

Preliminary Multiple-Test Estimation, with applications to k -sample Covariance Estimation

Davy Paindaveine^{*,†}, Jos ea Rasoafaraniaina^{*} and Thomas Verdebout^{*}
ECARES and Mathematics Department, Universit e Libre de Bruxelles (ULB)^{*}
Toulouse School of Economics, Universit e Toulouse Capitole[†]

December 18, 2020

Abstract

Multisample covariance estimation—that is, estimation of the covariance matrices associated with k distinct populations—is a classical problem in multivariate statistics. A common solution is to base estimation on the outcome of a test that these covariance matrices show some given pattern. Such a preliminary test may, e.g., investigate whether or not the various covariance matrices are equal to each other (test of homogeneity), or whether or not they have common eigenvectors (test of common principal components), etc. Since it is usually unclear what the possible pattern might be, it is natural to consider a collection of such patterns, leading to a collection of preliminary tests, and to base estimation on the outcome of such a multiple testing rule. In the present work, we therefore study preliminary test estimation based on multiple tests. Since this is of interest also outside k -sample covariance estimation, we do so in a very general framework where it is only assumed that the sequence of models at hand is locally asymptotically normal. In this general setup, we define the proposed estimators and derive their asymptotic properties. We come back to k -sample covariance estimation to illustrate the asymptotic and finite-sample behaviours of our estimators. Finally, we treat a real data example that allows us to show their practical relevance in a supervised classification framework.

Keywords: Covariance matrix estimation; Local asymptotic normality; Multisample problems; Preliminary test estimation; Valid post-selection inference.

1 Introduction

The present paper is motivated by the problem of estimating the covariance matrices $\Sigma_1, \dots, \Sigma_k$ associated with k distinct p -dimensional populations. This is a very classical point estimation problem in multivariate analysis. It is, e.g., of paramount importance when building discriminant analysis rules or when performing MANOVA or MANOCOVA. When $\Sigma_1, \dots, \Sigma_k$ are unconstrained, this multisample problem of course reduces to a collection of k separate estimation problems. In many applications, however, it is assumed or suspected that there is some link between the various covariance matrices. In line with this, [Boente and Orellana \(2004\)](#) and [Jensen and Madsen \(2004\)](#) considered k -sample covariance estimation under the assumption of *proportionality*, that specifies that $\Sigma_1, \dots, \Sigma_k$ are equal to a common covariance matrix Σ up to group-specific scalar factors. [Flury \(1984\)](#) tackled the same estimation problem in the *common principal components (CPC)* model, under which $\Sigma_1, \dots, \Sigma_k$ share the same eigenvectors. Many later works focused on this CPC model: [Flury \(1986\)](#) derived what is now the textbook Gaussian asymptotic theory; [Boente et al. \(2002\)](#) defined robust estimation procedures for this model, whereas [Hallin et al. \(2014\)](#) focused on rank-based estimation; [Browne and McNicholas \(2014\)](#) considered the problem in high dimensions, while functional extensions were proposed in [Benko et al. \(2009\)](#).

Prior to performing k -sample covariance estimation under some specific assumption (proportionality, CPC, etc.), it is of course natural to first perform a test to investigate whether or not the data are compatible with that particular assumption. If one assumes that the k covariance matrices are equal to each other, then one should accordingly perform a test of *homogeneity*, among those from [Schott \(2001\)](#) or [Hallin and Paindaveine \(2009\)](#), etc. For tests of proportionality, one may refer to [Liu et al. \(2014\)](#), [Tsukuda and Matsuura](#)

(2019) and the references therein, while tests for the CPC structure were proposed in Flury (1986), Schott (1991), Boente et al. (2009), and Hallin et al. (2013), to cite only a few.

Estimation is then based on the decision of such a preliminary test, as we now illustrate by considering the assumption of homogeneity. If ϕ_{cov} is a test of the null hypothesis of homogeneity $\mathcal{H}_0^{\text{cov}} : \boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_k$, then, writing $\mathbb{I}[A]$ for the indicator function of A , the resulting natural estimator of $(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_k)$ is

$$\mathbb{I}[\phi_{\text{cov}} = 1](\hat{\boldsymbol{\Sigma}}_1, \dots, \hat{\boldsymbol{\Sigma}}_k) + \mathbb{I}[\phi_{\text{cov}} = 0](\hat{\boldsymbol{\Sigma}}, \dots, \hat{\boldsymbol{\Sigma}}), \quad (1.1)$$

where the $\hat{\boldsymbol{\Sigma}}_1, \dots, \hat{\boldsymbol{\Sigma}}_k$ are “unconstrained” estimators of $\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_k$ and where $\hat{\boldsymbol{\Sigma}}$ is an estimator of the common value of the $\boldsymbol{\Sigma}_\ell$ ’s under the null hypothesis of homogeneity; as usual, $\phi_{\text{cov}} = 1$ (resp., $\phi_{\text{cov}} = 0$) indicates rejection (resp., non-rejection). The estimator in (1.1) is a *preliminary test estimator* (PTE) in the sense of Saleh (2006); we refer to Maeyama et al. (2011) and Paindaveine et al. (2017, 2020) for recent contributions on such estimators. PTEs typically achieve a good compromise between $(\hat{\boldsymbol{\Sigma}}_1, \dots, \hat{\boldsymbol{\Sigma}}_k)$ and $(\hat{\boldsymbol{\Sigma}}, \dots, \hat{\boldsymbol{\Sigma}})$ in the vicinity of $\mathcal{H}_0^{\text{cov}}$ and, provided that ϕ_{cov} is a consistent test, PTEs are also asymptotically equivalent to the classical estimator $(\hat{\boldsymbol{\Sigma}}_1, \dots, \hat{\boldsymbol{\Sigma}}_k)$ “away from the null hypothesis $\mathcal{H}_0^{\text{cov}}$ ”, that is, for fixed parameter values that do not satisfy $\mathcal{H}_0^{\text{cov}}$.

Motivation for the present work lies in the fact that, in many situations, *multiple* constraints may be considered. To provide an example in the above k -sample covariance estimation framework, let us factorize the k covariance matrices as $\boldsymbol{\Sigma}_\ell = \sigma_\ell^2 \mathbf{V}_\ell := (\det \boldsymbol{\Sigma}_\ell)^{1/p} \{ \boldsymbol{\Sigma}_\ell / (\det \boldsymbol{\Sigma}_\ell)^{1/p} \}$ to emphasize their “scale” σ_ℓ and their “shape” \mathbf{V}_ℓ . With this notation, one may consider the constraints associated with the null hypotheses of scale homogeneity $\mathcal{H}_0^{\text{scale}} : \sigma_1^2 = \dots = \sigma_k^2$ and of shape homogeneity $\mathcal{H}_0^{\text{shape}} : \mathbf{V}_1 = \dots = \mathbf{V}_k$ (note that $\mathcal{H}_0^{\text{shape}}$ coincides with the null hypothesis of proportionality). If ϕ_{scale} and ϕ_{shape} are tests for these null hypotheses, then a natural estimator is the *preliminary multiple-test*

estimator (PMTE)

$$\begin{aligned} & \mathbb{I}[\phi_{\text{scale}} = 1, \phi_{\text{shape}} = 1](\hat{\sigma}_1^2 \hat{\mathbf{V}}_1, \dots, \hat{\sigma}_k^2 \hat{\mathbf{V}}_k) + \mathbb{I}[\phi_{\text{scale}} = 1, \phi_{\text{shape}} = 0](\hat{\sigma}_1^2 \hat{\mathbf{V}}, \dots, \hat{\sigma}_k^2 \hat{\mathbf{V}}) \\ & + \mathbb{I}[\phi_{\text{scale}} = 0, \phi_{\text{shape}} = 1](\hat{\sigma}^2 \hat{\mathbf{V}}_1, \dots, \hat{\sigma}^2 \hat{\mathbf{V}}_k) + \mathbb{I}[\phi_{\text{scale}} = 0, \phi_{\text{shape}} = 0](\hat{\sigma}^2 \hat{\mathbf{V}}, \dots, \hat{\sigma}^2 \hat{\mathbf{V}}), \end{aligned} \quad (1.2)$$

where the $\hat{\sigma}_\ell^2$'s and $\hat{\mathbf{V}}_\ell$'s are unconstrained estimators (these are the scale and shape of the unconstrained estimators $\hat{\Sigma}_\ell$), $\hat{\sigma}^2$ is an estimator of the common value of the σ_ℓ^2 's under $\mathcal{H}_0^{\text{scale}}$, and where $\hat{\mathbf{V}}$, similarly, is an estimator of the common value of the \mathbf{V}_ℓ 's under $\mathcal{H}_0^{\text{shape}}$. The estimator in (1.2) is a PTE that involves two constraints, whose intersection is associated with the null hypothesis $\mathcal{H}_0^{\text{cov}}$ of homogeneity of the k covariance matrices. Obviously, more than two constraints may be considered. For instance, factorizing further the covariance matrices into $\Sigma_\ell = \sigma_\ell^2 \mathbf{V}_\ell = \sigma_\ell^2 \mathbf{O}_\ell \mathbf{\Lambda}_\ell \mathbf{O}_\ell'$ based on the usual spectral decomposition of the shape matrix \mathbf{V}_ℓ (here, \mathbf{O}_ℓ is an orthogonal matrix and $\mathbf{\Lambda}_\ell$ is a diagonal matrix collecting the eigenvalues of \mathbf{V}_ℓ on the diagonal), one may consider the constraints associated with the null hypothesis of scale homogeneity $\mathcal{H}_0^{\text{scale}}$, the null hypothesis of shape eigenvalue homogeneity $\mathcal{H}_0^{\text{eig}} : \mathbf{\Lambda}_1 = \dots = \mathbf{\Lambda}_k$, and the CPC null hypothesis $\mathcal{H}_0^{\text{CPC}} : \mathbf{O}_1 = \dots = \mathbf{O}_k$. Combining the outcomes of tests for the three null hypotheses allows one to define a three-constraint PMTE of the same nature as in (1.2). Such an estimator actually formalizes the estimator practitioners would use in practice in the present k -sample covariance estimation setup. It remains unclear, however, how this estimator behaves since, to the best of our knowledge, such PMTEs have not been considered in the literature.

The main objective of the present work is therefore to introduce and to study PMTEs, and show their practical relevance. The outline of the paper is as follows. We first introduce PMTEs in a general context (Section 2). Then, we derive the asymptotic behaviour of

such estimators in the common framework of locally asymptotically normal (LAN) models (Section 3). In particular, we show that away from all constraints, there is no asymptotic loss to consider PMTEs rather than standard PTEs, since both types of estimators then turn out to be asymptotically equivalent to the unconstrained estimator. As we show, however, PMTEs dominate their competitors in the vicinity of the considered constraints. To demonstrate the practical relevance of PMTEs, we mainly focus on the multisample covariance estimation problem that motivated this work (Section 4). We first derive the various estimators in this context (Section 4.1), then compare them theoretically and empirically, respectively through the computation of asymptotic efficiencies and through simulations (Section 4.2). Last, we illustrate our methodology in a real data example involving different species of voles (Section 4.3). We conclude with final comments (Section 5). In a supplementary material, we perform finite-sample comparisons with estimators resulting from a BIC-based model selection (Section A) and provide the technical proofs (Section B).

2 Preliminary multiple-test estimators (PMTEs)

Consider a model that is indexed by a parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d) \in \boldsymbol{\Theta} \subset \mathbb{R}^d$ of dimension d and assume that, in line with the situation considered in the introduction, m possible constraints on $\boldsymbol{\theta}$ are suspected to hold. For the sake of clarity, we first introduce PMTEs in the case involving $m = 2$ constraints only. Since we will actually restrict throughout to linear constraints on $\boldsymbol{\theta}$, these two constraints take the form $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_j)$, $j = 1, 2$, for some $d \times r_j$ matrices $\boldsymbol{\Upsilon}_j$ ($r_j < d$), where $\mathcal{M}(\mathbf{A})$ denotes the vector subspace of \mathbb{R}^d that is spanned by the columns of \mathbf{A} (without any loss of generality, we will assume in the sequel that the $\boldsymbol{\Upsilon}_j$'s have full rank). To make the notation lighter, we will throughout tacitly restrict to values of $\boldsymbol{\theta}$ that belong to $\boldsymbol{\Theta}$, which allows us to write $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_j)$ instead of $\boldsymbol{\theta} \in \boldsymbol{\Theta} \cap \mathcal{M}(\boldsymbol{\Upsilon}_j)$,

or $\boldsymbol{\theta} \in \mathbb{R}^d \setminus \mathcal{M}(\boldsymbol{\Upsilon}_j)$ instead of $\boldsymbol{\theta} \in \Theta \setminus \mathcal{M}(\boldsymbol{\Upsilon}_j)$, etc. For $d = 2$, the two constraints are vectorial lines that are respectively spanned by the d -vectors $\boldsymbol{\Upsilon}_j$, $j = 1, 2$. In this framework, we assume, for each $j = 1, 2$, that a test ϕ_j for the null hypothesis $\mathcal{H}_{j0} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_j)$ is available. The outcome of these tests is coded as $\boldsymbol{\phi} := (\phi_1, \phi_2) \in \{0, 1\}^2$; as in the introduction, $\phi_j = 1$ (resp., $\phi_j = 0$) indicates that ϕ_j leads to rejection (resp., non-rejection) of \mathcal{H}_{j0} . If $\boldsymbol{\phi} = (1, 1)$, then it is natural to adopt an unconstrained estimator $\hat{\boldsymbol{\theta}}_U$ (taking values in \mathbb{R}^d), whereas other values of $\boldsymbol{\phi}$ would lead to considering various constrained estimators of $\boldsymbol{\theta}$, namely a constrained estimator $\hat{\boldsymbol{\theta}}_{(0,1)}$ taking values in $\mathcal{M}(\boldsymbol{\Upsilon}_1)$ if $\boldsymbol{\phi} = (0, 1)$, a constrained estimator $\hat{\boldsymbol{\theta}}_{(1,0)}$ taking values in $\mathcal{M}(\boldsymbol{\Upsilon}_2)$ if $\boldsymbol{\phi} = (1, 0)$, and a constrained estimator $\hat{\boldsymbol{\theta}}_{(0,0)}$ taking values in $\mathcal{M}(\boldsymbol{\Upsilon}_1) \cap \mathcal{M}(\boldsymbol{\Upsilon}_2)$ if $\boldsymbol{\phi} = (0, 0)$. Summing up, this leads to the PMTE

$$\hat{\boldsymbol{\theta}}_{\text{PMTE}} := \mathbb{I}[\boldsymbol{\phi} = (1, 1)]\hat{\boldsymbol{\theta}}_U + \mathbb{I}[\boldsymbol{\phi} = (0, 1)]\hat{\boldsymbol{\theta}}_{(0,1)} + \mathbb{I}[\boldsymbol{\phi} = (1, 0)]\hat{\boldsymbol{\theta}}_{(1,0)} + \mathbb{I}[\boldsymbol{\phi} = (0, 0)]\hat{\boldsymbol{\theta}}_{(0,0)}, \quad (2.3)$$

which is obtained by taking into account the $2^m = 4$ possible (joint) outcomes of the tests ϕ_1 and ϕ_2 .

