (1980) considered estimation of regression coefficients after a preliminary test of homoscedasticity. Preliminary test estimation in elliptical models has been considered in the contexts of linear regression and of principal component analysis; see Arashi et al. (2014) and Paindaveine, Rasoafaraniaina and Verdebout (2017), respectively. It has also been widely considered in time series analysis; see, e.g., Ahmed and Basu (2000), Maeyama, Tamaki and Taniguchi (2011), and the references therein. For a general overview of the topic, we refer to Giles and Giles (1993) and Saleh (2006).

Despite the many works on the topic, it seems that no general theory for the asymptotic behavior of preliminary test estimators is available in the literature. The main objective of the present paper is therefore to derive such a general theory and to do so in a broad class of models (that will include in particular all models mentioned above). Assuming that the underlying model is regular in the sense that it is uniformly locally asymptotically normal (ULAN), we will derive the asymptotic behavior of a general preliminary test estimator; more precisely, we will consider preliminary test estimators based on estimators $\hat{\boldsymbol{\theta}}_{\mathrm{U}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{C}}$ that admit Bahadur-type representations. Our asymptotic results do cover many of the existing results in the literature but also allow us to consider more modern and involved models, as we will illustrate in a multisample covariance estimation framework

As expected, the asymptotic behavior of preliminary test estimators will depend on the true value of the parameter $\boldsymbol{\theta}$. We first show that when this true value is fixed outside $\boldsymbol{\Theta}_0$, then, provided that the test ϕ_n is consistent, a preliminary test estimator is asymptotically equivalent in probability to the unconstrained estimator $\hat{\boldsymbol{\theta}}_{\mathrm{U}}$. Second, we show that when the true value of $\boldsymbol{\theta}$ asymptotically belongs to contiguous regions of $\boldsymbol{\Theta}_0$ (in a sense that involves the asymptotic concept of contiguity, as we will make precise below), a preliminary test estimator exhibits an asymptotic behavior that achieves a nice compromise between $\hat{\boldsymbol{\theta}}_{\mathrm{U}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{C}}$.

The paper is organized as follows. In Section 2, we describe the assumptions that will be considered in the sequel. In Section 3, we state our asymptotic results and derive explicit forms for the asymptotic mean square error of preliminary test estimators based on asymptotically efficient estimators. In Section 4, we illustrate these general results in two particular setups. First, we show that, in a simple linear regression context, our results allow us to recover the classical results from Saleh (2006). Then, we consider preliminary test estimation of m covariance matrices in a multisample Gaussian setup. Preliminary test estimators associated with the constraints of *covariance homogeneity*, *shape homogeneity* and *scale homogeneity* are studied. Monte Carlo simulations confirm our theoretical results. In Section 5, we provide a real data illustration that, in the aforementioned multisample covariance framework, shows that preliminary test estimators are practically relevant when performing supervised classification. Finally, an appendix collects the proofs.

2 ULAN models and Preliminary Test Estimators

As mentioned in the introduction, our objective is to derive the asymptotic behavior of preliminary test estimators (PTEs) in a very general context. We will throughout assume that the underlying parametric model $\{\mathrm{P}_{\boldsymbol{\theta}}^{(n)} : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p\}$ under investigation is *uniformly locally and asymptotically normal (ULAN)* in the following sense (throughout, convergences are as $n \to \infty$).

Assumption (A). There exist a sequence $(\boldsymbol{\nu}_n)$ of full-rank non-random $p \times p$ matrices that is $o(1)$, a sequence of random p-vectors $\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)}$ (the central sequence) and a symmetric positive semidefinite $p \times p$ matrix $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$ (the information matrix), such that, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, any sequence $(\boldsymbol{\theta}_n)$ in $\boldsymbol{\Theta}$ with $\boldsymbol{\nu}_n^{-1}(\boldsymbol{\theta}_n - \boldsymbol{\theta}) = O(1)$, and any bounded sequence $(\boldsymbol{\tau}_n)$ in $\mathbb{R}^p$ such that $\boldsymbol{\theta}_n + \boldsymbol{\nu}_n \boldsymbol{\tau}_n \in \boldsymbol{\Theta}$ for any n, we have

$$(i) \quad \Lambda^{(n)} := \log \frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n + \boldsymbol{\nu}_n \boldsymbol{\tau}_n}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}} = \boldsymbol{\tau}_n' \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}_n} - \frac{1}{2} \boldsymbol{\tau}_n' \boldsymbol{\Gamma}_{\boldsymbol{\theta}} \boldsymbol{\tau}_n + o_{\mathrm{P}}(1) \qquad (2.2)$$

and

$$\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\theta}})$$

under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$.

An extensive list of models do satisfy Assumption (A). This list includes hidden Markov models (Bickel and Ritov, 1996), quantum mechanics models (Kahn and Guta, 2009, Guta and Kiukas, 2015), time series models (Drost, Klaassen and Werker, 1997, Hallin et al., 1999, Francq and Zakoian, 2013), elliptical models (Hallin and Paindaveine, 2006, Hallin, Paindaveine and Verdebout, 2010), multisample elliptical models (Hallin and Paindaveine, 2008, Hallin, Paindaveine and Verdebout, 2013, 2014), models for directional data (Ley et al., 2013, Garcia-Portugues, Paindaveine and Verdebout, 2020), to mention only a few.

As explained in the introduction, the construction of a PTE involves an unconstrained estimator $\hat{\boldsymbol{\theta}}_{\rm U}$ taking values in $\boldsymbol{\Theta}$, a constrained estimator $\hat{\boldsymbol{\theta}}_{\rm C}$ taking values in $\boldsymbol{\Theta}_0$, and a test ϕ_n for $\mathcal{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ against $\mathcal{H}_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$. Throughout, we will assume that $\boldsymbol{\Theta}_0$ is a linear subspace of $\mathbb{R}^p$ of the form

$$\boldsymbol{\Theta}_0 = (\boldsymbol{\theta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})) \cap \boldsymbol{\Theta},$$

where $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is fixed and $\mathcal{M}(\boldsymbol{\Upsilon})$ denotes the vector subspace of $\mathbb{R}^p$ that is spanned by the columns of the $p \times r$ full-rank matrix $\boldsymbol{\Upsilon}$ $(r < p)$. We will restrict to the case $\boldsymbol{\theta}_0 = \mathbf{0}$, which is without loss of generality (a reparametrization of the model always allows us to reduce to this case). We will consider PTEs of the form

$$\hat{\boldsymbol{\theta}}_{\rm PTE} = \mathbb{I}[\phi_n = 1]\hat{\boldsymbol{\theta}}_{\rm U} + \mathbb{I}[\phi_n = 0]\hat{\boldsymbol{\theta}}_{\rm C},$$

based on estimators $\hat{\boldsymbol{\theta}}_{\rm U}$, $\hat{\boldsymbol{\theta}}_{\rm C}$ and on a test ϕ_n that satisfy the following assumption (throughout, $\chi^2_{\ell,\beta}$ denotes the upper β-quantile of the χ^2_ℓ distribution).

Assumption (B). With $\boldsymbol{\nu}_n$, $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}$ and $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$ as in Assumption (A), there exists, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, a random p-vector $\mathbf{S}^{(n)}_{\boldsymbol{\theta}}$ for which

$$\begin{pmatrix} \mathbf{S}^{(n)}_{\boldsymbol{\theta}} \\ \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} \end{pmatrix} \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} & \boldsymbol{\Omega}_{\boldsymbol{\theta}} \\ \boldsymbol{\Omega}_{\boldsymbol{\theta}} & \boldsymbol{\Gamma}_{\boldsymbol{\theta}} \end{pmatrix}\right)$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$ and for which the following holds:

(i) $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm U} - \boldsymbol{\theta}) = \mathbf{A}_{\boldsymbol{\theta}}\mathbf{S}^{(n)}_{\boldsymbol{\theta}} + o_{\rm P}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$ for some $p \times p$ matrix $\mathbf{A}_{\boldsymbol{\theta}}$;

(ii) if $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, then $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm C} - \boldsymbol{\theta}) = \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\mathbf{S}^{(n)}_{\boldsymbol{\theta}} + o_{\rm P}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$ for some $r \times p$ matrix $\mathbf{B}_{\boldsymbol{\theta}}$;

(iii) ϕ_n rejects $\mathcal{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ at asymptotic level α when $Q^{(n)} := \|\mathbf{D}^{(n)}\|^2 > \chi^2_{p-r,1-\alpha}$, where, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, the random p-vector $\mathbf{D}^{(n)}$ satisfies $\mathbf{D}^{(n)} = \mathbf{C}_{\boldsymbol{\theta}}\mathbf{S}^{(n)}_{\boldsymbol{\theta}} + o_{\rm P}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, with a $p \times p$ matrix $\mathbf{C}_{\boldsymbol{\theta}}$ for which (i) $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ and (ii) $\mathrm{tr}[\mathbf{C}'_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}] = p - r$. Moreover, ϕ is consistent under any $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$.

As complex as it may look, Assumption (B) is actually extremely mild. Indeed, provided that the underlying model is ULAN as in Assumption (A), it merely only imposes that an unconstrained estimator $\hat{\boldsymbol{\theta}}_{\rm U}$ admitting a Bahadur-type representation is available. To show this, let us restrict to the usual contiguity rate $\boldsymbol{\nu}_n = n^{-1/2}\mathbf{I}_p$ where $\mathbf{I}_p$ is the p-dimensional identity matrix (extension to a general $\boldsymbol{\nu}_n$ is direct) and let us assume that, under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{U}} - \boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{T}_i^{(n)} + o_{\mathrm{P}}(1), \tag{2.3}$$

where the random p-vectors $\mathbf{T}_i^{(n)} = \mathbf{T}_i^{(n)}(\boldsymbol{\theta})$, $i = 1, \ldots, n$, are mutually independent and share a common distribution that has mean zero and has finite second-order moments. Obviously, Assumption (B)(i) then holds with $\mathbf{A}_{\boldsymbol{\theta}} := \mathbf{I}_p$ and $\mathbf{S}_{\boldsymbol{\theta}}^{(n)} := n^{-1/2} \sum_{i=1}^{n} \mathbf{T}_i^{(n)}$. Under very mild assumptions (only needed to check the Levy–Lindeberg condition), a CLT for triangular arrays will then ensure that $(\mathbf{S}_{\boldsymbol{\theta}}^{(n)\prime}, \boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)\prime})'$ is asymptotically normal under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, as required in Assumption (B). Now, letting $\mathbf{P}_{\boldsymbol{\Upsilon}} := \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}'\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}'$ be the matrix of the orthogonal projection onto the constraint $\boldsymbol{\Theta}_0 = \mathcal{M}(\boldsymbol{\Upsilon}) \cap \boldsymbol{\Theta}$, the constrained estimator $\hat{\boldsymbol{\theta}}_{\mathrm{C}} := \mathbf{P}_{\boldsymbol{\Upsilon}}\hat{\boldsymbol{\theta}}_{\mathrm{U}}$ readily satisfies, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{C}} - \boldsymbol{\theta}) = \mathbf{P}_{\boldsymbol{\Upsilon}}\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{U}} - \boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{P}_{\boldsymbol{\Upsilon}}\mathbf{T}_i^{(n)} + o_{\mathrm{P}}(1)$$

under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, so that Assumption (B)(ii) is fulfilled, too (with $\mathbf{B}_{\boldsymbol{\theta}} := (\boldsymbol{\Upsilon}'\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}'$). Finally, Assumption (B)(iii) will be satisfied by Wald tests for $\mathcal{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ against $\mathcal{H}_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$ constructed in the usual way from (2.3). Wrapping up, the only key point in Assumption (B) is its part (i), which itself holds as soon as an unconstrained estimator $\hat{\boldsymbol{\theta}}_{\mathrm{U}}$ admitting a Bahadur-type representation is available. In regular models, M-, R-, and S-estimation, as usual, will provide such unconstrained estimators, so that Assumption (B) is not at all restrictive.