We now discuss the general case involving an arbitrary number m of linear constraints. To do so, let $\boldsymbol{\Upsilon}_j$, $j = 1, \dots, m$, be full-rank $d \times r_j$ ($r_j < d$) matrices such that the j th constraint takes the form $\mathcal{H}_{j0} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_j)$ and let ϕ_j be a test for the corresponding null hypothesis. Any element $\mathbf{c} = (c_1, \dots, c_m) \in \mathcal{C}^m := \{0, 1\}^m$ may be used to indicate which constraints are satisfied by a given parameter value $\boldsymbol{\theta}$: letting $\mathcal{J}_{\mathbf{c}} = \{j = 1, \dots, m : c_j = 0\}$, $\boldsymbol{\theta} \in \bigcap_{j \in \mathcal{J}_{\mathbf{c}}} \mathcal{M}(\boldsymbol{\Upsilon}_j)$ means that $\boldsymbol{\theta}$ meets the constraints indexed by $\mathcal{J}_{\mathbf{c}}$ but not the other ones. Any $\mathbf{c} \in \mathcal{C}^m$ also corresponds to a possible decision for the m -tuple of tests $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)$. With this notation, the resulting PMTE is then

$$\hat{\boldsymbol{\theta}}_{\text{PMTE}} := \sum_{\mathbf{c} \in \mathcal{C}^m} \mathbb{I}[\boldsymbol{\phi} = \mathbf{c}] \hat{\boldsymbol{\theta}}_{\mathbf{c}}, \quad (2.4)$$

where $\hat{\boldsymbol{\theta}}_{\mathbf{c}}$ is a constrained estimator taking values in $\bigcap_{j \in \mathcal{J}_{\mathbf{c}}} \mathcal{M}(\boldsymbol{\Upsilon}_j)$. In the sequel, it will be

convenient to fix, for any $\mathbf{c} \in \mathcal{C}^m$, a full-rank matrix $\mathbf{\Upsilon}_{\mathbf{c}}$ such that $\mathcal{M}(\mathbf{\Upsilon}_{\mathbf{c}}) = \cap_{j \in \mathcal{J}_{\mathbf{c}}} \mathcal{M}(\mathbf{\Upsilon}_j)$. Note that for $\mathbf{c} = \mathbf{1}_m - \mathbf{e}_j$, where $\mathbf{1}_m$ is the m -vector of ones and \mathbf{e}_j is the j th vector of the canonical basis of \mathbb{R}^m , we may simply take $\mathbf{\Upsilon}_{\mathbf{c}} = \mathbf{\Upsilon}_j$ (so that $\hat{\boldsymbol{\theta}}_{\mathbf{c}}$ is a constrained estimator based on the j th constraint only), whereas for $\mathbf{c} = \mathbf{1}_m$, then we may take $\mathbf{\Upsilon}_{\mathbf{c}} = \mathbf{I}_d$ (so that $\hat{\boldsymbol{\theta}}_{\mathbf{c}} = \boldsymbol{\theta}_{\text{U}}$ is an unconstrained estimator), where \mathbf{I}_ℓ stands for the ℓ -dimensional identity matrix. Clearly, for $m = 2$, the PMTE in (2.4) reduces to the one in (2.3).

3 Asymptotic results

The objective of this section is to derive the asymptotic behaviour of the PMTE in (2.4) under mild assumptions. We first describe the assumptions we will adopt.

3.1 Assumptions

Our first assumption imposes regularity of the parametric model $\{\mathbb{P}_{\boldsymbol{\theta}}^{(n)} : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d\}$ at hand. More precisely, it requires that this model is *locally asymptotically normal (LAN)* (throughout, all convergences are as $n \rightarrow \infty$ and $\xrightarrow{\mathcal{D}}$ denotes weak convergence).

Assumption (A). There exist a deterministic sequence $(\boldsymbol{\nu}_n)$ of full-rank $d \times d$ matrices converging to zero (in Frobenius norm, say), a sequence of random d -vectors $\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)}$ (the *central sequence*) and a symmetric positive semidefinite $d \times d$ matrix $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$ (the *information matrix*), such that, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and any bounded sequence $(\boldsymbol{\tau}_n)$ in \mathbb{R}^d ,

$$(i) \quad \Lambda^{(n)} := \log \frac{d\mathbb{P}_{\boldsymbol{\theta} + \boldsymbol{\nu}_n \boldsymbol{\tau}_n}^{(n)}}{d\mathbb{P}_{\boldsymbol{\theta}}^{(n)}} = \boldsymbol{\tau}_n' \boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} - \frac{1}{2} \boldsymbol{\tau}_n' \boldsymbol{\Gamma}_{\boldsymbol{\theta}} \boldsymbol{\tau}_n + o_{\mathbb{P}}(1) \quad (3.5)$$

and (ii) $\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\theta}})$ under $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$.

This assumption is very mild as it is satisfied by numerous parametric models, including hidden Markov models (Bickel and Ritov, 1996), many time series models such as, e.g., ARMA and GARCH ones (Francq and Zakoïan, 2013), location-scatter elliptical models (Hallin and Paindaveine, 2006), multisample elliptical models (Hallin et al., 2013, Hallin et al., 2014), models for directional data (Ley et al., 2013, García-Portugués et al., 2020), and many more. Actually, any parametric model with smooth densities and continuous Fisher information satisfies Assumption (A); see, e.g., Theorems 12.2.2–12.2.3 in Lehmann and Romano (2005).

Our second assumption, that relates to the constrained/unconstrained estimators and preliminary tests to be used in the proposed PMTE, is as follows (throughout, $\text{tr}[\mathbf{A}]$ will denote the trace of \mathbf{A} and $\chi_{\ell,1-\alpha}^2$ will stand for the upper α -quantile of the chi-square distribution with ℓ degrees of freedom).

Assumption (B). The estimators $\hat{\boldsymbol{\theta}}_{\mathbf{c}}$, $\mathbf{c} \in \mathcal{C}^m$, and preliminary tests ϕ_j , $j = 1, \dots, m$, satisfy the following properties:

- (i) for any $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\mathbf{c}})$, we have $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta}) = \boldsymbol{\Upsilon}_{\mathbf{c}}\mathbf{B}_{\mathbf{c},\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}^{(n)} + o_{\mathbb{P}}(1)$ under $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$ for some $d \times d$ matrix $\mathbf{B}_{\mathbf{c},\boldsymbol{\theta}}$ and some sequence of random d -vectors $\mathbf{S}_{\boldsymbol{\theta}}^{(n)}$ such that

$$\begin{pmatrix} \mathbf{S}_{\boldsymbol{\theta}}^{(n)} \\ \boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} & \boldsymbol{\Omega}_{\boldsymbol{\theta}} \\ \boldsymbol{\Omega}_{\boldsymbol{\theta}} & \boldsymbol{\Gamma}_{\boldsymbol{\theta}} \end{pmatrix}\right) \quad (3.6)$$

under $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$ (here, $\boldsymbol{\nu}_n^{-1}$ and $\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)}$ are as in Assumption (A));

- (ii) the test ϕ_j rejects the null hypothesis $\mathcal{H}_{j0} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_j)$ at asymptotic level α when $Q_j^{(n)} := \|\mathbf{D}_j^{(n)}\|^2 := (\mathbf{D}_j^{(n)})'\mathbf{D}_j^{(n)} > \chi_{d-r_j,1-\alpha}^2$, where, for any $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_j)$, the random d -vector $\mathbf{D}_j^{(n)}$ is such that $\mathbf{D}_j^{(n)} = \mathbf{C}_{j,\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}^{(n)} + o_{\mathbb{P}}(1)$ under $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$, with $\mathbf{S}_{\boldsymbol{\theta}}^{(n)}$ as in (i) and with $\mathbf{C}_{j,\boldsymbol{\theta}}$ a $d \times d$ matrix satisfying $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}_{j,\boldsymbol{\theta}}'\mathbf{C}_{j,\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}_{j,\boldsymbol{\theta}}'\mathbf{C}_{j,\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{C}_{j,\boldsymbol{\theta}}'\mathbf{C}_{j,\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$

and $\text{tr}[\mathbf{C}'_{j,\boldsymbol{\theta}}\mathbf{C}_{j,\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}] = d - r_j$ for any $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_j)$. Furthermore, ϕ_j is consistent under $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$ for any $\boldsymbol{\theta} \notin \mathcal{M}(\boldsymbol{\Upsilon}_j)$.

Assumption (B) might look restrictive but it is actually extremely mild: provided that Assumption (A) holds, it indeed only requires the existence of an unconstrained estimator $\hat{\boldsymbol{\theta}}_{\text{U}}$ admitting a Bahadur-type representation. To see this, assume for the sake of simplicity that $\boldsymbol{\nu}_n$ is as usual given by $\boldsymbol{\nu}_n = n^{-1/2}\mathbf{I}_d$ (extension to a general $\boldsymbol{\nu}_n$ is trivial), and that, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the estimator $\hat{\boldsymbol{\theta}}_{\text{U}}$ satisfies the Bahadur representation

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{U}} - \boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{T}_{i,\boldsymbol{\theta}}^{(n)} + o_{\mathbb{P}}(1) \quad (3.7)$$

under $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$, where the random d -vectors $\mathbf{T}_{i,\boldsymbol{\theta}}^{(n)}$, $i = 1, \dots, n$, are mutually independent and share a common distribution that has mean zero and has finite second-order moments. Since the CLT for triangular arrays entails that $\mathbf{S}_{\boldsymbol{\theta}}^{(n)} := n^{-1/2} \sum_{i=1}^n \mathbf{T}_{i,\boldsymbol{\theta}}^{(n)}$ satisfies (3.6) under the usual mild Lévy–Lindeberg condition, this already ensures that Assumption (B)(i) holds for $\mathbf{c} = \mathbf{1}_m$, with $\mathbf{B}_{\mathbf{1}_m,\boldsymbol{\theta}} := \mathbf{I}_d$ (recall that $\boldsymbol{\Upsilon}_{\mathbf{1}_m} = \mathbf{I}_d$). Now, for any other $\mathbf{c} \in \mathcal{C}^m$, the constrained estimator $\hat{\boldsymbol{\theta}}_{\mathbf{c}} := \mathbf{P}_{\mathbf{c}}\hat{\boldsymbol{\theta}}_{\text{U}}$, where $\mathbf{P}_{\mathbf{c}} := \boldsymbol{\Upsilon}_{\mathbf{c}}(\boldsymbol{\Upsilon}'_{\mathbf{c}}\boldsymbol{\Upsilon}_{\mathbf{c}})^{-1}\boldsymbol{\Upsilon}'_{\mathbf{c}}$ is the matrix of the orthogonal projection onto the constraint $\mathcal{M}(\boldsymbol{\Upsilon}_{\mathbf{c}})$, is such that, for any $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\mathbf{c}})$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta}) = \mathbf{P}_{\mathbf{c}}\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{U}} - \boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{P}_{\mathbf{c}}\mathbf{T}_{i,\boldsymbol{\theta}}^{(n)} + o_{\mathbb{P}}(1) = \mathbf{P}_{\mathbf{c}}\mathbf{S}_{\boldsymbol{\theta}}^{(n)} + o_{\mathbb{P}}(1)$$

under $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$, which shows that Assumption (B)(i) is fulfilled. As for Assumption (B)(ii), it will be satisfied by Wald tests for $\mathcal{H}_{j0} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_j)$ constructed in the usual way from (3.7). This confirms that the only key point in Assumption (B) is the existence of an unconstrained estimator $\hat{\boldsymbol{\theta}}_{\text{U}}$ satisfying (3.7). In virtually all models, M-, L-, and R-estimation will provide an unconstrained estimator of this type, so that Assumption (B) is indeed extremely mild.

In the LAN framework of Assumption (A), one may alternatively want to rely on (constrained) asymptotically efficient estimation. The resulting estimators of $\boldsymbol{\theta}$ will satisfy Assumption (B)(i) with $\mathbf{S}_{\boldsymbol{\theta}}^{(n)} = \boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)}$ (hence with $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \boldsymbol{\Omega}_{\boldsymbol{\theta}} = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}$), since, for the constraint associated with $\mathbf{c} \in \mathcal{C}^m$ and any corresponding $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\mathbf{c}})$, the constrained asymptotically efficient estimator $\hat{\boldsymbol{\theta}}_{\mathbf{c}}$ is such that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta}) = \boldsymbol{\Upsilon}_{\mathbf{c}}(\boldsymbol{\Upsilon}'_{\mathbf{c}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon}_{\mathbf{c}})^{-1}\boldsymbol{\Upsilon}'_{\mathbf{c}}\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} + o_{\mathbb{P}}(1)$$

under $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$ (for unconstrained estimation, this provides the usual result $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathbf{U}} - \boldsymbol{\theta}) = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} + o_{\mathbb{P}}(1)$ under $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$ for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$). For testing $\mathcal{H}_0 : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\mathbf{c}})$ against $\mathcal{H}_1 : \boldsymbol{\theta} \notin \mathcal{M}(\boldsymbol{\Upsilon}_{\mathbf{c}})$, the locally asymptotically most stringent test rejects $\mathcal{H}_0 : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\mathbf{c}})$ at asymptotic level α when

$$Q_j^{(n)} := \|\mathbf{C}_{j,\hat{\boldsymbol{\theta}}_{\mathbf{c}}}\boldsymbol{\Delta}_{\hat{\boldsymbol{\theta}}_{\mathbf{c}}}^{(n)}\|^2 > \chi_{d-r_j,1-\alpha}^2, \quad \text{with } \mathbf{C}_{j,\boldsymbol{\theta}} := (\mathbf{I}_d - \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{\Upsilon}_{\mathbf{c}}(\boldsymbol{\Upsilon}'_{\mathbf{c}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Upsilon}_{\mathbf{c}})^{-1}\boldsymbol{\Upsilon}'_{\mathbf{c}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2})\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2};$$

see, e.g., Chapter 5 of [Ley and Verdebout \(2017\)](#). Remarkably, Assumption B(ii) then holds with $\mathbf{D}_j^{(n)} := \mathbf{C}_{j,\hat{\boldsymbol{\theta}}_{\mathbf{c}}}\boldsymbol{\Delta}_{\hat{\boldsymbol{\theta}}_{\mathbf{c}}}^{(n)}$, provided that, under any $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$, with $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\mathbf{c}})$,

$$\mathbf{C}_{j,\hat{\boldsymbol{\theta}}_{\mathbf{c}}}\boldsymbol{\Delta}_{\hat{\boldsymbol{\theta}}_{\mathbf{c}}}^{(n)} = \mathbf{C}_{j,\boldsymbol{\theta}}\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} + o_{\mathbb{P}}(1) \tag{3.8}$$

(direct computations indeed provide $\mathbf{C}'_{j,\boldsymbol{\theta}}\mathbf{C}_{j,\boldsymbol{\theta}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\mathbf{C}'_{j,\boldsymbol{\theta}}\mathbf{C}_{j,\boldsymbol{\theta}} = \mathbf{C}'_{j,\boldsymbol{\theta}}\mathbf{C}_{j,\boldsymbol{\theta}}$ and $\text{tr}[\mathbf{C}'_{j,\boldsymbol{\theta}}\mathbf{C}_{j,\boldsymbol{\theta}}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}] = d - r_j$). Note that for locally and asymptotically discrete estimators $\hat{\boldsymbol{\theta}}_{\mathbf{c}}$, (3.8) holds in particular when the LAN property in Assumption (A) is reinforced into a ULAN (uniform local asymptotic normality) one; see, e.g., Chapter 5 of [Ley and Verdebout, 2017](#) for details.

These considerations explain that Assumptions (A)–(B) will hold for a wide variety of models and corresponding estimators, including efficient estimators for the large class of Gaussian processes considered in [Dahlhaus \(1989\)](#), efficient estimators of regression models with long memory disturbances ([Hallin et al., 1999](#)), adaptive estimators in semiparametric

ARMA, ARCH or TAR time series models (Drost et al., 1997), M-estimators (Lopuhaa, 1992; Paindaveine and Van Bever, 2014) and R-estimators (Hallin et al., 2006, 2014) of scatter in elliptical models, efficient estimators and R-estimators of location in rotationally symmetric models for directional data (Ley et al., 2013), to cite only a few.

3.2 Asymptotic results

We now study the asymptotic behaviour of the PMTE estimator in (2.4). Fix $\mathbf{c}_0 \in \mathcal{C}^m$ and consider a parameter value $\boldsymbol{\theta}$ that is such that $\boldsymbol{\theta} \notin \mathcal{M}(\boldsymbol{\Upsilon}_j)$ for any j such that $c_{0j} = 1$. Using the notation introduced in Section 2, this rewrites $\boldsymbol{\theta} \in \mathbb{R}^d \setminus \cup_{j \notin \mathcal{J}_{\mathbf{c}_0}} \mathcal{M}(\boldsymbol{\Upsilon}_j)$. Consider then the oracle PMTE, $\hat{\boldsymbol{\theta}}_{\text{PMTE}, \mathbf{c}_0}$ say, that would be the natural one to consider if it would be known that $\boldsymbol{\theta} \in \mathbb{R}^d \setminus \cup_{j \notin \mathcal{J}_{\mathbf{c}_0}} \mathcal{M}(\boldsymbol{\Upsilon}_j)$, that is, the PMTE that does not involve tests of the constraints that are known not to be met. Letting $\mathcal{C}_{\mathbf{c}_0}^m := \{\mathbf{c} \in \mathcal{C}^m : c_j = 1 \text{ for any } j \notin \mathcal{J}_{\mathbf{c}_0}\}$, this oracle PMTE is given by

$$\hat{\boldsymbol{\theta}}_{\text{PMTE}, \mathbf{c}_0} := \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}] \hat{\boldsymbol{\theta}}_{\mathbf{c}}, \quad (3.9)$$

where $\boldsymbol{\phi}_j =_{\mathbf{c}_0} \mathbf{c}$ means that $\phi_j = c_j$ for any $j \in \mathcal{J}_{\mathbf{c}_0}$ (so that, as intended, this oracle PMTE does not involve the tests for the constraints that are known not to be met). We then have the following result.

Theorem 3.1. *Let Assumptions (A) and (B) hold. Fix $\mathbf{c}_0 \in \mathcal{C}^m$. Then, for any $\boldsymbol{\theta} \in \mathbb{R}^d \setminus \cup_{j \notin \mathcal{J}_{\mathbf{c}_0}} \mathcal{M}(\boldsymbol{\Upsilon}_j)$,*

$$\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}} - \boldsymbol{\theta}) = \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}, \mathbf{c}_0} - \boldsymbol{\theta}) + o_{\text{P}}(1)$$

as $n \rightarrow \infty$ under $\text{P}_{\boldsymbol{\theta}}^{(n)}$.

This result, that interestingly only requires consistency of the tests ϕ_j , $j = 1, \dots, m$, shows that the proposed PMTE is asymptotically equivalent in probability to the oracle PMTE constructed on the basis of the constraints associated with \mathbf{c}_0 as soon as the true value of the parameter $\boldsymbol{\theta}$ is fixed outside these constraints. Note that taking $\mathbf{c}_0 = \mathbf{1}_m$ readily yields the following corollary.

Corollary 3.1. *Let Assumptions (A) and (B) hold. Then, for any $\boldsymbol{\theta} \in \mathbb{R}^d \setminus \cup_{j=1}^m \mathcal{M}(\Upsilon_j)$,*

$$\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}} - \boldsymbol{\theta}) = \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{U}} - \boldsymbol{\theta}) + o_{\text{P}}(1)$$

as $n \rightarrow \infty$ under $\text{P}_{\boldsymbol{\theta}}^{(n)}$.

It directly follows from this result that, away from the constraints, there is no cost, asymptotically, to use the proposed PMTE rather than its unconstrained antecedent $\hat{\boldsymbol{\theta}}_{\text{U}}$. Finally we stress that it is only in the very particular case $\#\mathcal{J}_{\mathbf{c}_0} = 1$ that the oracle PMTE of Theorem 3.1 is a standard single-constraint PTE, with a known asymptotic behaviour; consequently, in cases where $\#\mathcal{J}_{\mathbf{c}_0} > 1$, Theorem 3.1 on its own does not allow one to deduce the asymptotic behaviour of $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ from the single-constraint theory.

We thus turn to the study of this asymptotic behaviour in the general case. We will actually derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ in two types of asymptotic scenarios, namely under fixed parameter values meeting (at least) one constraint or under sequences of local perturbations of such fixed parameter values. To be more precise, fix again $\mathbf{c}_0 \in \mathcal{C}^m \setminus \{\mathbf{1}_m\}$ and consider now a parameter value $\boldsymbol{\theta}$ that meets the corresponding constraints; in other words, $\boldsymbol{\theta} \in \mathcal{M}(\Upsilon_{\mathbf{c}_0})$. The aforementioned local perturbations are then of the form $\boldsymbol{\theta}_n = \boldsymbol{\theta} + \boldsymbol{\nu}_n \boldsymbol{\tau}_n$, where $\boldsymbol{\tau}_n$ is a bounded sequence in \mathbb{R}^d . Since we do not exclude the case $\boldsymbol{\tau}_n \equiv \mathbf{0}$, this actually also covers the case for which $\boldsymbol{\theta} \in \mathcal{M}(\Upsilon_{\mathbf{c}_0})$ is fixed. To describe the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ in these asymptotic scenarios, we need to

introduce the following notation. Recalling that $\mathcal{J}_{\mathbf{c}_0} = \{j = 1, \dots, m : c_{0j} = 0\}$, we will denote the (ordered) elements of $\mathcal{J}_{\mathbf{c}_0}$ as j_1, \dots, j_s , and the corresponding 2^s elements of $\mathcal{C}_{\mathbf{c}_0}^m = \{\mathbf{c} \in \mathcal{C}^m : c_j = 1 \text{ for any } j \notin \mathcal{J}_{\mathbf{c}_0}\}$ as $\mathbf{c}_1, \dots, \mathbf{c}_{2^s}$ (the ordering is here arbitrary but fixed). Based on this notation, let further

$$\mathbf{F} := \begin{pmatrix} \Upsilon_{\mathbf{c}_1} \mathbf{B}_{\mathbf{c}_1, \boldsymbol{\theta}} \\ \vdots \\ \Upsilon_{\mathbf{c}_{2^s}} \mathbf{B}_{\mathbf{c}_{2^s}, \boldsymbol{\theta}} \end{pmatrix}, \quad \mathbf{G} := \begin{pmatrix} \mathbf{C}_{j_1, \boldsymbol{\theta}} \\ \vdots \\ \mathbf{C}_{j_s, \boldsymbol{\theta}} \end{pmatrix}, \quad \text{and } \mathbf{W}(\mathbf{D}) := \begin{pmatrix} \mathbb{I}[\boldsymbol{\phi}(\mathbf{D}) =_{\mathbf{c}_0} \mathbf{c}_1] \\ \vdots \\ \mathbb{I}[\boldsymbol{\phi}(\mathbf{D}) =_{\mathbf{c}_0} \mathbf{c}_{2^s}] \end{pmatrix} \otimes \mathbf{I}_d, \quad (3.10)$$

where $\mathbf{D} := (\mathbf{D}'_1, \dots, \mathbf{D}'_s)'$ takes values in \mathbb{R}^{sd} and where, for any $\ell = 1, \dots, s$, the test $\phi_{j_\ell}(\mathbf{D})$ (in the test vector $\boldsymbol{\phi}(\mathbf{D})$) is defined as $\phi_{j_\ell}(\mathbf{D}) := \mathbb{I}[\|\mathbf{D}_\ell\|^2 > \chi_{d-r_{j_\ell}, 1-\alpha}^2]$ (note that only the tests $\phi_{j_\ell}(\mathbf{D})$, $\ell = 1, \dots, s$, are involved in $\mathbf{W}(\mathbf{D})$). We then have the following result.

Theorem 3.2. *Let Assumptions (A) and (B) hold. Fix $\mathbf{c}_0 \in \mathcal{C}^m \setminus \{\mathbf{1}_m\}$ and $\boldsymbol{\theta} \in \mathcal{M}(\Upsilon_{\mathbf{c}_0})$. Consider the sequence of local perturbations $\boldsymbol{\theta}_n = \boldsymbol{\theta} + \boldsymbol{\nu}_n \boldsymbol{\tau}_n$, where $(\boldsymbol{\tau}_n)$ is a sequence in \mathbb{R}^d converging to $\boldsymbol{\tau}$. Let \mathbf{R} and $\mathbf{D}_{\mathbf{c}_0}$ be random vectors whose joint distribution is described as*

$$\begin{pmatrix} \mathbf{R} \\ \mathbf{D}_{\mathbf{c}_0} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{F} \boldsymbol{\Omega}_\theta \boldsymbol{\tau} - (\mathbf{1}_{2^s} \otimes \boldsymbol{\tau}) \\ \mathbf{G} \boldsymbol{\Omega}_\theta \boldsymbol{\tau} \end{pmatrix}, \begin{pmatrix} \mathbf{F} \boldsymbol{\Sigma}_\theta \mathbf{F}' & \mathbf{F} \boldsymbol{\Sigma}_\theta \mathbf{G}' \\ \mathbf{G} \boldsymbol{\Sigma}_\theta \mathbf{F}' & \mathbf{G} \boldsymbol{\Sigma}_\theta \mathbf{G}' \end{pmatrix} \right).$$

Then, under $\mathbb{P}_{\boldsymbol{\theta}_n}^{(n)}$, the sequence of random d -vectors $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}} - \boldsymbol{\theta}_n)$ converges weakly to $\mathbf{W}'\mathbf{R}$, with $\mathbf{W} := \mathbf{W}(\mathbf{D}_{\mathbf{c}_0})$.