Now, in the ULAN framework of Assumption (A), an asymptotically efficient (unconstrained) estimator $\hat{\boldsymbol{\theta}}_{\mathrm{U}}$—that is, an estimator satisfying

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{U}} - \boldsymbol{\theta}) = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} + o_{\mathrm{P}}(1) \tag{2.4}$$

under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$ (see, e.g., Chapter 3 of Tanigushi and Kakizawa, 2000)—also satisfies Assumption (B)(i), with $\mathbf{A}_{\boldsymbol{\theta}} = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}$ and $\mathbf{S}_{\boldsymbol{\theta}}^{(n)} = \boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)}$ (which provides $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \boldsymbol{\Omega}_{\boldsymbol{\theta}} = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}$). An asymptotically efficient constrained estimator $\hat{\boldsymbol{\theta}}_{\mathrm{C}}$, that is such that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{C}} - \boldsymbol{\theta}) = \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}'\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} + o_{\mathrm{P}}(1) \tag{2.5}$$

under any $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, satisfies Assumption (B)(ii), with $\mathbf{B}_{\boldsymbol{\theta}} = (\boldsymbol{\Upsilon}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}'$ and $\mathbf{S}_{\boldsymbol{\theta}}^{(n)} = \boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)}$. For testing $\mathcal{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ against $\mathcal{H}_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$, the locally asymptotically most stringent test rejects $\mathcal{H}_0$ at asymptotic level α when

$$Q^{(n)} = \left\| \mathbf{C}_{\hat{\boldsymbol{\theta}}_{\mathrm{C}}} \boldsymbol{\Delta}_{\hat{\boldsymbol{\theta}}_{\mathrm{C}}}^{(n)} \right\|^2 > \chi^2_{p-r,1-\alpha}, \tag{2.6}$$

with

$$\mathbf{C}_{\boldsymbol{\theta}} := (\mathbf{I}_p - \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2})\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2} = (\mathbf{I}_p - \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2})\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}; \tag{2.7}$$

see, e.g., Chapter 5 of Ley and Verdebout (2017). Under Assumption (A), it is easy to check that, provided that $\hat{\boldsymbol{\theta}}_{\rm C}$ is locally and asymptotically discrete (a technical requirement with no practical impact), $\mathbf{C}_{\hat{\boldsymbol{\theta}}_{\rm C}}\boldsymbol{\Delta}^{(n)}_{\hat{\boldsymbol{\theta}}_{\rm C}} = \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} + o_{\rm P}(1)$ under any ${\rm P}^{(n)}_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, so that Assumption (B)(iii) then holds, still with $\mathbf{S}^{(n)}_{\boldsymbol{\theta}} = \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}$, and with the $\mathbf{C}_{\boldsymbol{\theta}}$ in (2.7) (one can indeed check that $\mathbf{C}'_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}} = \mathbf{C}'_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}$ and that ${\rm tr}[\mathbf{C}'_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}] = {\rm tr}[\mathbf{I}_p] - {\rm tr}[(\boldsymbol{\Upsilon}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon})^{-1}(\boldsymbol{\Upsilon}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon})] = p - r$). To summarize, Assumptions (A)–(B) cover many existing models and estimators. In the next section, our objective is to derive asymptotic results for PTEs in the general framework covered by these assumptions.

3 Asymptotic results

In this section, we derive, in a parametric model $\{{\rm P}^{(n)}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p\}$ satisfying Assumption (A), the asymptotic behavior of a PTE of the form

$$\hat{\boldsymbol{\theta}}_{\rm PTE} := \mathbb{I}[\phi_n = 1]\hat{\boldsymbol{\theta}}_{\rm U} + \mathbb{I}[\phi_n = 0]\hat{\boldsymbol{\theta}}_{\rm C}, \tag{3.8}$$

based on estimators $\hat{\boldsymbol{\theta}}_{\rm U}$, $\hat{\boldsymbol{\theta}}_{\rm C}$ and on a test ϕ_n that satisfy Assumption (B). Letting $\lambda(v) := \mathbb{I}[v \leq \chi^2_{p-r,1-\alpha}]$, the estimator in (3.8) rewrites

$$\hat{\boldsymbol{\theta}}_{\rm PTE} := (1 - \lambda(Q^{(n)}))\hat{\boldsymbol{\theta}}_{\rm U} + \lambda(Q^{(n)})\hat{\boldsymbol{\theta}}_{\rm C}. \tag{3.9}$$

When deriving the asymptotic behavior of $\hat{\boldsymbol{\theta}}_{\rm PTE}$ under ${\rm P}^{(n)}_{\boldsymbol{\theta}}$, we will discriminate between three cases: (i) $\boldsymbol{\theta}$ is fixed in the constraint $\boldsymbol{\Theta}_0$, (ii) $\boldsymbol{\theta} = \boldsymbol{\theta}_n$ belongs to the $\boldsymbol{\nu}_n$-vicinity of the constraint (that is, $\boldsymbol{\theta}_n = \boldsymbol{\theta} + \boldsymbol{\nu}_n\boldsymbol{\tau}_n$, with $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ and $(\boldsymbol{\tau}_n) = O(1)$), and (iii) $\boldsymbol{\theta}$ is fixed outside the constraint $\boldsymbol{\Theta}_0$; see Figure 1.

Our first result shows that, in case (iii), $\hat{\boldsymbol{\theta}}_{\rm PTE}$ is asymptotically equivalent in probability to the unconstrained estimator $\hat{\boldsymbol{\theta}}_{\rm U}$ (see the appendix for a proof).

Theorem 1. *Let Assumptions (A)–(B) hold. Fix $\boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$ and assume that $\hat{\boldsymbol{\theta}}_{\rm C} = O_{\rm P}(1)$ under ${\rm P}^{(n)}_{\boldsymbol{\theta}}$. Then, $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm PTE} - \boldsymbol{\theta}) = \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm U} - \boldsymbol{\theta}) + o_{\rm P}(1)$ under ${\rm P}^{(n)}_{\boldsymbol{\theta}}$.*

We now move to cases (i)–(ii), where we will actually consider parameter sequences of the form $\boldsymbol{\theta}_n = \boldsymbol{\theta} + \boldsymbol{\nu}_n\boldsymbol{\tau}_n \in \boldsymbol{\Theta}$, with $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ and $(\boldsymbol{\tau}_n) \to \boldsymbol{\tau}$ (note that case (i) is obtained for $\boldsymbol{\tau}_n \equiv \mathbf{0}$). We have the following result (see the appendix for a proof).

Theorem 2. *Let Assumptions (A)–(B) hold and consider sequences of the form $\boldsymbol{\theta}_n = \boldsymbol{\theta} + \boldsymbol{\nu}_n\boldsymbol{\tau}_n \in \boldsymbol{\Theta}$, with $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ and $(\boldsymbol{\tau}_n) \to \boldsymbol{\tau}$. Let $(\mathbf{Z}'_1, \mathbf{Z}'_2, \mathbf{D}')'$ be a Gaussian random vector with mean vector*

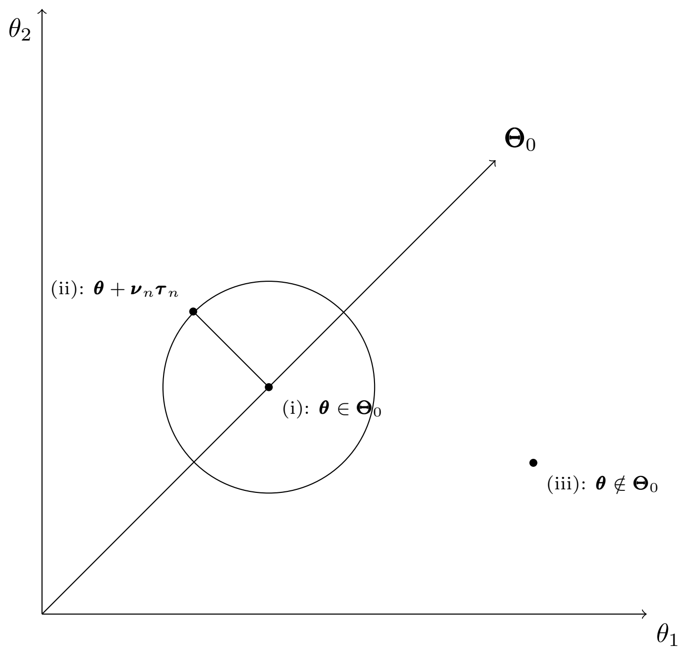

Figure 1: Illustration of the various situations where asymptotics are derived, for a bivariate parameter $\boldsymbol{\theta} = \binom{\theta_1}{\theta_2}$ and a constraint of the form $\boldsymbol{\Theta}_0 = \mathcal{M}(\boldsymbol{\Upsilon})$, with $\boldsymbol{\Upsilon} = \binom{1}{1}$.

$(\boldsymbol{\tau}'(\mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}} - \mathbf{I}_p)', \boldsymbol{\tau}'(\boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}} - \mathbf{I}_p)', \boldsymbol{\tau}'(\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}})')'$ *and covariance matrix*

$$\begin{pmatrix} \mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{A}'_{\boldsymbol{\theta}} & \mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{B}'_{\boldsymbol{\theta}}\boldsymbol{\Upsilon}' & \mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}} \\ \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{A}'_{\boldsymbol{\theta}} & \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{B}'_{\boldsymbol{\theta}}\boldsymbol{\Upsilon}' & \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}} \\ \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{A}'_{\boldsymbol{\theta}} & \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{B}'_{\boldsymbol{\theta}}\boldsymbol{\Upsilon}' & \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}} \end{pmatrix}.$$

Then, $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{PTE}} - \boldsymbol{\theta}_n)$ *converges weakly to*

$$\mathbf{Z} := (1 - \lambda(\|\mathbf{D}\|^2))\mathbf{Z}_1 + \lambda(\|\mathbf{D}\|^2)\mathbf{Z}_2. \tag{3.10}$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}$ *as* $n \to \infty$.