In this result, $\mathbf{W}'\mathbf{R}$, conditional on $\mathbf{D}_{\mathbf{c}_0}$, is asymptotically normal with mean vector

$$\boldsymbol{\mu}_{\text{PMTE}} = \mathbf{W}' \{ \mathbf{F} \boldsymbol{\Omega}_\theta \boldsymbol{\tau} - (\mathbf{1}_{2^s} \otimes \boldsymbol{\tau}) + \mathbf{F} \boldsymbol{\Sigma}_\theta \mathbf{G}' (\mathbf{G} \boldsymbol{\Sigma}_\theta \mathbf{G}')^{-1} (\mathbf{D}_{\mathbf{c}_0} - \mathbf{G} \boldsymbol{\Omega}_\theta \boldsymbol{\tau}) \} \quad (3.11)$$

and covariance matrix

$$\boldsymbol{\Gamma}_{\text{PMTE}} = \mathbf{W}' \mathbf{F} \boldsymbol{\Sigma}_\theta^{1/2} \{ \mathbf{I}_p - \mathbf{P}(\boldsymbol{\Sigma}_\theta^{1/2} \mathbf{G}') \} \boldsymbol{\Sigma}_\theta^{1/2} \mathbf{F}' \mathbf{W}, \quad (3.12)$$

with $\mathbf{P}(\mathbf{A}) := \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ (throughout, \mathbf{A}^{-} denotes the Moore–Penrose inverse of \mathbf{A}). This allows us to obtain the unconditional asymptotic distribution of $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}} - \boldsymbol{\theta}_n)$ under $\mathbb{P}_{\boldsymbol{\theta}_n}^{(n)}$: indeed, Theorem 3.2 implies that $\mathbf{D}_{\mathbf{c}_0}$ is asymptotically normal with mean vector $\boldsymbol{\mu}_{\mathbf{D}_{\mathbf{c}_0}} := \mathbf{G}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{D}_{\mathbf{c}_0}} = \mathbf{G}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{G}'$ under $\mathbb{P}_{\boldsymbol{\theta}_n}^{(n)}$, so that, under the same sequence of hypotheses, $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}} - \boldsymbol{\theta}_n)$ converges weakly to a random d -vector \mathbf{Z} with density

$$\mathbf{z} \mapsto \int_{\mathbb{R}^p} \phi_{\boldsymbol{\mu}_{\text{PMTE}}, \boldsymbol{\Gamma}_{\text{PMTE}}}(\mathbf{z}) \phi_{\boldsymbol{\mu}_{\mathbf{D}_{\mathbf{c}_0}}, \boldsymbol{\Sigma}_{\mathbf{D}_{\mathbf{c}_0}}}(\mathbf{x}) d\mathbf{x}, \quad (3.13)$$

where $\phi_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ stands for the density of the d -variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Comparing competing estimators will not be done on the basis of (3.13) but rather on asymptotic mean square errors (AMSEs). We define the AMSE of $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ under $\mathbb{P}_{\boldsymbol{\theta}_n}^{(n)}$ as

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

where \mathbf{Z} is the weak limit of $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}} - \boldsymbol{\theta}_n)$ under $\mathbb{P}_{\boldsymbol{\theta}_n}^{(n)}$. Since $\mathbb{E}[\mathbf{Z}] = \mathbb{E}[\mathbb{E}[\mathbf{Z}|\mathbf{D}_{\mathbf{c}_0}]] = \mathbb{E}[\boldsymbol{\mu}_{\text{PMTE}}]$ and $\text{Var}[\mathbf{Z}] = \mathbb{E}[\text{Var}[\mathbf{Z}|\mathbf{D}_{\mathbf{c}_0}]] + \text{Var}[\mathbb{E}[\mathbf{Z}|\mathbf{D}_{\mathbf{c}_0}]] = \mathbb{E}[\boldsymbol{\Gamma}_{\text{PTE}}] + \text{Var}[\boldsymbol{\mu}_{\text{PMTE}}]$, the AMSE of $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ takes the form $\mathbb{E}[\boldsymbol{\Gamma}_{\text{PTE}}] + \mathbb{E}[\boldsymbol{\mu}_{\text{PMTE}}\boldsymbol{\mu}'_{\text{PMTE}}]$, which can be obtained from Theorem 3.2. To do so, define, for $\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m$,

$$\lambda_{\mathbf{c}, \boldsymbol{\tau}} := \mathbb{P}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}], \quad \boldsymbol{\lambda}_{\mathbf{c}, \boldsymbol{\tau}} := \mathbb{E}[\mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}]\mathbf{D}_{\mathbf{c}_0}], \quad \text{and} \quad \boldsymbol{\Lambda}_{\mathbf{c}, \boldsymbol{\tau}} := \mathbb{E}[\mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}]\mathbf{D}_{\mathbf{c}_0}\mathbf{D}'_{\mathbf{c}_0}]$$

(to keep the notation light, we do not stress the dependence of these quantities on \mathbf{c}_0), where $\mathbf{D}_{\mathbf{c}_0}$ is multinormal with mean vector $\boldsymbol{\mu}_{\mathbf{D}_{\mathbf{c}_0}} = \mathbf{G}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{D}_{\mathbf{c}_0}} = \mathbf{G}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{G}'$ as above. We then have the following result.

Theorem 3.3. *Let Assumptions (A) and (B) hold. Fix $\mathbf{c}_0 \in \mathcal{C}^m \setminus \{\mathbf{1}_m\}$ and $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\mathbf{c}_0})$. Consider the sequence of local perturbations $\boldsymbol{\theta}_n = \boldsymbol{\theta} + \boldsymbol{\nu}_n\boldsymbol{\tau}_n$, where $(\boldsymbol{\tau}_n)$ is a sequence in \mathbb{R}^d*

converging to $\boldsymbol{\tau}$. Then, the AMSE of $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ under $\mathbb{P}_{\boldsymbol{\theta}_n}^{(n)}$ is given by

$$\begin{aligned} \text{AMSE}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}_{\text{PMTE}}) &= \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \left(\lambda_{\mathbf{c},\boldsymbol{\tau}} \boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{1/2} \{ \mathbf{I}_p - \mathbf{P}(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{1/2} \mathbf{G}') \} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{1/2} (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}})' \right. \\ &\quad + \lambda_{\mathbf{c},\boldsymbol{\tau}} (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta},\boldsymbol{\tau}} - \boldsymbol{\tau}) (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta},\boldsymbol{\tau}} - \boldsymbol{\tau})' + \boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{A}_{\boldsymbol{\theta}} \lambda_{\mathbf{c},\boldsymbol{\tau}} (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta},\boldsymbol{\tau}} - \boldsymbol{\tau})' \\ &\quad \left. + (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta},\boldsymbol{\tau}} - \boldsymbol{\tau}) (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{A}_{\boldsymbol{\theta}} \lambda_{\mathbf{c},\boldsymbol{\tau}})' + \boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{A}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}_{\mathbf{c},\boldsymbol{\tau}} \mathbf{A}_{\boldsymbol{\theta}}' \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}}' \boldsymbol{\Upsilon}_{\mathbf{c}}' \right), \end{aligned} \quad (3.14)$$

where we let $\mathbf{A}_{\boldsymbol{\theta}} := \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \mathbf{G}' (\mathbf{G} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \mathbf{G}')^{-}$ and $\mathbf{E}_{\boldsymbol{\theta},\boldsymbol{\tau}} := (\mathbf{I}_d - \mathbf{A}_{\boldsymbol{\theta}} \mathbf{G}) \boldsymbol{\Omega}_{\boldsymbol{\theta}} \boldsymbol{\tau}$.

Note that for $\boldsymbol{\tau} = \mathbf{0}$, (3.14) reduces to

$$\text{AMSE}_{\boldsymbol{\theta},\mathbf{0}}(\hat{\boldsymbol{\theta}}_{\text{PMTE}}) = \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \lambda_{\mathbf{c},\mathbf{0}} \boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{1/2} \{ \mathbf{I}_p - \mathbf{P}(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{1/2} \mathbf{G}') \} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{1/2} + \mathbf{A}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}_{\mathbf{c},\mathbf{0}} \mathbf{A}_{\boldsymbol{\theta}}' \right) (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}})'. \quad (3.15)$$

While the expressions (3.14)–(3.15) are quite complex, they allow one to compare theoretically the proposed PMTE with competing estimators, and in particular with single-constraint PTEs. In the next section, we illustrate this in the k -sample covariance estimation framework described in the introduction.

4 Multisample covariance estimation

4.1 PMTE based on scale and shape constraints

Consider $k (\geq 2)$ mutually independent samples of random p -vectors $\mathbf{X}_{\ell 1}, \dots, \mathbf{X}_{\ell n_{\ell}}$, $\ell = 1, \dots, k$, with respective sample sizes n_1, \dots, n_k , such that, for any ℓ , the $\mathbf{X}_{\ell i}$'s form a random sample from the p -variate normal distribution with mean vector $\mathbf{0}$ and non-singular covariance matrix $\boldsymbol{\Sigma}_{\ell}$ (due to the block-diagonality of the Fisher information matrix for location and scatter in elliptical models, all results below can easily be extended to the

case where observations in the ℓ th sample would have a common, unspecified, mean $\boldsymbol{\mu}_\ell$, $\ell = 1, \dots, k$; see, e.g., [Hallin and Paindaveine, 2006, 2009](#)). As explained in the introduction, the covariance matrices can be reparametrized into $\boldsymbol{\Sigma}_\ell = \sigma_\ell^2 \mathbf{V}_\ell$, where $\sigma_\ell := (\det \boldsymbol{\Sigma}_\ell)^{1/(2p)}$ is their “scale” and $\mathbf{V}_\ell := \boldsymbol{\Sigma}_\ell / (\det \boldsymbol{\Sigma}_\ell)^{1/p}$ is their “shape”. Under the only assumption that $\lambda_\ell := \lambda_\ell^{(n)} := n_\ell / n := n_\ell / (\sum_{r=1}^k n_r)$ converges in $(0, 1)$ for any ℓ (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

$$\boldsymbol{\theta} := (\sigma_1^2, \dots, \sigma_k^2, (\text{vech } \mathbf{V}_1)', \dots, (\text{vech } \mathbf{V}_k)')', \quad (4.16)$$

where $\text{vech } \mathbf{V} =: (V_{11}, \text{vech } \mathbf{V})'$ is the vector stacking the upper-triangular elements of \mathbf{V} , is ULAN; note that since $\det \mathbf{V} = 1$, the upper-left entry V_{11} of \mathbf{V} can be obtained from $\text{vech } \mathbf{V}$ (a vector with dimension $b_p := p(p+1)/2 - 1$), which explains that the upper-left entries of the various shape matrices do not enter the parametrization in (4.16). The dimension of $\boldsymbol{\theta}$ is thus $d := k(b_p + 1)$.

To provide more details, we need the following notation: denoting as \mathbf{e}_r the r th vector of the canonical basis of \mathbb{R}^p and by \otimes the Kronecker product, we let $\mathbf{K}_p := \sum_{r,s=1}^p (\mathbf{e}_r \mathbf{e}_s') \otimes (\mathbf{e}_s \mathbf{e}_r')$ be the $p^2 \times p^2$ commutation matrix, $\mathbf{J}_p := (\text{vec } \mathbf{I}_p)(\text{vec } \mathbf{I}_p)'$, and define $\mathbf{M}_p(\mathbf{V})$ as the $b_p \times p^2$ matrix such that $(\mathbf{M}_p(\mathbf{V}))'(\text{vech } \mathbf{v}) = \text{vec } \mathbf{v}$ for any symmetric $p \times p$ matrix \mathbf{v} such that $\text{tr}[\mathbf{V}^{-1} \mathbf{v}] = 0$. We further put $\mathbf{H}_p(\mathbf{V}) := \frac{1}{4} \mathbf{M}_p(\mathbf{V}) (\mathbf{I}_{p^2} + \mathbf{K}_p) (\mathbf{V} \otimes \mathbf{V})^{-1} (\mathbf{M}_p(\mathbf{V}))'$. Then, letting $\mathbf{S}_\ell := n_\ell^{-1} \sum_{i=1}^{n_\ell} \mathbf{X}_{\ell i} \mathbf{X}'_{\ell i}$ be the empirical covariance matrix in the ℓ th sample (with respect to the fixed location $\boldsymbol{\mu}_\ell = \mathbf{0}$), the central sequence is

$$\Delta_{\boldsymbol{\theta}} := \left(\Delta_{\boldsymbol{\theta}}^{I,1}, \dots, \Delta_{\boldsymbol{\theta}}^{I,k}, (\Delta_{\boldsymbol{\theta}}^{II,1})', \dots, (\Delta_{\boldsymbol{\theta}}^{II,k})' \right)',$$

where, for $\ell = 1, \dots, k$,

$$\Delta_{\boldsymbol{\theta}}^{I,\ell} := \frac{\sqrt{n_\ell}}{2\sigma_\ell^2} \text{tr}[\sigma_\ell^{-2} \mathbf{V}_\ell^{-1} (\mathbf{S}_\ell - \sigma_\ell^2 \mathbf{V}_\ell)] \quad \text{and} \quad \Delta_{\boldsymbol{\theta}}^{II,\ell} := \frac{\sqrt{n_\ell}}{2\sigma_\ell^2} \mathbf{M}_p(\mathbf{V}_\ell) (\mathbf{V}_\ell \otimes \mathbf{V}_\ell)^{-1} (\text{vec } \mathbf{S}_\ell),$$

whereas the non-singular information matrix takes the form $\mathbf{\Gamma}_\theta := \text{diag}(\mathbf{\Gamma}_\theta^I, \mathbf{\Gamma}_\theta^{II})$, with

$$\mathbf{\Gamma}_\theta^I := \frac{p}{2} \text{diag}(\sigma_1^{-4}, \dots, \sigma_k^{-4}) \quad \text{and} \quad \mathbf{\Gamma}_\theta^{II} := \text{diag}(\mathbf{H}_p(\mathbf{V}_1), \dots, \mathbf{H}_p(\mathbf{V}_k))$$

(here, $\text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_m)$ is the block-diagonal matrix with diagonal blocks $\mathbf{A}_1, \dots, \mathbf{A}_m$).