Theorem 2 provides the asymptotic behavior of $\hat{\boldsymbol{\theta}}_{\mathrm{PTE}}$ in the vicinity of $\boldsymbol{\Theta}_0$. Note that, using the identities $\lambda^2(v) = \lambda(v)$, $(1 - \lambda(v))^2 = 1 - \lambda(v)$, and $\lambda(v)(1 - \lambda(v)) = 0$, it is easy to see that, conditional on $\mathbf{D}$, the weak limit $\mathbf{Z}$ of $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{PTE}} - \boldsymbol{\theta}_n)$ in (3.10) is Gaussian with mean vector

$$\begin{aligned} \boldsymbol{\mu}^{\mathrm{Vic}}_{\mathrm{PTE}} &= (1 - \lambda(\|\mathbf{D}\|^2)) \left\{ (\mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}} - \mathbf{I}_p)\boldsymbol{\tau} + \mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}}(\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}})^-(\mathbf{D} - \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau}) \right\} \\ &\quad + \lambda(\|\mathbf{D}\|^2) \left\{ (\boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}} - \mathbf{I}_p)\boldsymbol{\tau} + \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}}(\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}})^-(\mathbf{D} - \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau}) \right\} \end{aligned} \tag{3.11}$$

and covariance matrix

$$\boldsymbol{\Gamma}^{\mathrm{Vic}}_{\mathrm{PTE}} = (1 - \lambda(\|\mathbf{D}\|^2))\mathbf{A}_{\boldsymbol{\theta}}(\boldsymbol{\Sigma}_{\boldsymbol{\theta}} - \mathbf{L}_{\boldsymbol{\theta}})\mathbf{A}'_{\boldsymbol{\theta}} + \lambda(\|\mathbf{D}\|^2)\boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}(\boldsymbol{\Sigma}_{\boldsymbol{\theta}} - \mathbf{L}_{\boldsymbol{\theta}})\mathbf{B}'_{\boldsymbol{\theta}}\boldsymbol{\Upsilon}', \tag{3.12}$$

where we denoted as $\mathbf{A}^-$ the Moore-Penrose inverse of $\mathbf{A}$ and where we let $\mathbf{L}_{\boldsymbol{\theta}} := \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}}(\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}})^-\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$. Since $\mathbf{D}$ in Theorem 2 is Gaussian with mean vector $\boldsymbol{\mu}_{\mathbf{D}} := \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{D}} = \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}'_{\boldsymbol{\theta}}$, the probability density function (pdf) of the weak limit $\mathbf{Z}$ of $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm PTE} - \boldsymbol{\theta}_n)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}$ is given by

$$\mathbf{z} \mapsto \int_{\mathbb{R}^p} \phi_{\boldsymbol{\mu}^{\rm Vic}_{\rm PTE}, \boldsymbol{\Gamma}^{\rm Vic}_{\rm PTE}}(\mathbf{z})\phi_{\boldsymbol{\mu}_{\mathbf{D}}, \boldsymbol{\Sigma}_{\mathbf{D}}}(\mathbf{x})d\mathbf{x}, \tag{3.13}$$

where $\phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ stands for the pdf of the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Since the pdf (3.13) does not allow for a simple comparison between $\hat{\boldsymbol{\theta}}_{\rm PTE}$, $\hat{\boldsymbol{\theta}}_{\rm U}$ and $\hat{\boldsymbol{\theta}}_{\rm C}$, we will base such a comparison on the asymptotic mean square errors (MSEs) of these estimators.

A general expression for the asymptotic MSEs can be obtained by computing $\mathrm{E}[\boldsymbol{\mu}^{\rm Vic}_{\rm PTE}]$, $\mathrm{Var}[\boldsymbol{\mu}^{\rm Vic}_{\rm PTE}]$ and $\mathrm{E}[\boldsymbol{\Gamma}^{\rm Vic}_{\rm PTE}]$. We now derive these limiting MSEs when PTEs are based on the preliminary tests in (2.6) and on asymptotically efficient estimators satisfying (2.4)–(2.5) (limiting MSEs of PTEs based on other estimators can be obtained in the same way). For such estimators and preliminary tests, the random p-vector $\mathbf{D}$ in Theorem 2 is Gaussian with mean vector $\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon}}\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}$ and covariance matrix $\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon}}$, where $\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff} := \mathbf{I}_p - \mathbf{P}_{\boldsymbol{\Upsilon},\rm eff}$ is based on $\mathbf{P}_{\boldsymbol{\Upsilon},\rm eff} := \boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}'\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}$, and it follows from (3.11)–(3.12) that, conditional on $\mathbf{D}$, the random vector $\mathbf{Z}$ in (3.10) is Gaussian with mean vector

$$\boldsymbol{\mu}^{\rm Vic}_{\rm PTE,eff} = \boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\big\{(1-\lambda(\|\mathbf{D}\|^2))\mathbf{D} - \boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}\big\} \tag{3.14}$$

and covariance matrix

$$\boldsymbol{\Gamma}^{\rm Vic}_{\rm PTE,eff} = \boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}}\mathbf{P}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}}. \tag{3.15}$$

We then have the following result (see the appendix for a proof).

Proposition 1. *If $\boldsymbol{\mu}^{\rm Vic}_{\rm PTE,eff}$ in (3.14) is based on a random p-vector $\mathbf{D}$ that is Gaussian with mean vector $\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon}}\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}$ and covariance matrix $\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon}}$, then*

$$\mathrm{E}[\boldsymbol{\mu}^{\rm Vic}_{\rm PTE,eff}] = -\gamma_2\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}$$

and

$$\begin{aligned}\mathrm{Var}[\boldsymbol{\mu}^{\rm Vic}_{\rm PTE,eff}] &= (1-\gamma_2)\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}} \\ &\quad +((1-\gamma_4)-(1-\gamma_2)^2)\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}},\end{aligned}$$

where we let $\gamma_j := \mathrm{P}[V_j \leq \chi^2_{p-r,1-\alpha}]$, with $V_j \sim \chi^2_{p-r+j}(\boldsymbol{\tau}'\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau})$ (throughout, $\chi^2_\ell(\eta)$ will stand for the non-central chi-square distribution with ℓ degrees of freedom and with non-centrality parameter η).

We define the asymptotic MSE of $\hat{\boldsymbol{\theta}}_{\rm PTE}$ under ${\rm P}^{(n)}_{\boldsymbol{\theta}_n}$ as

$$\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm PTE}) := {\rm E}[\mathbf{Z}\mathbf{Z}'] = {\rm Var}[\mathbf{Z}] + {\rm E}[\mathbf{Z}]({\rm E}[\mathbf{Z}])',$$

where $\mathbf{Z}$ is the weak limit of $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm PTE} - \boldsymbol{\theta}_n)$ under ${\rm P}^{(n)}_{\boldsymbol{\theta}_n}$; see Theorem 2. Now, since ${\rm E}[\mathbf{Z}] = {\rm E}[{\rm E}[\mathbf{Z}|\mathbf{D}]] = {\rm E}[\boldsymbol{\mu}^{\rm Vic}_{\rm PTE,eff}]$ and ${\rm Var}[\mathbf{Z}] = {\rm E}[{\rm Var}[\mathbf{Z}|\mathbf{D}]] + {\rm Var}[{\rm E}[\mathbf{Z}|\mathbf{D}]] = \boldsymbol{\Gamma}^{\rm Vic}_{\rm PTE,eff} + {\rm Var}[\boldsymbol{\mu}^{\rm Vic}_{\rm PTE,eff}]$ (note that ${\rm Var}[\mathbf{Z}|\mathbf{D}] = \boldsymbol{\Gamma}^{\rm Vic}_{\rm PTE,eff}$ is non-random), Proposition 1 yields

$$\begin{aligned}\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm PTE}) &= \boldsymbol{\Gamma}^{\rm Vic}_{\rm PTE,eff} + {\rm Var}[\boldsymbol{\mu}^{\rm Vic}_{\rm PTE,eff}] + ({\rm E}[\boldsymbol{\mu}^{\rm Vic}_{\rm PTE,eff}])({\rm E}[\boldsymbol{\mu}^{\rm Vic}_{\rm PTE,eff}])' \\ &= \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1} - \gamma_2 \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2} \\ &\quad + (2\gamma_2 - \gamma_4)\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}. \end{aligned} \tag{3.16}$$

To enable proper comparison with the unconstrained and constrained antecedents of $\hat{\boldsymbol{\theta}}_{\rm PTE}$ (namely, the estimators $\hat{\boldsymbol{\theta}}_{\rm U}$ and $\hat{\boldsymbol{\theta}}_{\rm C}$ satisfying (2.4) and (2.5), respectively), the following result provides explicit expressions for the asymptotic MSEs of these estimators (see the appendix for a proof).

Proposition 2. *Let Assumptions (A)–(B) hold. Then, under* ${\rm P}^{(n)}_{\boldsymbol{\theta}_n}$,

$$\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm U}) = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}$$

and

$$\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm C}) = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2} + \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2},$$

where $\hat{\boldsymbol{\theta}}_{\rm U}$ *and* $\hat{\boldsymbol{\theta}}_{\rm C}$ *are estimators satisfying* (2.4) *and* (2.5)*, respectively.*

It is worthwhile to consider some boundary cases. For $\alpha = 1$, we have $\gamma_2 = \gamma_4 = 0$, so that $\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm PTE}) = \text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm U})$, which is compatible with the fact that $\hat{\boldsymbol{\theta}}_{\rm PTE} = \hat{\boldsymbol{\theta}}_{\rm U}$ almost surely when the preliminary test ϕ_n is performed at asymptotic level $\alpha = 1$. At the other extreme, for $\alpha = 0$, we rather have $\gamma_2 = \gamma_4 = 1$, which provides

$$\begin{aligned}\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm PTE}) &= \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1} - \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2} + \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2} \\ &= \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2} + \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2} \\ &= \text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm C}),\end{aligned}$$

in agreement with the fact that $\hat{\boldsymbol{\theta}}_{\rm PTE} = \hat{\boldsymbol{\theta}}_{\rm C}$ almost surely when ϕ_n is performed at asymptotic level $\alpha = 0$.