The corresponding matrices $\boldsymbol{\nu}_n$ in Assumption (A) are then given by

$$\boldsymbol{\nu}_n = n^{-1/2} \mathbf{r}_n := n^{-1/2} \text{diag}(\lambda_1^{-1/2}, \dots, \lambda_k^{-1/2}, \lambda_1^{-1/2} \mathbf{I}_{b_p}, \dots, \lambda_k^{-1/2} \mathbf{I}_{b_p}).$$

We consider here the estimation of $\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_k$ or, equivalently, the estimation of $\boldsymbol{\theta}$ in (4.16). An advantage of the $\boldsymbol{\theta}$ -parametrization is that it allows addressing situations in which one would suspect scale homogeneity $\mathcal{H}_0^{\text{scale}} : \sigma_1^2 = \dots = \sigma_k^2$, shape homogeneity $\mathcal{H}_0^{\text{shape}} : \mathbf{V}_1 = \dots = \mathbf{V}_k$, or (the intersection between scale and shape homogeneity:) covariance homogeneity $\mathcal{H}_0^{\text{cov}} : \sigma_1^2 \mathbf{V}_1 = \dots = \sigma_k^2 \mathbf{V}_k$, that is, $\mathcal{H}_0^{\text{cov}} : \boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_k$. In the present Gaussian model, an asymptotically efficient unconstrained estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_U := \left((\det \mathbf{S}_1)^{1/p}, \dots, (\det \mathbf{S}_k)^{1/p}, \frac{(\text{vech } \mathbf{S}_1)'}{(\det \mathbf{S}_1)^{1/p}}, \dots, \frac{(\text{vech } \mathbf{S}_k)'}{(\det \mathbf{S}_k)^{1/p}} \right)'. \quad (4.17)$$

Writing $\mathbf{S} := n^{-1} \sum_{\ell=1}^k \sum_{i=1}^{n_\ell} \mathbf{X}_{\ell i} \mathbf{X}'_{\ell i}$ for the pooled covariance matrix estimator (with respect to the fixed locations $\boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_k = \mathbf{0}$), 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

$$\hat{\boldsymbol{\theta}}_C^{\text{scale}} := \left((\det \mathbf{S})^{1/p} \mathbf{1}'_k, \frac{(\text{vech } \mathbf{S}_1)'}{(\det \mathbf{S}_1)^{1/p}}, \dots, \frac{(\text{vech } \mathbf{S}_k)'}{(\det \mathbf{S}_k)^{1/p}} \right)', \quad (4.18)$$

$$\hat{\boldsymbol{\theta}}_C^{\text{shape}} := \left((\det \mathbf{S}_1)^{1/p}, \dots, (\det \mathbf{S}_k)^{1/p}, \mathbf{1}'_k \otimes \frac{(\text{vech } \mathbf{S})'}{(\det \mathbf{S})^{1/p}} \right)' \quad (4.19)$$

and

$$\hat{\boldsymbol{\theta}}_C^{\text{cov}} := \left((\det \mathbf{S})^{1/p} \mathbf{1}'_k, \mathbf{1}'_k \otimes \frac{(\text{vech } \mathbf{S})'}{(\det \mathbf{S})^{1/p}} \right)', \quad (4.20)$$

respectively. Note that the three hypotheses $\mathcal{H}_0^{\text{scale}}$, $\mathcal{H}_0^{\text{shape}}$ and $\mathcal{H}_0^{\text{cov}}$ impose linear restrictions on $\boldsymbol{\theta}$: they rewrite

$$\mathcal{H}_0^{\text{scale}} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\text{scale}}), \quad \text{with } \boldsymbol{\Upsilon}_{\text{scale}} := \text{diag}(\mathbf{1}_k, \mathbf{I}_{kb_p}),$$

$$\mathcal{H}_0^{\text{shape}} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\text{shape}}), \quad \text{with } \boldsymbol{\Upsilon}_{\text{shape}} := \text{diag}(\mathbf{I}_k, \mathbf{1}_k \otimes \mathbf{I}_{b_p}),$$

and

$$\mathcal{H}_0^{\text{cov}} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\text{cov}}), \quad \text{with } \boldsymbol{\Upsilon}_{\text{cov}} := \text{diag}(\mathbf{1}_k, \mathbf{1}_k \otimes \mathbf{I}_{b_p}).$$

Now, if the $d \times r$ matrix $\boldsymbol{\Upsilon}$ stands for either $\boldsymbol{\Upsilon}_{\text{scale}}$, $\boldsymbol{\Upsilon}_{\text{shape}}$ or $\boldsymbol{\Upsilon}_{\text{cov}}$ (of course, each constraint matrix has its own r), the locally asymptotically most stringent test $\phi_{\mathbf{r}}$ for $\mathcal{H}_0 : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon})$ rejects the null hypothesis at asymptotic level α when

$$Q_{\mathbf{r}} := \boldsymbol{\Delta}'_{\hat{\boldsymbol{\theta}}} \left(\boldsymbol{\Gamma}_{\hat{\boldsymbol{\theta}}}^{-1} - \mathbf{r}_n^{-1} \boldsymbol{\Upsilon} (\boldsymbol{\Upsilon}' \mathbf{r}_n^{-1} \boldsymbol{\Gamma}_{\hat{\boldsymbol{\theta}}} \mathbf{r}_n^{-1} \boldsymbol{\Upsilon})^{-1} \boldsymbol{\Upsilon}' \mathbf{r}_n^{-1} \right) \boldsymbol{\Delta}_{\hat{\boldsymbol{\theta}}} > \chi_{d-r, 1-\alpha}^2, \quad (4.21)$$

where $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\mathbf{r}}$ is a constrained estimator (for the three constraints considered, these are the estimators in (4.18)–(4.20)). On the basis of these various tests, the PTEs involving a single constraint are

$$\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{scale}} := \mathbb{I}[\phi_{\mathbf{r}_{\text{scale}}} = 1] \hat{\boldsymbol{\theta}}_{\text{U}} + \mathbb{I}[\phi_{\mathbf{r}_{\text{scale}}} = 0] \hat{\boldsymbol{\theta}}_{\text{C}}^{\text{scale}}, \quad (4.22)$$

$$\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{shape}} := \mathbb{I}[\phi_{\mathbf{r}_{\text{shape}}} = 1] \hat{\boldsymbol{\theta}}_{\text{U}} + \mathbb{I}[\phi_{\mathbf{r}_{\text{shape}}} = 0] \hat{\boldsymbol{\theta}}_{\text{C}}^{\text{shape}} \quad (4.23)$$

and

$$\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}} := \mathbb{I}[\phi_{\mathbf{r}_{\text{cov}}} = 1] \hat{\boldsymbol{\theta}}_{\text{U}} + \mathbb{I}[\phi_{\mathbf{r}_{\text{cov}}} = 0] \hat{\boldsymbol{\theta}}_{\text{C}}^{\text{cov}}, \quad (4.24)$$

whereas the PMTE proposed in this work is given by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\text{PMTE}} &:= \mathbb{I}[\phi_{\mathbf{r}_{\text{scale}}} = 1, \phi_{\mathbf{r}_{\text{shape}}} = 1] \hat{\boldsymbol{\theta}}_{\text{U}} + \mathbb{I}[\phi_{\mathbf{r}_{\text{scale}}} = 1, \phi_{\mathbf{r}_{\text{shape}}} = 0] \hat{\boldsymbol{\theta}}_{\text{C}}^{\text{shape}} \\ &\quad + \mathbb{I}[\phi_{\mathbf{r}_{\text{scale}}} = 0, \phi_{\mathbf{r}_{\text{shape}}} = 1] \hat{\boldsymbol{\theta}}_{\text{C}}^{\text{scale}} + \mathbb{I}[\phi_{\mathbf{r}_{\text{scale}}} = 0, \phi_{\mathbf{r}_{\text{shape}}} = 0] \hat{\boldsymbol{\theta}}_{\text{C}}^{\text{cov}}. \end{aligned} \quad (4.25)$$

In the next section, we investigate how this PMTE compares with its single-constraint PTE competitors in (4.22)–(4.24), both asymptotically and in finite samples.

4.2 Comparing PMTE vs single-constraint PTEs

In the vicinity of the scale homogeneity constraint and away from the shape homogeneity constraint, $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ is asymptotically equivalent to $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{scale}}$ (Theorem 3.1). Similarly, in the vicinity of the shape homogeneity constraint and away from the scale homogeneity constraint, $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ is asymptotically equivalent to $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{shape}}$. In both cases, thus, the asymptotic properties of $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ can be deduced from those of a single-constraint PTE, so that the AMSEs of this estimator can then be obtained from Theorem 2 in [Paindaveine et al. \(2020\)](#). Recall from Section 3.2, however, that it is only when $\#\mathcal{J}_{\mathbf{c}_0} = 1$ that the asymptotic behaviour of the proposed PMTE can be obtained from the single-constraint theory. In the general case $\#\mathcal{J}_{\mathbf{c}_0} \geq 1$, the asymptotic results from Section 3.2 are the only ones that allow us to grasp the asymptotic behaviour of our PMTE.

In the present situation involving $m = 2$ constraints, a point that is not covered by the single-constraint theory is the comparison between $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ and $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}}$ in the vicinity of covariance homogeneity, that is, close to the null hypothesis $\mathcal{H}_0^{\text{cov}}$. There, using $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ rather than $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}}$ should intuitively have a cost, as the test for $\mathcal{H}_0^{\text{cov}} : \boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\text{cov}})$ is not used when defining $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$. We now evaluate this cost by comparing the asymptotic performances of both estimators, measured by the corresponding AMSEs. Since these AMSEs are matrix-valued, one needs to base this comparison on a scalar summary, such as, e.g., the trace of the AMSEs. In the present setup where asymptotically efficient estimators are used, the benchmark unconstrained estimator satisfies $\text{AMSE}_{\boldsymbol{\theta}, \tau}(\hat{\boldsymbol{\theta}}_{\text{U}}) = \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}$,

which makes it natural to consider the scalar summary

$$\text{AMSE}_{\boldsymbol{\theta}, \boldsymbol{\tau}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}) := \text{tr}[\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2} \text{AMSE}_{\boldsymbol{\theta}, \boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}) \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}], \quad (4.26)$$

which, irrespective of $\boldsymbol{\theta}$, yields $\text{AMSE}_{\boldsymbol{\theta}, \boldsymbol{\tau}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}_{\text{U}}) = d$ for the benchmark estimator. Under the considered covariance homogeneity constraint, we then have the following result.

Proposition 4.1. *For any $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\text{cov}})(= \mathcal{M}(\boldsymbol{\Upsilon}_{\text{scale}}) \cap \mathcal{M}(\boldsymbol{\Upsilon}_{\text{shape}}))$,*

$$\text{AMSE}_{\boldsymbol{\theta}, \mathbf{0}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}_{\text{PMTE}}) = d - (k - 1) \{ b_p \Psi_{(k-1)b_p+2}(\chi_{(k-1)b_p, 1-\alpha}^2) + \Psi_{k+1}(\chi_{k-1, 1-\alpha}^2) \}, \quad (4.27)$$

where Ψ_{ℓ} stands for the cumulative distribution function of the χ_{ℓ}^2 distribution.

It follows from [Paindaveine et al. \(2020\)](#) that, for any $\boldsymbol{\theta} \in \mathcal{M}(\boldsymbol{\Upsilon}_{\text{cov}})$,

$$\text{AMSE}_{\boldsymbol{\theta}, \mathbf{0}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}}) = d - (k - 1)(b_p + 1) \Psi_{(k-1)(b_p+1)+2}(\chi_{(k-1)(b_p+1), 1-\alpha}^2), \quad (4.28)$$

which allows for a direct comparison with the AMSE in (4.27). More precisely, [Figure 1](#) plots the asymptotic relative efficiency

$$\text{ARE}_{\boldsymbol{\theta}, \mathbf{0}}(\hat{\boldsymbol{\theta}}_{\text{PMTE}} / \hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}}) := \frac{\text{AMSE}_{\boldsymbol{\theta}, \mathbf{0}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}})}{\text{AMSE}_{\boldsymbol{\theta}, \mathbf{0}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}_{\text{PMTE}})}$$

as a function of the dimension p and of the number k of populations (here, the nominal level of all preliminary tests is fixed at $\alpha = 5\%$). Clearly, irrespective of p and k , the loss that results from using $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ rather than $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}}$ under covariance homogeneity is extremely small (the minimal ARE, which is obtained for $p = 2$ and $k = 13$, is about .928). Remarkably, this loss actually converges to zero as p or k diverges to infinity.

We turn to Monte Carlo exercises that aim at comparing the finite-sample performances of the proposed PMTE with those of its single-constraint PTE competitors. Throughout, these exercises focus on two populations ($k = 2$) that are two-dimensional ($p = 2$)

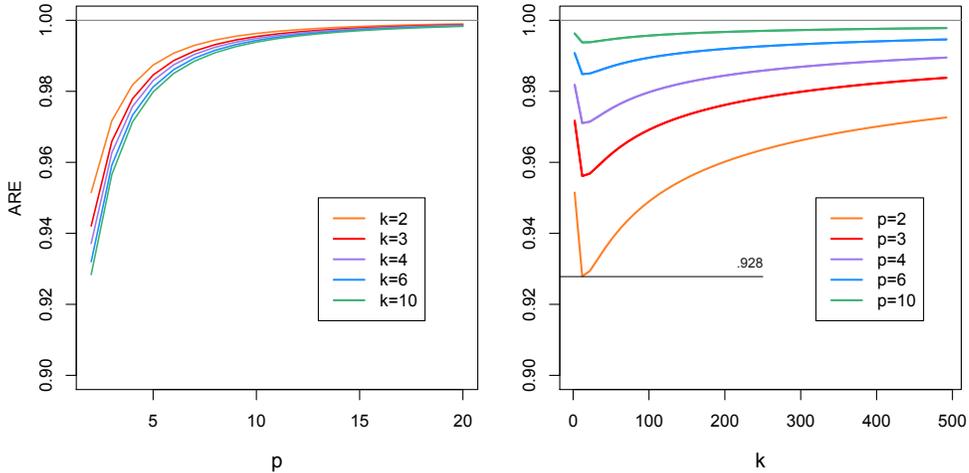


Figure 1: Plots of $\text{ARE}_{\theta,0}(\hat{\theta}_{\text{PMTE}}/\hat{\theta}_{\text{PTE}}^{\text{cov}})$ under covariance homogeneity, as a function of the dimension p for various values of the number k of populations (left), or as a function of the number k of populations for various values of the dimension p (right). All preliminary tests are conducted at nominal level $\alpha = 5\%$.