To conclude this section, we provide a comparison between $\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm PTE})$, $\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm U})$, and $\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm C})$. These asymptotic MSEs being matrix-valued, it is needed to base this comparison on a scalar summary, such as, e.g., their trace. In the present case, where the unconstrained estimator satisfies $\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm U}) = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}$ (see Proposition 2), it is natural to measure the asymptotic performance of an estimator $\hat{\boldsymbol{\theta}}$ through the scalar quantity

$$\text{AMSE}^{\rm s}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}) := {\rm tr}[\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}})\,\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}], \tag{3.17}$$

which, for $\hat{\boldsymbol{\theta}}_{\rm U}$, will provide the "normalized" perfomance $\mathrm{AMSE}^{\rm s}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm U}) = p$, that does not depend on the value of $\boldsymbol{\theta}$ at which the contiguous alternatives $\boldsymbol{\theta}_n = \boldsymbol{\theta} + \boldsymbol{\nu}_n\boldsymbol{\tau}_n$ are localized. Proposition 2 also entails that

$$\mathrm{AMSE}^{\rm s}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm C}) = \mathrm{tr}[\mathbf{P}_{\boldsymbol{\Upsilon},\rm eff}] + \mathrm{tr}[\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}] = r + \|\boldsymbol{\delta}\|^2,$$

with $\boldsymbol{\delta} := \mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\rm eff}\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}$. Note that, at $\boldsymbol{\tau} = \mathbf{0}$, this shows that $\mathrm{AMSE}^{\rm s}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm C}) = r < p = \mathrm{AMSE}^{\rm s}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm U})$, which confirms the intuition that $\hat{\boldsymbol{\theta}}_{\rm C}$ dominates $\hat{\boldsymbol{\theta}}_{\rm U}$ when the true parameter value belongs to $\boldsymbol{\Theta}_0$. Now, it easily follows from (3.16) that

$$\mathrm{AMSE}^{\rm s}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm PTE}) = p - \gamma_2(p - r) + (2\gamma_2 - \gamma_4)\|\boldsymbol{\delta}\|^2,$$

where $\gamma_j = \mathrm{P}[V_j \leq \chi^2_{p-r,1-\alpha}]$, with $V_j \sim \chi^2_{p-r+j}(\|\boldsymbol{\delta}\|^2)$. Figure 2 plots, for $p = 10$, $r = 1$ and $\alpha = .05$, the quantities $\mathrm{AMSE}^{\rm s}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm U})$, $\mathrm{AMSE}^{\rm s}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm C})$ and $\mathrm{AMSE}^{\rm s}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\rm PTE})$ as functions of $\|\boldsymbol{\delta}\|^2$. The figure reveals that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ (which corresponds to $\boldsymbol{\delta} = \mathbf{0}$), the constrained estimator $\hat{\boldsymbol{\theta}}_{\rm C}$ has the best performance, as expected. The PTE performs better than $\hat{\boldsymbol{\theta}}_{\rm U}$ in the vicinity of the constraint ($\|\boldsymbol{\delta}\|$ small to moderate) and it is asymptotically equivalent to $\hat{\boldsymbol{\theta}}_{\rm U}$ far from the constraint ($\|\boldsymbol{\delta}\|$ large).

4 Two specific illustrations

In this section, we illustrate the general results obtained above in two particular cases. First, we consider preliminary test estimation in the simple linear regression model and show that we recover for this model and for the considered estimation problem the classical results of Saleh (2006) (Section 4.1). Then, we consider the joint estimation of m covariance matrices $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m$ in a context where it is suspected that these covariance matrices might be equal, might be proportional, or might share a common "scale" (Section 4.2).

4.1 Simple linear regression

Consider the simple linear regression model

$$\mathbf{Y} = \rho\mathbf{1}_n + \beta\mathbf{x} + \boldsymbol{\varepsilon}, \tag{4.18}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ is a response vector, $\mathbf{x} = (x_1, \ldots, x_n)'$ is a vector of non-random covariates, and where the error vector $\boldsymbol{\varepsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ is multinormal with mean zero and covariance matrix $\sigma^2\mathbf{I}_n$, for some $\sigma^2 > 0$. This is the classical simple linear model with intercept ρ, slope β,

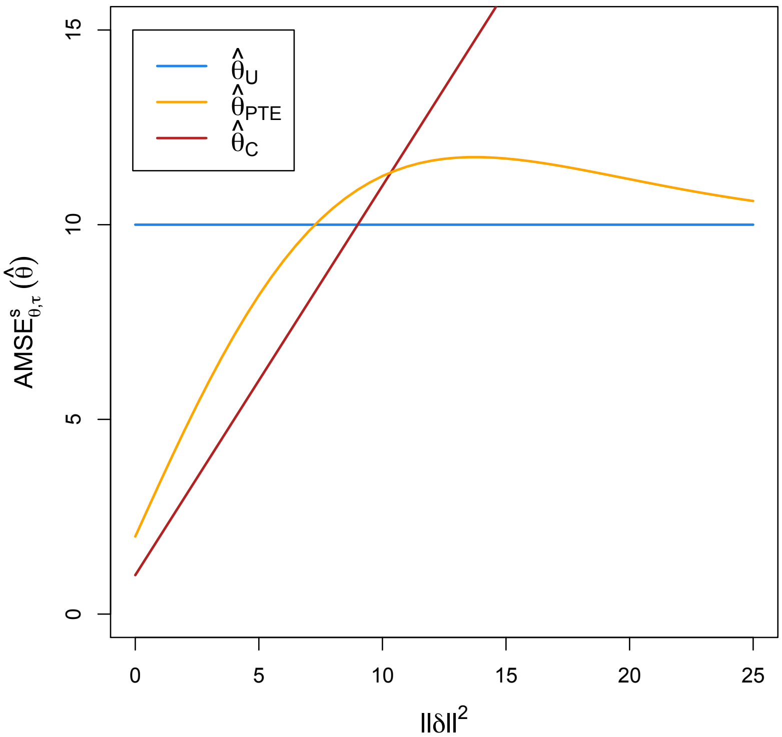

Figure 2: Plots of $\mathrm{AMSE}^{\mathrm{s}}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\mathrm{U}})$, $\mathrm{AMSE}^{\mathrm{s}}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\mathrm{C}})$ and $\mathrm{AMSE}^{\mathrm{s}}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\mathrm{PTE}})$ as functions of $\|\boldsymbol{\delta}\|^2$, for $p = 10$, $r = 1$ and $\alpha = .05$.

and Gaussian homoscedastic errors with variance σ^2. Throughout, we consider the parameter $\boldsymbol{\theta} := (\rho, \beta)'$, as we will assume that σ^2 is known (this is actually no restriction, since the block-diagonality of the Fisher information matrix in this model entails that replacing σ^2 with a root-n consistent estimator will have no asymptotic cost, so that all results we obtain below extend to the case where σ^2 would remain an unspecified nuisance). Under mild assumptions on the limiting behavior of the x_i's (ensuring that the quantities $\bar{x}_0$ and s_0 below do exist and are finite), one can easily show that this model is ULAN, with a central sequence $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}$ that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, is asymptotically normal with mean zero and covariance matrix

$$\boldsymbol{\Gamma}_{\boldsymbol{\theta}} = \frac{1}{\sigma^2}\begin{pmatrix} 1 & \bar{x}_0 \\ \bar{x}_0 & s_0 + \bar{x}_0^2 \end{pmatrix},$$

where $\bar{x}_0 := \lim_{n\to\infty} n^{-1}\sum_{i=1}^n x_i$ and $s_0 := \lim_{n\to\infty} s_x^{(n)}$, with $s_x^{(n)} := n^{-1}\mathbf{x}'\mathbf{x} - n^{-2}(\mathbf{1}_n'\mathbf{x})^2$. We consider here preliminary test estimation of $\boldsymbol{\theta}$ when it is suspected that $\beta = \beta_0$ for some given β_0.

In this context, the classical, unconstrained, estimator of $\boldsymbol{\theta}$ is the maximum likelihood estimator

$$\hat{\boldsymbol{\theta}}_{\mathrm{U}} := \begin{pmatrix} \hat{\rho} \\ \hat{\beta} \end{pmatrix} := \begin{pmatrix} n^{-1}(\mathbf{1}_n'\mathbf{Y} - \hat{\beta}\mathbf{1}_n'\mathbf{x}) \\ (\mathbf{x}'\mathbf{Y} - n^{-1}\mathbf{x}'\mathbf{1}_n\mathbf{1}_n'\mathbf{Y})/ns_x^{(n)} \end{pmatrix},$$

whereas the natural constrained estimator is $\hat{\boldsymbol{\theta}}_{\mathrm{C}} := \binom{\tilde{\rho}}{\beta_0}$, with $\tilde{\rho} := n^{-1}(\mathbf{1}_n'\mathbf{Y} - \beta_0\mathbf{1}_n'\mathbf{x})$. Since the locally asymptotically optimal test for $\mathcal{H}_0 : \beta = \beta_0$ against $\mathcal{H}_1 : \beta \neq \beta_0$ rejects the null hypothesis at asymptotic level α when

$$Q^{(n)} := \frac{n(\hat{\beta} - \beta_0)^2 s_x^{(n)}}{\sigma^2} > \chi^2_{1,1-\alpha},$$

the resulting PTE is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{PTE}} = \begin{pmatrix} \hat{\rho}_{\mathrm{PTE}} \\ \hat{\beta}_{\mathrm{PTE}} \end{pmatrix} := \mathbb{I}[Q^{(n)} > \chi^2_{1,1-\alpha}]\hat{\boldsymbol{\theta}}_{\mathrm{U}} + \mathbb{I}[Q^{(n)} \leq \chi^2_{1,1-\alpha}]\hat{\boldsymbol{\theta}}_{\mathrm{C}}.$$

Letting $\boldsymbol{\theta}_0 = \binom{\rho}{\beta_0}$ be an arbitrary value of the parameter of interest corresponding to the constraint, the null hypothesis can be written as $\mathcal{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\theta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$, with $\boldsymbol{\Upsilon} := \binom{1}{0}$. Since

$$\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1} = \sigma^2 \begin{pmatrix} 1 + \frac{\bar{x}_0^2}{s_0} & -\frac{\bar{x}_0}{s_0} \\ -\frac{\bar{x}_0}{s_0} & \frac{1}{s_0} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}_{\boldsymbol{\Upsilon}}^{\perp}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2} = \begin{pmatrix} 0 & -\bar{x}_0 \\ 0 & 1 \end{pmatrix},$$

it follows from (3.16) that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\tau}}$, with $\boldsymbol{\tau} = \binom{0}{\delta}$, the quantity $\mathrm{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\mathrm{PTE}})$ is here given by

$$\begin{pmatrix} \sigma^2(1 + \frac{\bar{x}_0^2}{s_0} - \frac{\gamma_2\bar{x}_0^2}{s_0}) + (2\gamma_2 - \gamma_4)\bar{x}_0^2\delta^2 & \frac{\sigma^2(\gamma_2-1)\bar{x}_0}{s_0} - (2\gamma_2 - \gamma_4)\bar{x}_0\delta^2 \\ \frac{\sigma^2(\gamma_2-1)\bar{x}_0}{s_0} - (2\gamma_2 - \gamma_4)\bar{x}_0\delta^2 & \frac{\sigma^2(1-\gamma_2)}{s_0} + (2\gamma_2 - \gamma_4)\delta^2 \end{pmatrix},$$

where the γ_j's are computed with $p = 2$ and $r = 1$. This is in perfect agreement with the result in Theorem 4, p.p. 94–96 in Saleh (2006).