and balanced ($n_1 = n_2$). We considered two scenarios. In the first one, we generated, for each ξ in $\{0, 1, \dots, 10\}$ and for each value of the common sample size $n_1 = n_2$ in $\{100, 400, 1000\}$, a collection of $M = 10\,000$ samples of mutually independent observations $\mathbf{X}_{11}, \dots, \mathbf{X}_{1n_1}, \mathbf{X}_{21,\xi}, \dots, \mathbf{X}_{2n_2,\xi}$, where the \mathbf{X}_{1i} 's are $\mathcal{N}(\mathbf{0}, \Sigma_1)$ and the $\mathbf{X}_{2i,\xi}$'s are $\mathcal{N}(\mathbf{0}, \Sigma_{2,\xi})$, with $\Sigma_1 = \mathbf{I}_2$ and $\Sigma_{2,\xi} = \sigma_{2,\xi}^2 \mathbf{V}_{2,\xi}$ based on

$$\sigma_{2,\xi}^2 = \exp\left(\frac{\xi}{2\sqrt{n}}\right), \quad \mathbf{V}_{2,\xi} = \frac{\mathbf{N}_\xi}{\sqrt{\det \mathbf{N}_\xi}}, \quad \text{and} \quad \mathbf{N}_\xi = \begin{pmatrix} 1 & \frac{\xi}{3\sqrt{n}} \\ \frac{\xi}{3\sqrt{n}} & 1 \end{pmatrix}. \quad (4.29)$$

The value $\xi = 0$ provides covariance homogeneity, hence also scale and shape homogeneity, whereas $\xi = 1, \dots, 10$ provide both increasingly distinct scales and increasingly distinct shapes. For any estimator $\hat{\theta}$ of the corresponding parameter value θ , the finite-sample

performance of $\hat{\boldsymbol{\theta}}$ can be measured through the scalar quantity

$$\text{MSE}_{\boldsymbol{\theta}, \boldsymbol{\Gamma}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}) := \text{tr}[\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2} \text{MSE}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2}], \quad (4.30)$$

with

$$\text{MSE}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) := \frac{1}{M} \sum_{m=1}^M \{\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}^{(m)} - \boldsymbol{\theta})\} \{\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}^{(m)} - \boldsymbol{\theta})\}', \quad (4.31)$$

where $\hat{\boldsymbol{\theta}}^{(m)}$ denotes the value of $\hat{\boldsymbol{\theta}}$ in the m th replication. The left panels of Figure 2 then plot $\text{MSE}_{\boldsymbol{\theta}, \boldsymbol{\Gamma}}^{\text{scalar}}(\hat{\boldsymbol{\theta}})$ as a function of ξ , for the PTE $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{shape}}$ in (4.23) based on the single shape homogeneity constraint, for the PTE $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}}$ in (4.24) based on the single covariance homogeneity constraint, for the proposed PMTE $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ in (4.25) based on the shape and scale homogeneity constraints, and for their unconstrained antecedent $\hat{\boldsymbol{\theta}}_{\text{U}}$ (all preliminary tests were performed at asymptotic level $\alpha = 5\%$). In the present setup involving deviations from covariance homogeneity, the estimator $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}}$ is an oracle one, that is expected to outperform its competitors. Remarkably, the results show that the multiple-constraint estimator $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ show virtually the same performances as $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}}$. Under covariance homogeneity ($\xi = 0$), this is in line with our theoretical results above, as Figure 2 indeed confirms the very close values $\text{AMSE}_{\boldsymbol{\theta}, \mathbf{0}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}_{\text{PMTE}}) = 3.679$ and $\text{AMSE}_{\boldsymbol{\theta}, \mathbf{0}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}}) = 3.500$ that result from (4.27) and (4.28), respectively.

We repeated the exercise above in a second scenario, that is obtained from the first one by replacing $\sigma_{2, \xi}^2$ in (4.29) with $\sigma_{2, \xi}^2 = 2$ for any ξ . Irrespective of ξ , thus, this new scenario stays away from scale homogeneity (hence also from covariance homogeneity), whereas $\xi = 1, \dots, 10$ provide increasingly severe departures from the shape homogeneity situation obtained for $\xi = 0$. The right panels of Figure 2 show the resulting values of $\text{MSE}_{\boldsymbol{\theta}, \boldsymbol{\Gamma}}^{\text{scalar}}(\hat{\boldsymbol{\theta}})$ for the same four estimators as above. The results clearly support Theorem 3.1, that states that $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ and $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{shape}}$ are asymptotically equivalent, as one cannot

discriminate between the MSE curves of these estimators (further simulations revealed that one needs to consider sample sizes as small as $n_1 = n_2 = 100$ to see a tiny difference in these MSE curves). Overall, thus, $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ dominates $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{shape}}$ since the former was performing better than the latter in the first scenario. Incidentally, note that, in this second scenario, $\hat{\boldsymbol{\theta}}_{\text{PTE}}^{\text{cov}}$ and $\hat{\boldsymbol{\theta}}_{\text{U}}$ behave very similarly, which is reasonable since all values of ξ considered provide a setup that is far from covariance homogeneity (Corollary 3.1).

4.3 CPC and homogeneity of eigenvalues: a real data example

Studies of microtus population biology have attracted a lot of attention in the past decades; see, e.g., Wallace (2006) and Conroy and Gupta (2011), and the references therein. In this real data illustration, we discuss the estimation of covariance matrices for two samples ($k = 2$) of different species of voles: a sample of $n_1 = 43$ *Microtus multiplex* and a sample of $n_2 = 46$ *Microtus subterraneus*. Eight measurements ($p = 8$) are made on each animal: (i)–(iii) the width of upper left molar 1–3, (iv) the length of incisive foramen, (v) the length of palatal bone, (vi) the Condylar incisive length or skull length, (vii) the skull height above bullae, and (viii) the skull width across rostrum; see Airolidi et al. (1996). The dataset is available in the R package *Flury* (data *microtus*).

We consider estimation of the underlying covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ using preliminary tests of the following three constraints ($m = 3$): homogeneity of scales (\mathcal{H}_0^a), homogeneity of the shape matrices' eigenvectors (\mathcal{H}_0^b , described in the introduction as the CPC hypothesis), and homogeneity of the shape matrices' eigenvalues (\mathcal{H}_0^c). We performed the optimal Gaussian test for \mathcal{H}_0^a from Hallin and Paindaveine (2009), the optimal Gaussian test for \mathcal{H}_0^b from Hallin et al. (2013), and the optimal Gaussian test for \mathcal{H}_0^c we derived from the local asymptotic normality result in Hallin et al. (2013). The respective p -values

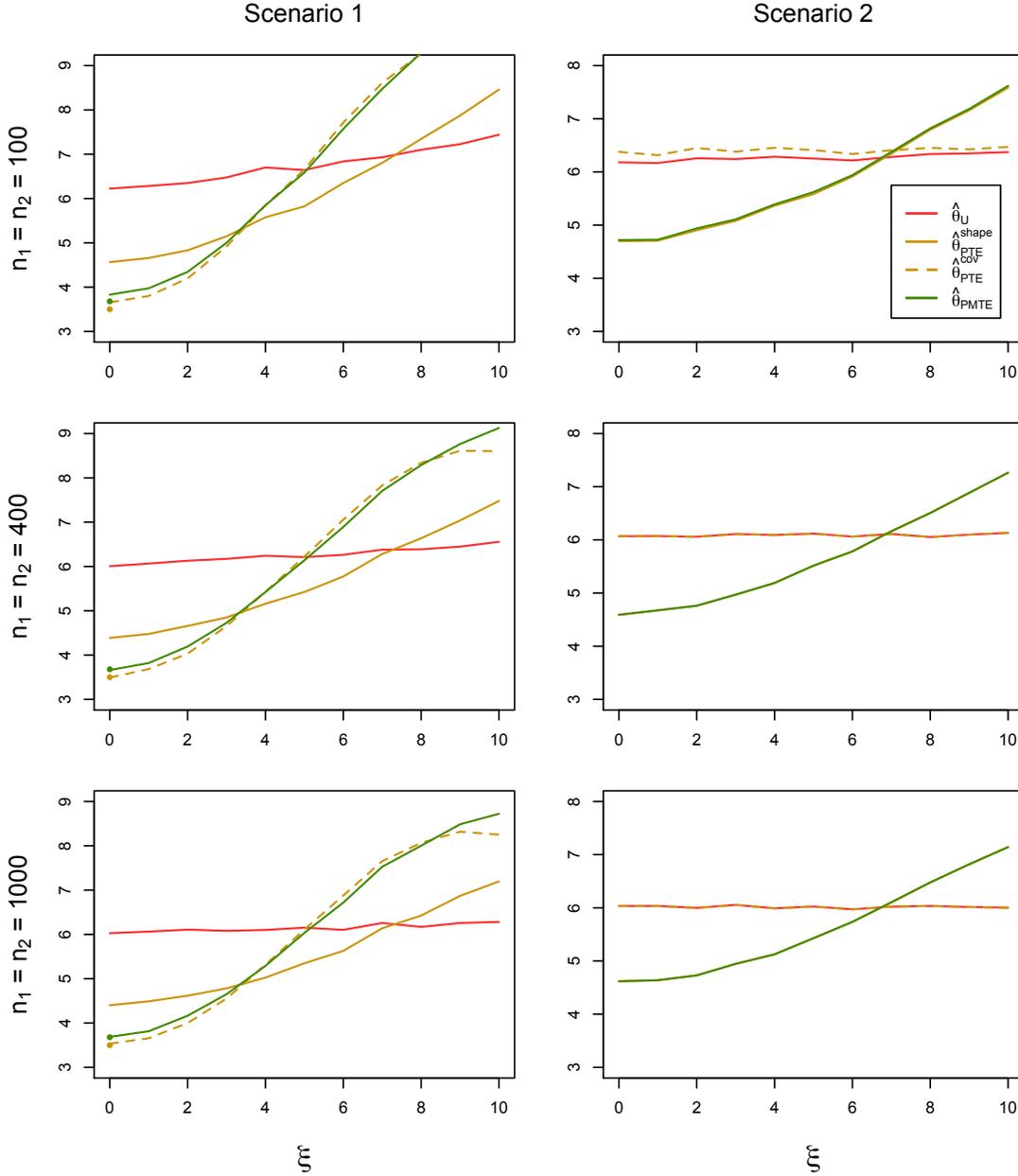


Figure 2: Plots, as functions of $\xi = 0, 1, \dots, 10$, of $\text{MSE}_{\theta, \Gamma}^{\text{scalar}}(\hat{\theta})$ for the single-constraint PTEs $\hat{\theta}_{\text{PTE}}^{\text{shape}}$ and $\hat{\theta}_{\text{PTE}}^{\text{cov}}$, the PMTE $\hat{\theta}_{\text{PMTE}}$ based on the shape and scale homogeneity constraints, and their unconstrained antecedent $\hat{\theta}_{\text{U}}$ (all preliminary tests are performed at asymptotic level $\alpha = 5\%$). In Scenario 1, larger values of ξ provide increasingly severe deviations—both in terms of scale and shape²⁴—from the covariance homogeneity obtained at $\xi = 0$, whereas, in Scenario 2, scale heterogeneity holds for any ξ and larger values of ξ provide increasingly severe deviations from the shape homogeneity obtained at $\xi = 0$; see Section 4.2 for details. The green and orange points in the left panels indicate the values of $\text{AMSE}_{\theta, 0}^{\text{scalar}}(\hat{\theta}_{\text{PMTE}})$ and $\text{AMSE}_{\theta, 0}^{\text{scalar}}(\hat{\theta}_{\text{PTE}}^{\text{cov}})$, respectively.

are .0121, .0010, and .2004, indicating that, for $\alpha = 5\%$, the PMTE based on the three constraints above would be an estimator of $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ assuming common shape eigenvalues only, whereas for $\alpha = 1\%$, it would be an estimator assuming both common scales and common shape eigenvalues—that is, assuming that $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ share the same eigenvalues.

As explained in [Airoldi et al. \(1996\)](#), *Microtus multiplex* and *Microtus subterraneus* are difficult to distinguish morphologically. Actually, it is only since [Krapp \(1982\)](#) and [Niethammer \(1982\)](#) that they are considered two distinct species—as a reaction to the vision in [Ellerman and Morrison-Scott \(1951\)](#). One possible way to explore the practical relevance of the PMTE above is thus to perform supervised classification. To do so, we randomly sampled 30 observations in each group and trained various classifiers on the resulting training set of size 60. The misclassification rate of each classifier was then evaluated on the basis of the test set made of the remaining 29 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 $M = 2000$ times; [Figure 3](#) provides, for each classifier, a boxplot of the resulting M misclassification rates. The considered classifiers all perform quadratic discriminant analysis (QDA) using plain sample averages as estimates $\hat{\boldsymbol{\mu}}_\ell$, $\ell = 1, 2$, of the group-specific mean vectors, hence only differ through the estimates $\hat{\boldsymbol{\Sigma}}_\ell$, $\ell = 1, 2$, of the corresponding covariance matrices. The classical QDA procedure, that will be the benchmark, is based on the unconstrained sample covariance matrices $\hat{\boldsymbol{\Sigma}}_{\ell,U}$, $\ell = 1, 2$. The other classifiers are based on various preliminary (single or multiple) test estimators $\hat{\boldsymbol{\Sigma}}_\ell$, $\ell = 1, 2$, using the same unconstrained estimators as in the benchmark and the constrained estimators obtained from the following estimators: letting $\hat{\boldsymbol{\Sigma}}_{\text{pool}} := (n_1 \hat{\boldsymbol{\Sigma}}_{1,U} + n_2 \hat{\boldsymbol{\Sigma}}_{2,U}) / (n_1 + n_2)$ be the pooled covariance matrix, the common value of the scale parameter under \mathcal{H}_0^a is estimated by $(\det \hat{\boldsymbol{\Sigma}}_{\text{pool}})^{1/p}$ (recall that $p = 8$), the common eigenvectors matrix under \mathcal{H}_0^b is estimated

using the eigenvectors matrix $\hat{\beta}_{\text{pool}}$ of $\hat{\Sigma}_{\text{pool}}$, whereas the common value of the eigenvalues matrix under \mathcal{H}_0^c is estimated using $\hat{\Lambda}_{\text{pool}}/(\det \hat{\Sigma}_{\text{pool}})^{1/p}$, with $\hat{\Lambda}_{\text{pool}} := \hat{\beta}'_{\text{pool}} \hat{\Sigma}_{\text{pool}} \hat{\beta}_{\text{pool}}$.

Figure 3 provides the boxplots of the misclassification rates (and reports the average misclassification rates) for the resulting five QDA classifiers, namely the ones based (i) on unconstrained estimators (that is, the classical QDA classifier), (ii)–(iv) on PTE estimators associated with the single constraint of homogeneity of scales (\mathcal{H}_0^a), homogeneity of the shape matrices' eigenvectors (\mathcal{H}_0^b), and homogeneity of the shape matrices' eigenvalues (\mathcal{H}_0^c), and (v) on the PMTE involving these three constraints. Each of the classifiers (ii)–(v) was considered in four versions, according to the nominal level α used for preliminary tests: $\alpha = 0.1\%$, 1% , 5% , and a value of α obtained from 6-fold cross-validation. Clearly, the results indicate that the best classifiers are the ones based on the PMTEs and that those based on PTEs only marginally improve over the benchmark unconstrained classifier. Moreover, it is seen that cross-validation provides an effective way to choose the tuning parameter α . Figure 4 also provides the boxplots of the p -values of the tests for \mathcal{H}_0^a , \mathcal{H}_0^b , and \mathcal{H}_0^c , obtained in the collection of $M = 2000$ training samples above. These boxplots reveal that it is not uncommon that the null hypotheses \mathcal{H}_0^a and \mathcal{H}_0^c both fail to be rejected, which explains that our PMTE has an edge in the present classification exercise.

5 Final comments

When demonstrating the practical relevance of our PMTE in Section 4, we focused (i) on multisample covariance estimation and (ii) on constrained and unconstrained estimators that are of a Gaussian nature (more precisely, the estimators there were all Gaussian maximum likelihood estimators). As explained in Section 3, however, our methodology is very widely applicable as it merely only requires that the considered model is locally asymptoti-

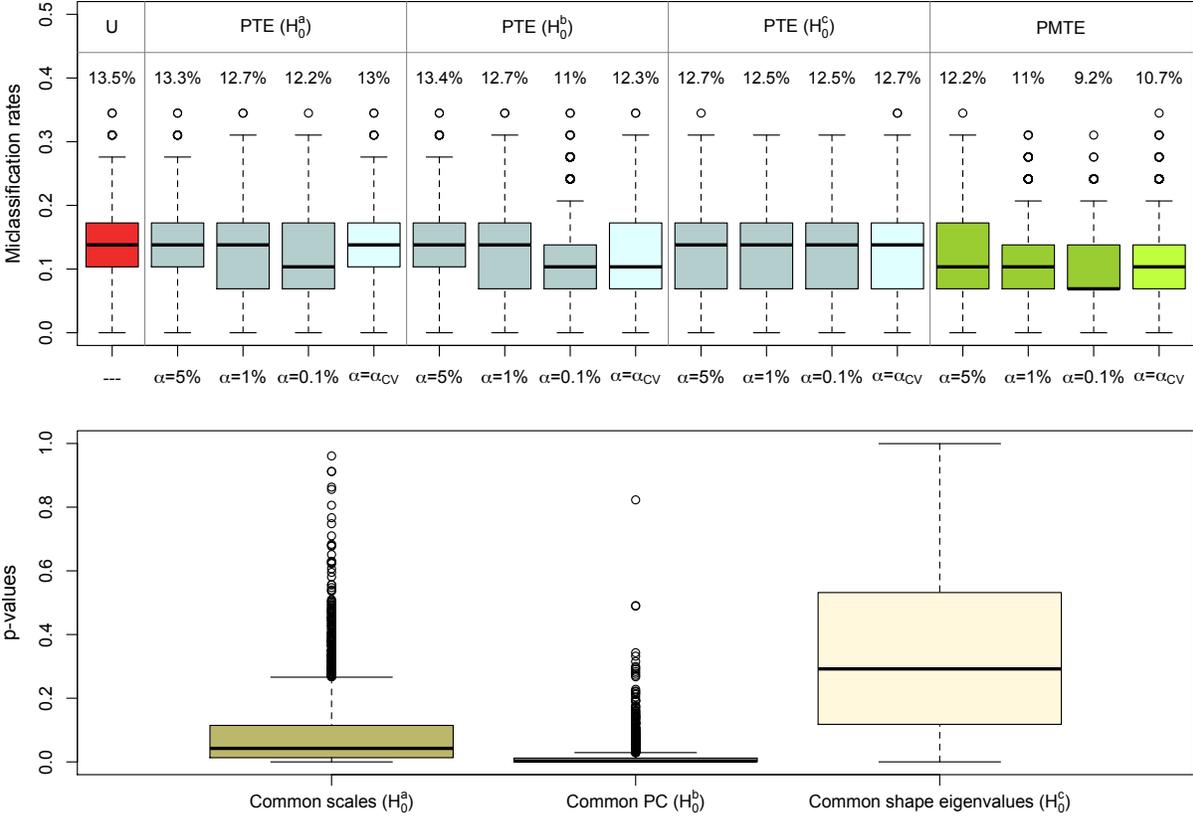


Figure 3: (Top:) Boxplots of the misclassification rates obtained by applying different QDA classifiers in $M = 2000$ random partitions of the microtus dataset into training and test sets. The QDA classifiers use different estimators of the group-specific covariance matrices, namely the usual unconstrained estimators, the PMTE involving the constraints \mathcal{H}_0^a , \mathcal{H}_0^b and \mathcal{H}_0^c from Section 4.3, and the three classical PTEs involving one of these constraints only. Each PMTE/PTE comes in four versions, according to the nominal level α used in preliminary tests, the last version being based on 6-fold cross-validation. Percentages above the boxplots report the average misclassification rate of the corresponding classifiers; see Section 4.3. (Bottom:) Boxplots of the p -values of the tests for the constraints \mathcal{H}_0^a , \mathcal{H}_0^b , and \mathcal{H}_0^c , obtained in the same collection of 2000 training samples.

cally normal and that unconstrained estimators of the corresponding parameters admitting Bahadur representation results are available. To showcase practical use of PMTEs in another situation and based on estimators of another nature, we consider robust one-sample covariance matrix estimation.

Testing for linear constraints on a covariance matrix Σ has been much considered in the literature; see, e.g., [Zhang et al. \(1991\)](#) and the references therein (see also [Dryden et al., 2009](#) for testing constraints with observed covariance matrices). Commonly considered constraints on Σ are associated, e.g., with the assumptions of *sphericity* ($\mathcal{H}_0^{\text{sph}} : \Sigma = \lambda \mathbf{I}_p$ for some $\lambda > 0$) or *multivariate independence* ($\mathcal{H}_0^{\text{ind}} : \Sigma = \text{diag}(\Sigma_1, \Sigma_2)$, where Σ_1 and Σ_2 are $q \times q$ and $(p-q) \times (p-q)$ covariance matrices, respectively). When aiming at robust estimation based on preliminary testing, it is natural to consider robust tests for these constraints, such as the [Hallin and Paindaveine \(2006\)](#) sign test for $\mathcal{H}_0^{\text{sph}}$ and the [Taskinen et al. \(2003\)](#) rank test for $\mathcal{H}_0^{\text{ind}}$. Obviously, one then also needs to rely on a robust unconstrained estimator of Σ , such as the celebrated Minimum Covariance Determinant (MCD) estimator from [Rousseeuw \(1985\)](#) (which, as required, satisfies a Bahadur representation result of the form (3.7); see [Cator and Lopuhaa, 2010](#)). To explore the performance of a corresponding robust PMTE, we generated, for each ξ in $\{0, 1, \dots, 5\}$ and for each $n \in \{100, 400, 1000\}$, $M = 10\,000$ independent random samples $\mathbf{X}_{1,\xi}, \dots, \mathbf{X}_{n,\xi}$ of size n from the four-dimensional ($p = 4$) multinormal distribution with mean vector zero and covariance matrix

$$\Sigma_\xi = \frac{1}{p} \left(\left(1 - \frac{2\xi}{3\sqrt{n}} \right) \mathbf{I}_p + \frac{2\xi}{3\sqrt{n}} \mathbf{1}_p \mathbf{1}_p' \right),$$

with $\mathbf{1}_p = (1, \dots, 1)' \in \mathbb{R}^p$. In each replication, we evaluated the MCD and a robust PMTE based on the constraints $\mathcal{H}_0^{\text{sph}}$ and $\mathcal{H}_0^{\text{ind}}$, with $q = 2$. For this PMTE, the preliminary tests (performed at asymptotic level $\alpha = 5\%$) are the robust ones mentioned above, and the MCD was used to obtain the needed unconstrained and constrained estimators of Σ (for instance,

the constrained estimator of Σ under $\mathcal{H}_0^{\text{ind}}$ is $\text{diag}(\hat{\Sigma}_1, \hat{\Sigma}_2)$, where $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ stand for the MCD estimators obtained from the first q marginals and last $p - q$ marginals, respectively). Denoting as $\hat{\Sigma}^{(m)}$ the value of a given estimator in the m th replication, Figure 4 provides, both for the MCD and for this robust PMTE, the MSE quantities

$$\text{MSE}_{\Sigma}(\hat{\Sigma}) := \frac{1}{M} \sum_{m=1}^M \text{tr}[\{\sqrt{n}(\hat{\Sigma}^{(m)} - \Sigma)\}^2], \quad (5.32)$$

with respect to the value of $\Sigma = \Sigma_{\xi}$ at hand. As Figure 4 clearly shows, this robust PMTE shows the same dominance over its unconstrained antecedent in the vicinity of the constraints (note that the constraints are met at $\xi = 0$) as the Gaussian ones considered in the multisample framework of Section 4.

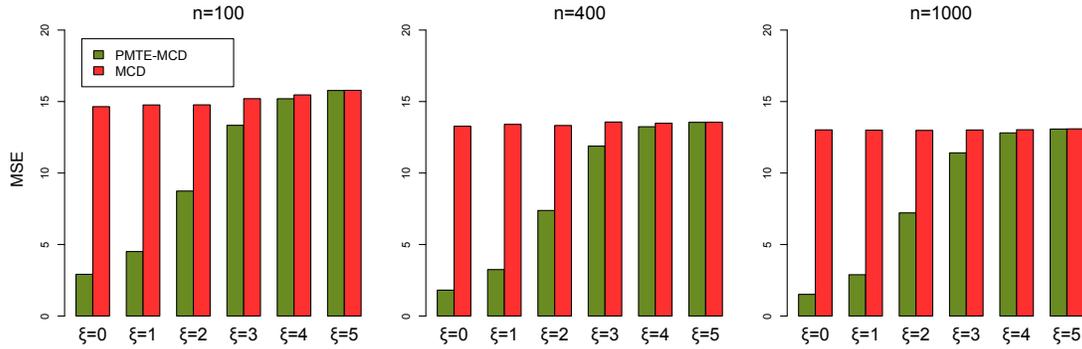


Figure 4: Bar plots, as functions of $\xi = 0, 1, \dots, 5$, of the MSE quantities $\text{MSE}_{\Sigma}(\hat{\Sigma})$ in (5.32) for (in red:) the MCD estimator of Σ and (in green:) the MCD-based PMTE associated with the sphericity and multivariate independence constraints (with preliminary tests performed at asymptotic level $\alpha = 5\%$). Here, the larger the value of ξ , the more severe the departure from the sphericity and multivariate independence constraints (that are both met at $\xi = 0$); see Section 5 for details.

We conclude the paper with a brief discussion of possible research perspectives. First, implementing the PMTE proposed in this work requires selecting the nominal level α at which to perform the preliminary tests, and it would be natural to develop methods for this choice. One way to tackle this problem is to try to adapt the “minimax regret” strategy proposed in [Giles et al. \(1992\)](#) using the general value of the AMSE measure in (3.14). In some specific contexts, though, a suitable value of α may simply be chosen through cross-validation, as we showed in the supervised classification exercise conducted in Section 4.3. Second, while we focused on asymptotic scenarios where the dimension d of the parameter remains fixed as the sample size n diverges to infinity, it would be interesting to tackle the high-dimensional case where the dimension $d = d_n$ diverges to infinity with n . This is of course quite challenging, particularly so for covariance estimation since it is well-known that covariance matrices cannot always be estimated consistently in high dimensions. Finally, another interesting venue for future research on PMTEs, that would be especially relevant in high dimensions, would be to consider asymptotics as the number of constraints $m = m_n$ increases with n . The need to resort to multiple testing corrections would then maybe be more imperious than in the fixed- m framework we considered. These research perspectives all are quite ambitious and are left for future work.

References

- Airoldi, J.-P., Flury, Bernard, N., and Salvioni, M. (1996). Discrimination between two species of *Microtus* using both classified and unclassified observations. *Journal of Theoretical Biology*, 177:247–262.
- Benko, M., Härdle, W., Kneip, A., et al. (2009). Common functional principal components.

Annals of Statistics, 37(1):1–34.

Bickel, P. J. and Ritov, Y. (1996). Inference in hidden markov models I: Local asymptotic normality in the stationary case. *Bernoulli*, 2(3):199–228.

Boente, G. and Orellana, L. (2004). Robust plug-in estimators in proportional scatter models. *Journal of Statistical Planning and Inference*, 122(1-2):95–110.

Boente, G., Pires, A. M., and Rodrigues, I. M. (2002). Influence functions and outlier detection under the common principal components model: a robust approach. *Biometrika*, 89(4):861–875.

Boente, G., Pires, A. M., and Rodrigues, I. M. (2009). Robust tests for the common principal components model. *Journal of Statistical Planning and Inference*, 139(4):1332–1347.

Browne, R. P. and McNicholas, P. D. (2014). Estimating common principal components in high dimensions. *Advances in Data Analysis and Classification*, 8(2):217–226.