4.2 Multisample estimation of covariance matrices

Consider $m(\geq 2)$ mutually independent samples of random k-vectors $\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i}$, $i = 1, \ldots, m$, with respective sample sizes $n_1, \ldots, n_m$, such that, for any i, the $\mathbf{X}_{ij}$'s form a random sample from the multinormal distribution with mean vector $\mathbf{0}$ and (invertible) covariance matrix $\boldsymbol{\Sigma}_i$ (all results below extend to the case where observations in the ith sample would have a common, unspecified, mean $\boldsymbol{\mu}_i$, $i = 1, \ldots, n$, due to the block-diagonality of the Fisher information matrix for location and scatter in elliptical models; see, e.g., Hallin and Paindaveine, 2006). In the sequel, we decompose the covariance matrices into $\boldsymbol{\Sigma}_i = \sigma_i^2\mathbf{V}_i$, where $\sigma_i := (\det\boldsymbol{\Sigma}_i)^{1/(2k)}$ is their "scale" and $\mathbf{V}_i := \boldsymbol{\Sigma}_i/(\det\boldsymbol{\Sigma}_i)^{1/k}$ is their "shape". Under the only assumption that $\lambda_i := \lambda_i^{(n)} := n_i/n := n_i/(\sum_{\ell=1}^m n_\ell)$ converges in $(0, 1)$ for any i (to make the notation lighter, we will not stress the dependence in n in many quantities below), it follows from Hallin and Paindaveine (2009) that the sequence of Gaussian models indexed by

$$\boldsymbol{\theta} := \big(\sigma_1^2, \ldots, \sigma_m^2, (\overset{\circ}{\mathrm{vech}}\,\mathbf{V}_1)', \ldots, (\overset{\circ}{\mathrm{vech}}\,\mathbf{V}_m)'\big)', \tag{4.19}$$

where $\overset{\circ}{\mathrm{vech}}\,\mathbf{V}(\in \mathbb{R}^{d_k}$, with $d_k := k(k+1)/2-1)$ stands for the vector obtained by depriving $\mathrm{vech}\mathbf{V}$ of its first entry $\mathbf{V}_{11}$, is ULAN in the sense of Assumption (A). To describe the corresponding central sequence and Fisher information matrix, we need the following notation.

Denoting as $\mathbf{e}_r$ the rth vector of the canonical basis of $\mathbb{R}^k$, let $\mathbf{K}_k := \sum_{r,s=1}^k (\mathbf{e}_r\mathbf{e}_s') \otimes (\mathbf{e}_s\mathbf{e}_r')$ be the $k^2 \times k^2$ *commutation matrix*, put $\mathbf{J}_k := (\mathrm{vec}\,\mathbf{I}_k)(\mathrm{vec}\,\mathbf{I}_k)'$, and define $\mathbf{M}_k(\mathbf{V})$ as the $(d_k \times k^2)$ matrix such that $(\mathbf{M}_k(\mathbf{V}))'(\overset{\circ}{\mathrm{vech}}\,\mathbf{v}) = \mathrm{vec}\,\mathbf{v}$ for any symmetric $k \times k$ matrix $\mathbf{v}$ such that $\mathrm{tr}[\mathbf{V}^{-1}\mathbf{v}] = 0$. We further put

$$\mathbf{H}_k(\mathbf{V}) := \frac{1}{4}\,\mathbf{M}_k(\mathbf{V})\,(\mathbf{I}_{k^2} + \mathbf{K}_k)\,(\mathbf{V} \otimes \mathbf{V})^{-1}\,(\mathbf{M}_k(\mathbf{V}))'.$$

Then, letting $\mathbf{S}_i := n_i^{-1} \sum_{j=1}^{n_i} \mathbf{X}_{ij}\mathbf{X}_{ij}'$ be the empirical covariance matrix in sample i (with respect to the fixed location $\boldsymbol{\mu}_i = \mathbf{0}$), the central sequence is

$$\boldsymbol{\Delta}_{\boldsymbol{\theta}} = \big(\Delta_{\boldsymbol{\theta}}^{I,1}, \ldots, \Delta_{\boldsymbol{\theta}}^{I,m}, (\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{II,1})', \ldots, (\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{II,m})'\big)',$$

where, for $i = 1, \ldots, m$, we wrote

$$\Delta_{\boldsymbol{\theta}}^{I,i} := \frac{\sqrt{n_i}}{2\sigma_i^2}\,\mathrm{tr}\big[\sigma_i^{-2}\mathbf{V}_i^{-1}(\mathbf{S}_i - \sigma_i^2\mathbf{V}_i)\big] \quad \text{and} \quad \boldsymbol{\Delta}_{\boldsymbol{\theta}}^{II,i} := \frac{\sqrt{n_i}}{2\sigma_i^2}\,\mathbf{M}_k(\mathbf{V}_i)(\mathbf{V} \otimes \mathbf{V})^{-1}(\mathrm{vec}\,\mathbf{S}_i),$$

whereas the (full-rank) information matrix takes the block-diagonal form $\boldsymbol{\Gamma}_{\boldsymbol{\theta}} := \mathrm{diag}(\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{I}, \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{II})$, with

$$\boldsymbol{\Gamma}^{I} := \frac{k}{2}\mathrm{diag}\big(\sigma_1^{-4}, \ldots, \sigma_m^{-4}\big) \quad \text{and} \quad \boldsymbol{\Gamma}^{II} := \mathrm{diag}\big(\mathbf{H}_k(\mathbf{V}_1), \ldots, \mathbf{H}_k(\mathbf{V}_m)\big).$$

The corresponding contiguity rate $\boldsymbol{\nu}_n$ in Assumption (A) is given by $\boldsymbol{\nu}_n = n^{-1/2}\mathbf{r}_n$, with

$$\mathbf{r}_n := \mathrm{diag}\big(\lambda_1^{-1/2}, \ldots, \lambda_m^{-1/2}, \lambda_1^{-1/2}\mathbf{I}_{d_k}, \ldots, \lambda_m^{-1/2}\mathbf{I}_{d_k}\big).$$

We consider here estimation of $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m$ or, equivalently, estimation of $\boldsymbol{\theta}$ in (4.19). An advantage of the $\boldsymbol{\theta}$-parametrization is that it allows the construction of various PTEs: one may suspect, e.g., *scale homogeneity* $\mathcal{H}_0^{\mathrm{scale}} : \sigma_1^2 = \ldots = \sigma_m^2$, *shape homogeneity* $\mathcal{H}_0^{\mathrm{shape}} : \mathbf{V}_1 = \ldots = \mathbf{V}_m$, or full *covariance homogeneity* $\mathcal{H}_0^{\mathrm{cov}} : \sigma_1^2\mathbf{V}_1 = \ldots = \sigma_m^2\mathbf{V}_m$, that is, $\mathcal{H}_0^{\mathrm{cov}} : \boldsymbol{\Sigma}_1 = \ldots = \boldsymbol{\Sigma}_m$. An asymptotically efficient unconstrained estimator in this Gaussian model is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{U}} := \left((\det \mathbf{S}_1)^{1/k}, \ldots, (\det \mathbf{S}_m)^{1/k}, \frac{(\overset{\circ}{\mathrm{vech}}\,\mathbf{S}_1)'}{(\det \mathbf{S}_1)^{1/k}}, \ldots, \frac{(\overset{\circ}{\mathrm{vech}}\,\mathbf{S}_m)'}{(\det \mathbf{S}_m)^{1/k}}\right)', \tag{4.20}$$

whereas, writing $\mathbf{S} := n^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{X}_{ij}\mathbf{X}_{ij}'$ for the pooled covariance matrix estimator (with respect to the fixed locations $\boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_m = \mathbf{0}$), asymptotically efficient constrained estimators, for the three constraints $\mathcal{H}_0^{\text{scale}}$, $\mathcal{H}_0^{\text{shape}}$ and $\mathcal{H}_0^{\text{cov}}$ above, are given by

$$\hat{\boldsymbol{\theta}}_{\text{C}}^{\text{scale}} := \left((\det \mathbf{S})^{1/k}\mathbf{1}_m', \frac{(\overset{\circ}{\text{vech}}\,\mathbf{S}_1)'}{(\det \mathbf{S}_1)^{1/k}}, \ldots, \frac{(\overset{\circ}{\text{vech}}\,\mathbf{S}_m)'}{(\det \mathbf{S}_m)^{1/k}}\right)', \tag{4.21}$$

$$\hat{\boldsymbol{\theta}}_{\text{C}}^{\text{shape}} := \left((\det \mathbf{S}_1)^{1/k}, \ldots, (\det \mathbf{S}_m)^{1/k}, \mathbf{1}_m' \otimes \frac{(\overset{\circ}{\text{vech}}\,\mathbf{S})'}{(\det \mathbf{S})^{1/k}}\right)' \tag{4.22}$$

and

$$\hat{\boldsymbol{\theta}}_{\text{C}}^{\text{cov}} := \left((\det \mathbf{S})^{1/k}\mathbf{1}_m', \mathbf{1}_m' \otimes \frac{(\overset{\circ}{\text{vech}}\,\mathbf{S})'}{(\det \mathbf{S})^{1/k}}\right)', \tag{4.23}$$

respectively. The three hypotheses $\mathcal{H}_0^{\text{scale}}$, $\mathcal{H}_0^{\text{shape}}$ and $\mathcal{H}_0^{\text{cov}}$ impose linear restrictions on $\boldsymbol{\theta}$, hence can be written as

$$\mathcal{H}_0^{\text{scale}} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\text{scale}}), \quad \mathcal{H}_0^{\text{shape}} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\text{shape}}) \quad \text{and} \quad \mathcal{H}_0^{\text{cov}} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\text{cov}})$$

(more specifically, $\boldsymbol{\Upsilon}_{\text{scale}} := \text{diag}(\mathbf{1}_m, \mathbf{I}_{md_k})$, $\boldsymbol{\Upsilon}_{\text{shape}} := \text{diag}(\mathbf{I}_m, \mathbf{1}_m \otimes \mathbf{I}_{d_k})$ and $\boldsymbol{\Upsilon}_{\text{cov}} := \text{diag}(\mathbf{1}_m, \mathbf{1}_m \otimes \mathbf{I}_{d_k})$). Now, if the $p \times r$ matrix $\boldsymbol{\Upsilon}$ stands for either of $\boldsymbol{\Upsilon}_{\text{scale}}$, $\boldsymbol{\Upsilon}_{\text{shape}}$ or $\boldsymbol{\Upsilon}_{\text{cov}}$ (of course, each constraint matrix has its own r), the locally asymptotically most stringent test $\phi_{\boldsymbol{\Upsilon}}^{(n)}$ for $\mathcal{H}_0 : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon})$ rejects the null hypothesis at asymptotic level α when

$$\begin{aligned} Q_{\boldsymbol{\theta},\boldsymbol{\Upsilon}}^{(n)} &:= \boldsymbol{\Delta}_{\boldsymbol{\theta}}' \left[\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1} - (\mathbf{r}^{(n)})^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}'(\mathbf{r}^{(n)})^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}(\mathbf{r}^{(n)})^{-1}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}'(\mathbf{r}^{(n)})^{-1}\right]\boldsymbol{\Delta}_{\boldsymbol{\theta}} \\ &> \chi^2_{m(d_k+1)-r,1-\alpha}. \end{aligned} \tag{4.24}$$

This allows us to consider the PTEs

$$\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{scale}} := \mathbb{I}[\phi_{\boldsymbol{\Upsilon}_{\text{scale}}}^{(n)} = 1]\hat{\boldsymbol{\theta}}_{\text{U}} + \mathbb{I}[\phi_{\boldsymbol{\Upsilon}_{\text{scale}}}^{(n)} = 0]\hat{\boldsymbol{\theta}}_{\text{C}}^{\text{scale}},$$

$$\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{shape}} := \mathbb{I}[\phi_{\boldsymbol{\Upsilon}_{\text{shape}}}^{(n)} = 1]\hat{\boldsymbol{\theta}}_{\text{U}} + \mathbb{I}[\phi_{\boldsymbol{\Upsilon}_{\text{shape}}}^{(n)} = 0]\hat{\boldsymbol{\theta}}_{\text{C}}^{\text{shape}}$$

and

$$\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}} := \mathbb{I}[\phi_{\boldsymbol{\Upsilon}_{\text{cov}}}^{(n)} = 1]\hat{\boldsymbol{\theta}}_{\text{U}} + \mathbb{I}[\phi_{\boldsymbol{\Upsilon}_{\text{cov}}}^{(n)} = 0]\hat{\boldsymbol{\theta}}_{\text{C}}^{\text{cov}}.$$

To compare these PTEs with their unconstrained and constrained antecedents, we performed the following Monte Carlo exercise, that focuses on the case $m = 2$ and $k = 2$. For each $\ell = 0, \ldots, 9$ and for each value of $n_1 = n_2 (= n/2) \in \{200, 2\,000, 20\,000\}$, we generated independently $M = 10\,000$ collections of mutually independent observations $\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}, \mathbf{Y}_{1,\ell}, \ldots, \mathbf{Y}_{n_2,\ell}$, where the $\mathbf{X}_i$'s are $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_1)$ and the $\mathbf{Y}_{i,\ell}$'s are $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{2,\ell})$, with $\boldsymbol{\Sigma}_1 = \mathbf{I}_k$ and with $\boldsymbol{\Sigma}_{2,\ell} = \sigma_{2,\ell}^2 \mathbf{V}_{2,\ell}$ based on

$$\sigma_{2,\ell}^2 = e^{\ell/(2\sqrt{n})} \quad \text{and} \quad \mathbf{V}_{2,\ell} = \begin{pmatrix} 1 & \frac{2\ell}{3\sqrt{n}} \\ \frac{2\ell}{3\sqrt{n}} & 1 \end{pmatrix} \Big/ \sqrt{\det \begin{pmatrix} 1 & \frac{2\ell}{3\sqrt{n}} \\ \frac{2\ell}{3\sqrt{n}} & 1 \end{pmatrix}}.$$

For $\ell = 0$, both populations share the same covariance matrix, hence also the same scales and shapes, whereas $\ell = 1, \ldots, 9$ provide increasingly distinct scales and shapes. In other words, the constraints above are met for $\ell = 0$ and they are more and more severely violated for $\ell = 1, \ldots, 9$. For any estimator $\hat{\boldsymbol{\theta}}$ of the corresponding true parameter value $\boldsymbol{\theta}$, we measure the performance of $\hat{\boldsymbol{\theta}}$ through

$$\mathrm{MSE}^s_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) := \mathrm{tr}[\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2} \mathrm{MSE}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}], \tag{4.25}$$

with

$$\mathrm{MSE}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) := \frac{1}{M} \sum_{m=1}^{M} \{\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}^{(m)} - \boldsymbol{\theta})\}\{\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}^{(m)} - \boldsymbol{\theta})\}',$$

where $\hat{\boldsymbol{\theta}}^{(m)}$ is the value the estimator takes in the mth of the M replications. Figure 3 then plots the values of $\mathrm{MSE}^s_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ for $\hat{\boldsymbol{\theta}}^{\mathrm{scale}}_{\mathrm{PTE}}$, $\hat{\boldsymbol{\theta}}^{\mathrm{shape}}_{\mathrm{PTE}}$ and $\hat{\boldsymbol{\theta}}^{\mathrm{cov}}_{\mathrm{PTE}}$ (with all preliminary tests performed at asymptotic level $\alpha = 5\%$) and for their constrained and unconstrained antecedents $\hat{\boldsymbol{\theta}}^{\mathrm{scale}}_{\mathrm{C}}$, $\hat{\boldsymbol{\theta}}^{\mathrm{shape}}_{\mathrm{C}}$, $\hat{\boldsymbol{\theta}}^{\mathrm{cov}}_{\mathrm{C}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{U}}$. To match what was done in Figure 2, in Figure 3, these quantities are not plotted as functions of ℓ, but rather as functions of the induced quantity $\|\boldsymbol{\delta}\|^2$. Irrespective of the sample size and of the constraint, the PTEs achieve a nice trade-off between the constrained and unconstrained estimators. For large sample sizes, the finite-sample MSEs are clearly in an excellent agreement with their asymptotic versions in (3.17), that are also plotted in Figure 3.

5 Real data example

To demonstrate the practical relevance of the PTEs we introduced for multisample covariance estimation in Section 4.2, we now provide a real data example. The dataset involves $n_1 = 49$ pairs of monozygotic male twins and $n_2 = 36$ pairs of monozygotic female twins (group 1 and group 2, respectively, say). For each pair of twins, six variables are available, namely the stature, hip width, and chest circumference for each twin. The resulting dataset, that thus collects $n_1 + n_2 = n = 85$ vectors $\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n_1}, \mathbf{x}_{21}, \ldots, \mathbf{x}_{2n_2}$ in $\mathbb{R}^6$, has been analyzed in Flury (2013) and is available in the R package `Flury` (see the data frames `m.twins` and `f.twins`, for male and female twins, respectively).

We consider estimation of the underlying covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ of the two groups. Of couse, the natural unconstrained estimator is the one associated with the group-specific sample covariance matrices $\hat{\boldsymbol{\Sigma}}_{\ell,\mathrm{U}} := n_\ell^{-1} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \hat{\boldsymbol{\mu}}_\ell)(\mathbf{x}_{\ell j} - \hat{\boldsymbol{\mu}}_\ell)'$, $\ell = 1, 2$, where $\hat{\boldsymbol{\mu}}_\ell := n_\ell^{-1} \sum_{j=1}^{n_\ell} \mathbf{x}_{\ell j}$, $\ell = 1, 2$, estimate the corresponding mean vectors. In the same spirit as in Section 4.2, we consider the PTEs associated with each of the following three constraints: homogeneity of scales ($\mathcal{H}_0^a$), homogeneity of shape matrices ($\mathcal{H}_0^b$), and homogeneity of covariance matrices ($\mathcal{H}_0^c$). These PTEs are based on the unconstrained estimator above and on constrained estimators obtained from

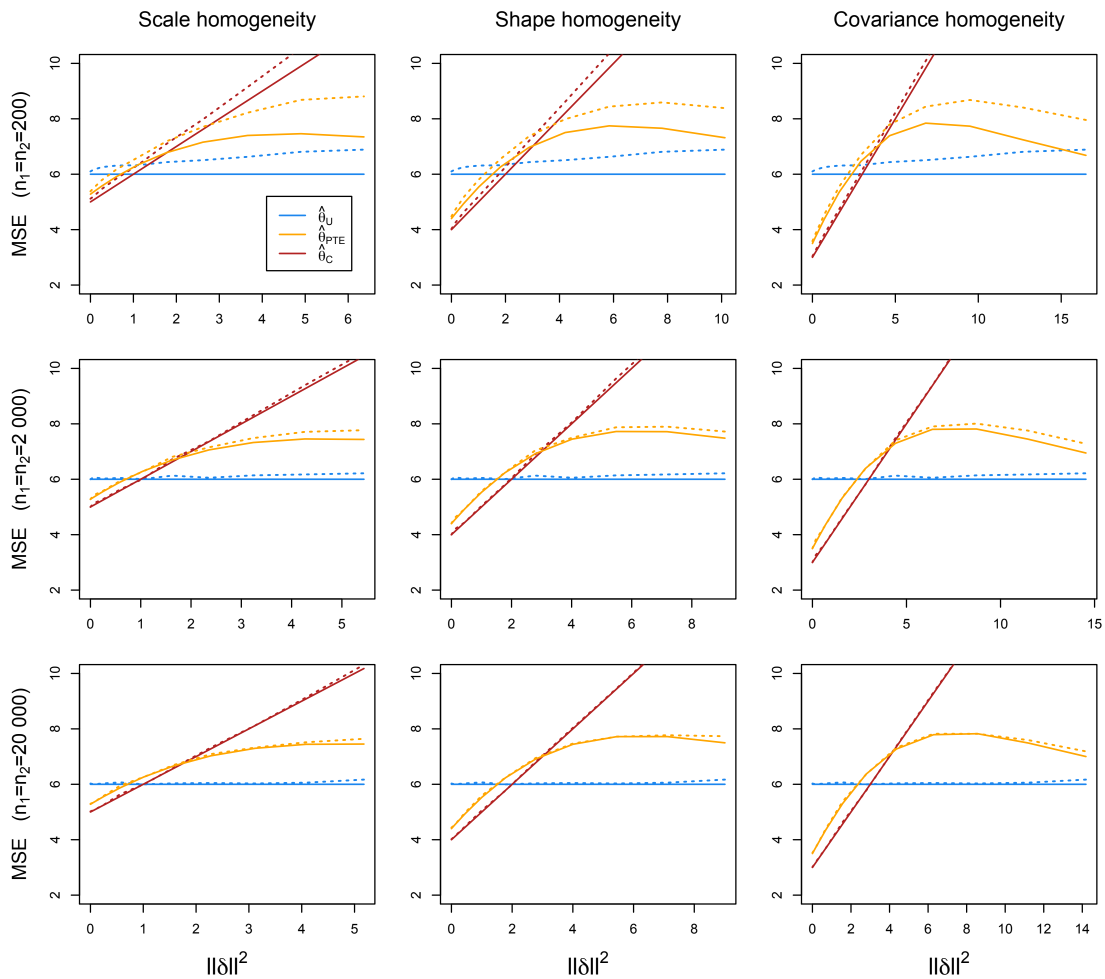

Figure 3: For three different constraints, namely *scale homogeneity* (left), *shape homogeneity* (middle), and *covariance homogeneity* (right), plots, as functions of $\|\boldsymbol{\delta}\|^2$ (which measures distance to the constraint), of the finite-sample MSEs in (4.25) (dotted lines) and of their theoretical asymptotic versions in (3.17) (solid lines) for the corresponding PTE estimator $\hat{\boldsymbol{\theta}}_{\rm PTE}$ (with preliminary tests performed at asymptotic level $\alpha = 5\%$) and for their constrained and unconstrained antecedents, $\hat{\boldsymbol{\theta}}_{\rm C}$ and $\hat{\boldsymbol{\theta}}_{\rm U}$; see Section 4.2.

the following pooled quantities: letting $\hat{\boldsymbol{\Sigma}}_{\rm pool} := (n_1\hat{\boldsymbol{\Sigma}}_{1,{\rm U}} + n_2\hat{\boldsymbol{\Sigma}}_{2,{\rm U}})/n$ be the pooled covariance matrix used as the estimator of the common covariance matrix under $\mathcal{H}_0^c$, the common value of the scale parameter under $\mathcal{H}_0^a$ is estimated by $(\det\hat{\boldsymbol{\Sigma}}_{\rm pool})^{1/p}$ (recall that $p = 6$), whereas the common shape matrix under $\mathcal{H}_0^b$ is estimated by $(\det\hat{\boldsymbol{\Sigma}}_{\rm pool})^{-1/p}\hat{\boldsymbol{\Sigma}}_{\rm pool}$.

Practical relevance of these three PTEs will be shown through the following supervised classification exercise. All considered classifiers perform quadratic discriminant analysis (QDA) based on the sample means $\hat{\boldsymbol{\mu}}_1/\hat{\boldsymbol{\mu}}_2$ above, hence only differ through the estimates $\hat{\boldsymbol{\Sigma}}_1/\hat{\boldsymbol{\Sigma}}_2$ of the group-specific covariance matrices: this leads to four QDA classifiers, namely the one using unconstrained estimators (which is the usual QDA classifier) and those using each of the three PTEs above. To compare the performances of these classifiers, we randomly sampled 30 observations in group 1 and 25 observations in group 2, and we trained the various classifiers on the resulting training set of size 55 (the PTE-based classifiers were applied with asymptotic level $\alpha = 1\%$ and asymptotic level $\alpha = 0.1\%$ for the preliminary tests). The misclassification rate of each classifier was then evaluated on the basis of the test set made of the remaining $n_1 + n_2 - 55 = 30$ observations. To ensure that the results are not specific to a particular partition of the dataset into a training set and a test set, this was repeated $M = 1\,000$ times. Figure 4 provides, for each classifier, a boxplot of the resulting M misclassification rates (the average misclassification rate of each classifier is also given). Clearly, the results indicate that the PTE-based classifiers dominate the classical QDA procedure. For $\alpha = 0.1\%$, the PTE associated with the constraint of covariance homogeneity ($\mathcal{H}_0^c$) provides a classifier that reduces by more than 12% the average misclassification rate of the classical QDA classifier.

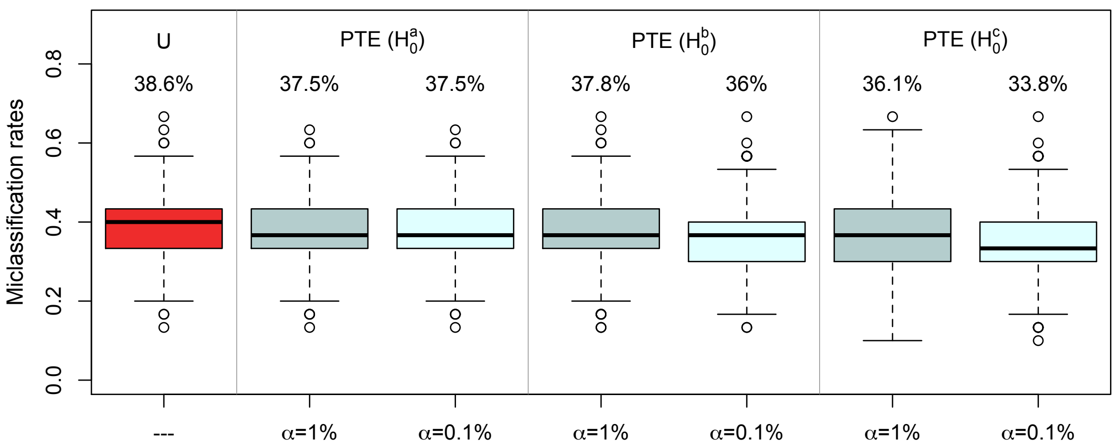

Figure 4: Boxplots of the misclassification rates obtained by applying different QDA classifiers in $M = 1\,000$ random partitions of the twin dataset into training and test sets. The QDA classifiers use different estimators of the group-specific covariance matrices, namely the usual unconstrained estimators (which provides the classical QDA classifier), and the PTEs associated with the constraints of scale homogeneity ($\mathcal{H}_0^a$), shape homogeneity ($\mathcal{H}_0^b$) and covariance homogeneity ($\mathcal{H}_0^c$). For PTEs, two versions were considered, that differ in the nominal level α at which preliminary tests were performed. Percentages above the boxplots indicate average misclassification rates; see Section 5 for details.

Acknowledgements

Davy Paindaveine's research is supported by a research fellowship from the Francqui Foundation and by the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles. Thomas Verdebout's research is supported by the ARC Program of the Université libre de Bruxelles and by the Crédit de Recherche J.0134.18 of the FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique.

Appendix: Proofs

In this appendix, we collect the proofs of the various results.

Proof of Theorem 1. First note that

$$\begin{aligned}\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm PTE}-\boldsymbol{\theta}) &= \lambda(Q^{(n)})\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm C}-\boldsymbol{\theta})+(1-\lambda(Q^{(n)}))\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm U}-\boldsymbol{\theta})\\ &= \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm U}-\boldsymbol{\theta})+\lambda(Q^{(n)})\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm C}-\hat{\boldsymbol{\theta}}_{\rm U}). \end{aligned} \tag{A.26}$$

For any $\varepsilon>0$, Assumption (A)(iii) ensures that

$$\mathrm{P}_{\boldsymbol{\theta}}^{(n)}[\lambda(Q^{(n)})\|\boldsymbol{\nu}_n^{-1}\|>\varepsilon]\leq \mathrm{P}_{\boldsymbol{\theta}}^{(n)}[\lambda(Q^{(n)})=1]\to 0,$$

so that $\lambda(Q^{(n)})\boldsymbol{\nu}_n^{-1}=o_{\rm P}(1)$ under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$. Since by assumption, $\hat{\boldsymbol{\theta}}_{\rm C}-\hat{\boldsymbol{\theta}}_{\rm U}=\hat{\boldsymbol{\theta}}_{\rm C}-\boldsymbol{\theta}+o_{\rm P}(1)=O_{\rm P}(1)$ under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, the result follows from (A.26). □

Proof of Theorem 2. Writing $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm U}-\boldsymbol{\theta}_n)=\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm U}-\boldsymbol{\theta})-\boldsymbol{\nu}_n^{-1}(\boldsymbol{\theta}_n-\boldsymbol{\theta})$ and $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm C}-\boldsymbol{\theta}_n)=\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm C}-\boldsymbol{\theta})-\boldsymbol{\nu}_n^{-1}(\boldsymbol{\theta}_n-\boldsymbol{\theta})$, Assumption (B) entails that

$$\begin{pmatrix}\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm U}-\boldsymbol{\theta}_n)\\ \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\rm C}-\boldsymbol{\theta}_n)\\ \mathbf{D}^{(n)}\\ \Lambda^{(n)}\end{pmatrix}=\begin{pmatrix}\mathbf{A}_{\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\tau}_n\\ \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\tau}_n\\ \mathbf{D}^{(n)}\\ \Lambda^{(n)}\end{pmatrix}+o_{\rm P}(1) \tag{A.27}$$

under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, $\boldsymbol{\theta}\in\boldsymbol{\Theta}_0$. Using Assumption (B) again, we have

$$\begin{pmatrix}\mathbf{A}_{\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\tau}_n\\ \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\tau}_n\\ \mathbf{D}^{(n)}\\ \Lambda^{(n)}\end{pmatrix}+o_{\rm P}(1)\xrightarrow{\mathcal{D}}\mathcal{N}\left(\begin{pmatrix}-\boldsymbol{\tau}\\ -\boldsymbol{\tau}\\ \mathbf{0}\\ -\frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau}\end{pmatrix},\mathbf{F}\right)$$

under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$, $\boldsymbol{\theta}\in\boldsymbol{\Theta}_0$, with

$$\mathbf{F}:=\begin{pmatrix}\mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}' & \mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{B}_{\boldsymbol{\theta}}'\boldsymbol{\Upsilon}' & \mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}' & \mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau}\\ \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}' & \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{B}_{\boldsymbol{\theta}}'\boldsymbol{\Upsilon}' & \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}' & \boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau}\\ \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}' & \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{B}_{\boldsymbol{\theta}}'\boldsymbol{\Upsilon}' & \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}' & \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau}\\ \boldsymbol{\tau}'\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{A}_{\boldsymbol{\theta}}' & \boldsymbol{\tau}'\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{B}_{\boldsymbol{\theta}}'\boldsymbol{\Upsilon}' & \boldsymbol{\tau}'\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}' & \boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau}\end{pmatrix}.$$

Thus, the third Le Cam Lemma (jointly with the fact that (A.27) also holds under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}$, from contiguity) directly yields that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}$,

$$\begin{pmatrix} \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{U}} - \boldsymbol{\theta}_n) \\ \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{C}} - \boldsymbol{\theta}_n) \\ \mathbf{D}^{(n)} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\begin{pmatrix} (\mathbf{A}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}} - \mathbf{I}_p)\boldsymbol{\tau} \\ (\boldsymbol{\Upsilon}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}} - \mathbf{I}_p)\boldsymbol{\tau} \\ \mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau} \end{pmatrix}, \tilde{\mathbf{F}}\right),$$

where $\tilde{\mathbf{F}}$ is obtained from $\mathbf{F}$ by deleting its last column and last row. Since

$$\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{PTE}} - \boldsymbol{\theta}_n) = (1 - \lambda(\|\mathbf{D}\|^2))\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{U}} - \boldsymbol{\theta}_n) + \lambda(\|\mathbf{D}\|^2)\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{C}} - \boldsymbol{\theta}_n),$$

the result then directly follows from the continuous mapping theorem. □

The proof of Proposition 1 requires the following preliminary result.

Lemma 1 (Saleh (2006), pp. 32)**.** *Let $\mathbf{Z}$ be a Gaussian random p-vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{I}_p$. Then, for any real measurable function φ,*

$$(i) \quad \mathrm{E}[\varphi(\|\mathbf{Z}\|^2)\mathbf{Z}] = \mathrm{E}[\varphi(V)]\boldsymbol{\mu}$$

and

$$(ii) \quad \mathrm{E}[\varphi(\|\mathbf{Z}\|^2)\mathbf{Z}\mathbf{Z}'] = \mathrm{E}[\varphi(V))]\mathbf{I}_p + \mathrm{E}[\varphi(W)]\boldsymbol{\mu}\boldsymbol{\mu}',$$

where $V \sim \chi^2_{p+2}(\|\boldsymbol{\mu}\|^2)$ and $W \sim \chi^2_{p+4}(\|\boldsymbol{\mu}\|^2)$.

Proof of Proposition 1. Since $\mathrm{E}[\mathbf{D}] = \mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}$ and since $\mathbf{P}_{\boldsymbol{\Upsilon},\mathrm{eff}}$ is idempotent, we have

$$\begin{aligned} \mathrm{E}[\boldsymbol{\mu}^{\mathrm{Vic}}_{\mathrm{PTE,eff}}] &= \mathrm{E}[\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}((1 - \lambda(\|\mathbf{D}\|^2))\mathbf{D} - \boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau})] \\ &= -\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\mathrm{E}[\lambda(\|\mathbf{D}\|^2)\mathbf{D}]. \end{aligned} \tag{A.28}$$

Since $\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}$ is a projection matrix with rank $p-r$, it decomposes into $\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}} = \mathbf{O}\boldsymbol{\Lambda}\mathbf{O}'$, where $\mathbf{O}$ is a $p \times p$ orthogonal matrix and $\boldsymbol{\Lambda} := \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$ is a diagonal matrix with $\mathrm{tr}[\boldsymbol{\Lambda}] = p-r$. The random vector $\mathbf{E} := \mathbf{O}'\mathbf{D}$ is then Gaussian with mean vector $\boldsymbol{\Lambda}\mathbf{O}'\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}$ and covariance matrix $\boldsymbol{\Lambda}$. Lemma 1(i) thus entails that

$$\mathrm{E}[\lambda(\|\mathbf{D}\|^2)\mathbf{D}] = \mathbf{O}\mathrm{E}[\lambda(\|\mathbf{E}\|^2)\mathbf{E}] = \gamma_2\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}, \tag{A.29}$$

where γ_2 is based on a non-central chi-square distribution with $p-r+2$ degrees of freedom and non-centrality parameter $(\boldsymbol{\Lambda}\mathbf{O}'\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau})'\boldsymbol{\Lambda}\mathbf{O}'\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau} = \boldsymbol{\tau}'\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\theta}}\boldsymbol{\tau}$. Plugging this into (A.28) provides the result for $\mathrm{E}[\boldsymbol{\mu}^{\mathrm{Vic}}_{\mathrm{PTE,eff}}]$.
We thus turn to $\mathrm{Var}[\boldsymbol{\mu}^{\mathrm{Vic}}_{\mathrm{PTE,eff}}]$. Since $(1-\lambda(v))^2 = 1 - \lambda(v)$, we have

$$\mathrm{Var}[\boldsymbol{\mu}^{\mathrm{Vic}}_{\mathrm{PTE,eff}}] = \boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\mathrm{Var}[(1 - \lambda(\|\mathbf{D}\|^2))\mathbf{D}]\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\theta}}$$

$$= \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\big\{\mathrm{E}[(1-\lambda(\|\mathbf{D}\|^2))\mathbf{D}\mathbf{D}'] \qquad (A.30)$$
$$-(1-\gamma_2)^2\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\big\}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2},$$

where we used (A.29). Now, by assumption, $\mathrm{E}[\mathbf{D}\mathbf{D}'] = \mathrm{Var}[\mathbf{D}] + \mathrm{E}[\mathbf{D}](\mathrm{E}[\mathbf{D}])' = \mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}} + \mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}$, and, applying Lemma 1(ii) along the same lines as above, we have that $\mathrm{E}[\lambda(\|\mathbf{D}\|^2)\mathbf{D}\mathbf{D}'] = \mathbf{O}\mathrm{E}[\lambda(\|\mathbf{E}\|^2)\mathbf{E}\mathbf{E}']\mathbf{O}' = \gamma_2\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}} + \gamma_4\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}$. Plugging these expressions into (A.30) then provides the result. □

Proof of Proposition 2. Contiguity implies that (2.4) also holds under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}$, so that

$$\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{U}} - \boldsymbol{\theta}_n) = \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{U}} - \boldsymbol{\theta}) - \boldsymbol{\tau} = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} - \boldsymbol{\tau} + o_{\mathrm{P}}(1)$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}$. Since Le Cam's third lemma entails that $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}$ is asymptotically normal with mean vector $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau}$ and covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}$, it follows that $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{U}} - \boldsymbol{\theta}_n)$ is asymptotically normal with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}$, which yields $\mathrm{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\mathrm{U}}) = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}$. Working along the same lines, we have that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}$,

$$\begin{aligned}\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{C}} - \boldsymbol{\theta}_n) &= \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{C}} - \boldsymbol{\theta}) - \boldsymbol{\tau}\\ &= \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}'\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} - \boldsymbol{\tau} + o_{\mathrm{P}}(1)\\ &= \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} - \boldsymbol{\tau} + o_{\mathrm{P}}(1).\end{aligned}$$

It directly follows that $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{C}} - \boldsymbol{\theta}_n)$ is, still under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n}$, asymptotically normal with mean vector $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\tau} - \boldsymbol{\tau} = -\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}^{\perp}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\tau}$ and covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{P}_{\boldsymbol{\Upsilon},\mathrm{eff}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}$. The expression for $\mathrm{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\mathrm{C}})$ given in Proposition 2 directly follows. □

References

Ahmed, S. E., and Basu, A. K. (2000). Least squares, preliminary test and Stein-type estimation in general vector AR(p) models. *Stat. Neerl.* **54**, 47–66.

Arashi, M., Kibria, B. G., Norouzirad, M., and Nadarajah, S. (2014). Improved preliminary test and Stein-rule Liu estimators for the ill-conditioned elliptical linear regression model. *J. Multivariate Anal.* **126**, 53–74.

Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Statist.* **15**, 190–204.

Bickel, P. J., and Ritov, Y. A. (1996). Inference in hidden Markov models I: Local asymptotic normality in the stationary case. *Bernoulli* **2**, 199–228.

Drost, F. C., Klaassen, C. A., and Werker, B. J. (1997). Adaptive estimation in time-series models. *Ann. Statist.* **25**, 786–817.

Flury, B. (2013). *A first course in multivariate statistics.* Springer Science and Business Media.

Francq, C., and Zakoian, J. M. (2013). Inference in nonstationary asymmetric GARCH models. *Ann. Statist.* **41**, 1970–1998.

Garcia-Portugues, E., Paindaveine, D. and Verdebout, Th. (2020). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *J. Amer. Statist. Assoc.*, **115**, 1873–1887 .

Giles, D. E., Lieberman, O., and Giles, J. A. (1992). The optimal size of a preliminary test of linear restrictions in a misspecified regression model. *J. Amer. Statist. Assoc.* **87**, 1153–1157.

Giles, J. A., and Giles, D. E. (1993). Pre-test estimation and testing in econometrics: recent developments. *J. Econ. Surveys* **7**, 145–197.

Guta, M., and Kiukas, J. (2015). Equivalence classes and local asymptotic normality in system identification for quantum Markov chains. *Comm. Math. Phys.* **335**, 1397–1428.

Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.* **34**, 2707–2756.

Hallin, M. and Paindaveine, D. (2008). Optimal rank-based tests for homogeneity of scatter. *Ann. Statist.* **34**, 1261–1298.

Hallin, M. and Paindaveine, D. (2009). Optimal tests for homogeneity of covariance, scale, and shape. *J. Multivariate Anal.* **29**, 422–444.

Hallin, M., Paindaveine, D. and Verdebout, Th. (2010). Optimal rank-based testing for principal components. *Ann. Statist.* **38**, 3245–3299.

Hallin, M., Paindaveine, D. and Verdebout, Th. (2013). Optimal rank-based tests for common principal components. *Bernoulli* **19**, 2524–2556.

Hallin, M., Paindaveine, D. and Verdebout, Th. (2014). Efficient R-estimation of principal and common principal components. *J. Amer. Statist. Assoc.* **109**, 1071–1083.

Hallin, M., Taniguchi, M., Serroukh, A., and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.* **27**, 2054–2080.

Kahn, J., and Guta, M. (2009). Local asymptotic normality for finite dimensional quantum systems. *Comm. Math. Phys.* **289**, 597–652.

Kibria, B. G., and Saleh, A. M. E. (2004). Preliminary test ridge regression estimators with student's t errors and conflicting test-statistics. *Metrika* **59**, 105–124.

Ley, Chr., Swan, Y, Thiam, B. and Verdebout, Th. (2013) Optimal R-estimation of a spherical location. *Statist. Sinica* **23**, 305–333.

Ley, Chr. and Verdebout, Th. (2017). *Modern Directional Statistics.* Chapman and Hall, CRC press.

Maeyama, Y., Tamaki, K., and Taniguchi, M. (2011). Preliminary test estimation for spectra. *Statist. Probab. Lett.* **81**, 1580–1587.

Ohtani, K., and Toyoda, T. (1980). Estimation of regression coefficients after a preliminary test for homoscedasticity. *J. Econometrics* **12**, 151–159.

Paindaveine, D., Rasoafaraniaina, J. and Verdebout, Th. (2017). Preliminary test estimation for multi-sample principal components. *Econom. Stat.* **2**, 106–116.

Saleh, A. M. E. (2006). *Theory of Preliminary Test and Stein-type Estimation with Applications* (Vol. 517). John Wiley and Sons.

Sen, P. K., and Saleh, A. M. E. (1979). Nonparametric estimation of location parameter after a preliminary test on regression in the multivariate case. *J. Multivariate Anal.* **9**, 322–331.

Sen, P. K., and Saleh, A. E. (1987). On preliminary test and shrinkage M-estimation in linear models. *Ann. Statist.* **15**, 1580–1592.

Tanigushi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series.* Springer.

Wan, A. T., Zou, G., and Ohtani, K. (2006). Further results on optimal critical values of pre-test when estimating the regression error variance. *Econom. J.* **9**, 159–176.