Cator, E. and Lopuhaa, H. P. (2010). Asymptotic expansion of the minimum covariance determinant estimators. *Journal of Multivariate Analysis*, 101:2372–2388.

Conroy, C. J. and Gupta, A. M. (2011). Cranial morphology of the California vole (*Microtus californicus*, cricetidae) in a contact zone. *Biological Journal of the Linnean Society*, 104(2):264–283.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics*, 17:1749–1766.

- Drost, F. C., Klaassen, C. A., and Werker, B. J. (1997). Adaptive estimation in time-series models. *Annals of Statistics*, 25(2):786–817.
- Dryden, I. L., Koloydenko, A., Zhou, D., et al. (2009). Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*, 3(3):1102–1123.
- Ellerman, J. and Morrison-Scott, T. (1951). Checklist of Palaeartic and indian mammals 1758 to 1946. *London British Museum (Natural History)*
- Flury, B. N. (1984). Common principal components in k groups. *Journal of the American Statistical Association*, 79:892–898.
- Flury, B. N. (1986). Asymptotic theory for common principal component analysis. *Annals of Statistics*, 14:418–430.
- Francq, C. and Zakoian, J.-M. (2013). Inference in nonstationary asymmetric GARCH models. *Annals of Statistics*, 41(4):1970–1998.
- García-Portugués, E., Paindaveine, D., and Verdebout, T. (2020). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *Journal of the American Statistical Association* page to appear.
- 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. *Journal of the American Statistical Association*, 87(420):1153–1157.
- Hallin, M., Oja, H., and Paindaveine, D. (2006). Semiparametrically efficient rank-based

- inference for shape II. optimal r-estimation of shape. *Annals of Statistics*, 34(6):2757–2789.
- Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape I. Optimal rank-based tests for sphericity. *Annals of Statistics*, 34(6):2707–2756.
- Hallin, M. and Paindaveine, D. (2009). Optimal tests for homogeneity of covariance, scale, and shape. *Journal of Multivariate Analysis*, 100(3):422–444.
- Hallin, M., Paindaveine, D., and Verdebout, T. (2013). Optimal rank-based tests for common principal components. *Bernoulli*, 19(5B):2524–2556.
- Hallin, M., Paindaveine, D., and Verdebout, T. (2014). Efficient R-estimation of principal and common principal components. *Journal of the American Statistical Association*, 109(507):1071–1083.
- Hallin, M., Taniguchi, M., Serroukh, A., Choy, K., et al. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Annals of Statistics*, 27(6):2054–2080.
- Jensen, S. T. and Madsen, J. (2004). Estimation of proportional covariances in the presence of certain linear restrictions. *Annals of Statistics*, 32(1):219–232.
- Krapp, F. (1982). *Microtus multiplex* (fatio, 1905)–alpen-kleinwühlmaus. *Handbuch der Säugetiere Europas*, 2(1):419–428.
- Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, New York.

- Ley, C., Swan, Y., Thiam, B., and Verdebout, T. (2013). Optimal R-estimation of a spherical location. *Statistica Sinica*, 23(1):305–333.
- Ley, C. and Verdebout, T. (2017). *Modern Directional Statistics*. Chapman and Hall/CRC.
- Liu, B., Xu, L., Zheng, S., and Tian, G.-L. (2014). A new test for the proportionality of two large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 131:293–308.
- Lopuhaa, H. P. (1992). Highly efficient estimators of multivariate location with high breakdown point. *Annals of Statistics*, 20:398–413.
- Maeyama, Y., Tamaki, K., and Taniguchi, M. (2011). Preliminary test estimation for spectra. *Statistics and Probability Letters*, 81(11):1580–1587.
- Niethammer, J. (1982). *Microtus subterraneus* (de sélys-longchamps, 1836)–kurzohrmaus. *Handbuch der Säugetiere Europas*, 2(1):397–418.
- Paindaveine, D., Rasoafaraniaina, J., and Verdebout, T. (2020). Preliminary test estimation in ULAN models. *Under revision*
- Paindaveine, D., Rasoafaraniaina, R. J., and Verdebout, T. (2017). Preliminary test estimation for multi-sample principal components. *Econometrics and Statistics*, 2:106–116.
- Paindaveine, D. and Van Bever, G. (2014). Inference on the shape of elliptical distributions based on the MCD. *Journal of Multivariate Analysis*, 129:125–144.
- Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical statistics and applications*, 8(283-297):37.

- Saleh, A. M. E. (2006). *Theory of Preliminary Test and Stein-type Estimation with Applications*, volume 517. John Wiley & Sons.
- Schott, J. R. (1991). Some tests for common principal component subspaces in several groups. *Biometrika*, 78(4):771–777.
- Schott, J. R. (2001). Some tests for the equality of covariance matrices. *Journal of Statistical Planning and Inference*, 94(1):25–36.
- Taskinen, S., Kankainen, A., and Oja, H. (2003). Sign test of independence between two random vectors. *Statistics and Probability Letters*, 62(1):9–21.
- Tsukuda, K. and Matsuura, S. (2019). High-dimensional testing for proportional covariance matrices. *Journal of Multivariate Analysis*, 171:412–420.
- Wallace, S. C. (2006). Differentiating *Microtus xanthognathus* and *Microtus pennsylvanicus* lower first molars using discriminant analysis of landmark data. *Journal of Mammalogy*, 87(6):1261–1269.
- Zhang, J., Pantula, S. G., and Boos, D. D. (1991). Robust methods for testing the pattern of a single covariance matrix. *Biometrika*, 78(4):787–795.

SUPPLEMENTARY MATERIAL

In Section A of this supplement, we provide a comparison of our method with a BIC model selection-based estimator. In Section B, we then provide the proofs of the various results obtained in the paper.

A Comparison with BIC-based model selection

Both in its classical single-constraint version and in the multiple one proposed in this paper, preliminary test estimation de facto realizes model selection. Another common model selection procedure is the one based on the Bayesian information criterion (BIC). Using the generic notation from Section 2 and denoting as $L_{\boldsymbol{\theta}}^{(n)}$ the likelihood function of the parametric model at hand, this criterion measures suitability of the submodel associated with the constraint $\boldsymbol{\theta} \in \cap_{j \in \mathcal{J}_{\mathbf{c}}} \mathcal{M}(\boldsymbol{\Upsilon}_j)$, with $\mathbf{c} \in \mathcal{C}^m$, through

$$\text{BIC}_{\mathbf{c}} := \max_{\boldsymbol{\theta}} \log L_{\boldsymbol{\theta}}^{(n)} - n \log d_{\mathbf{c}},$$

where, as in Section 2, the full-rank $d \times d_{\mathbf{c}}$ matrix $\boldsymbol{\Upsilon}_{\mathbf{c}}$ is such that $\mathcal{M}(\boldsymbol{\Upsilon}_{\mathbf{c}}) = \cap_{j \in \mathcal{J}_{\mathbf{c}}} \mathcal{M}(\boldsymbol{\Upsilon}_j)$ (so that $d_{\mathbf{c}}$ is the number of functionally independent parameters indexing the constrained model associated with \mathbf{c}). The estimator resulting from this criterion,

$$\hat{\boldsymbol{\theta}}_{\text{BIC}} := \sum_{\mathbf{c} \in \mathcal{C}^m} \mathbb{I}[\arg\max_{\mathbf{d} \in \mathcal{C}^m} \text{BIC}_{\mathbf{d}} = \mathbf{c}] \hat{\boldsymbol{\theta}}_{\mathbf{c}}, \quad (\text{A.33})$$

then simply is the constrained estimator associated with the submodel showing the largest BIC value among all possible submodels. It is natural to wonder how such an estimator compares with the preliminary multiple test estimator (2.4) proposed in this work.

To investigate this, we performed the following simulation, still in the bivariate two-sample covariance estimation framework considered above. More specifically, for each ξ in $\{0, 1, \dots, 5\}$ and for each value of the common sample size $n_1 = n_2$ in $\{100, 400, 1000\}$, we generated a collection of $M = 10\,000$ samples of mutually independent observations $\mathbf{X}_{11}, \dots, \mathbf{X}_{1n_1}, \mathbf{X}_{21,\xi}, \dots, \mathbf{X}_{2n_2,\xi}$, where the \mathbf{X}_{1i} 's are $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_1)$ and the $\mathbf{X}_{2i,\xi}$'s are $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{2,\xi})$, with $\boldsymbol{\Sigma}_1 = \mathbf{I}_2$ and $\boldsymbol{\Sigma}_{2,\xi} = \sigma_{2,\xi}^2 \mathbf{V}_{2,\xi}$ based on

$$\sigma_{2,\xi}^2 = 2, \quad \mathbf{V}_{2,\xi} = \frac{\mathbf{N}_\xi}{\sqrt{\det \mathbf{N}_\xi}}, \quad \text{and} \quad \mathbf{N}_\xi = \begin{pmatrix} 1 & \frac{2\xi}{\sqrt{n}} \\ \frac{2\xi}{\sqrt{n}} & 1 \end{pmatrix}.$$

The value $\xi = 0$ corresponds to shape homogeneity, whereas $\xi = 1, \dots, 5$ provide increasingly distinct shapes. In each replication, we evaluated the PMTE estimator $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ in (4.25) (still with preliminary tests performed at asymptotic level $\alpha = 5\%$) and the BIC estimator $\hat{\boldsymbol{\theta}}_{\text{BIC}}$ in (A.33), both based on the shape and scale homogeneity constraints (so that both estimators need to select the right submodel in a collection of $2^m = 2^2 = 4$ possible models). The parametric likelihood $L_{\boldsymbol{\theta}}^{(n)}$ used in the computation of the BIC is the multisample Gaussian likelihood; the model used in the BIC is therefore correctly specified. Here, the finite-sample performance of an estimator $\hat{\boldsymbol{\theta}}$ was assessed by the scalar MSE quantity

$$\text{MSE}_{\boldsymbol{\theta}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}) := \text{tr}[\text{MSE}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})], \tag{A.34}$$

where $\text{MSE}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ is the matrix-valued MSE in (4.31) based on the values $\hat{\boldsymbol{\theta}}^{(1)}, \dots, \hat{\boldsymbol{\theta}}^{(M)}$ of $\hat{\boldsymbol{\theta}}$ obtained in the M replications. Figure 5, which provides bar plots of the scalar MSEs in (A.34) for $\hat{\boldsymbol{\theta}}_{\text{PMTE}}$ and $\hat{\boldsymbol{\theta}}_{\text{BIC}}$, clearly shows that the proposed PMTE estimator compares favourably with its BIC competitor, especially far from the (shape homogeneity) constraint. While we do not report the results here, we checked that replacing the scalar MSE in (A.34) with the model-dependent one in (4.30)—that was adopted in Section 4.2 only to show the

agreement of finite-sample performances with our asymptotic results—has no qualitative impact on the results of Figure 5.

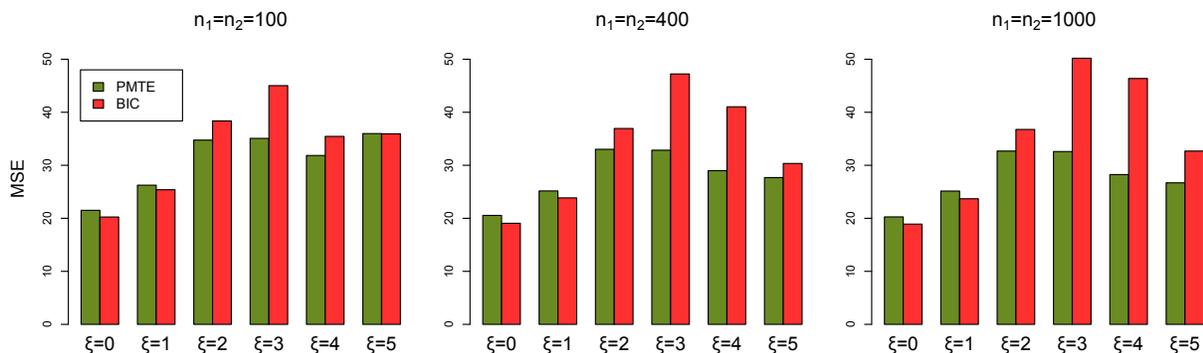


Figure 5: Bar plots, as functions of $\xi = 0, 1, \dots, 5$, of the scalar MSE quantity $\text{MSE}_{\hat{\theta}}^{\text{scalar}}(\hat{\theta})$ in (A.34) for the PMTE $\hat{\theta}_{\text{PMTE}}$ and its BIC competitor $\hat{\theta}_{\text{BIC}}$, both based on the shape and scale homogeneity constraints (with preliminary tests performed at asymptotic level $\alpha = 5\%$). Here, scale heterogeneity holds for any ξ and larger values of ξ provide increasingly severe deviations from the shape homogeneity obtained at $\xi = 0$; see Section A for details.

B Technical details

In this section, we provide the proofs of all results stated in the paper.

Proof of Theorem 3.1. Fix θ as in the statement of the theorem. In this proof, all stochastic convergences are then as $n \rightarrow \infty$ under $\mathbb{P}_{\theta}^{(n)}$. First note that since

$$\sum_{\mathbf{c} \in \mathcal{C}^m} \mathbb{I}[\phi = \mathbf{c}] = 1 \quad \text{and} \quad \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\phi =_{\mathbf{c}_0} \mathbf{c}] = 1$$

almost surely, we need to show that

$$\begin{aligned}
& \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}} - \boldsymbol{\theta}) - \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}, \mathbf{c}_0} - \boldsymbol{\theta}) \\
&= \sum_{\mathbf{c} \in \mathcal{C}^m} \mathbb{I}[\boldsymbol{\phi} = \mathbf{c}] \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta}) - \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}] \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta}) \\
&= \sum_{\mathbf{c} \in \mathcal{C}^m \setminus \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\boldsymbol{\phi} = \mathbf{c}] \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta}) + \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \{\mathbb{I}[\boldsymbol{\phi} = \mathbf{c}] - \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}]\} \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta})
\end{aligned}$$

is $o_{\mathbb{P}}(1)$. Fix first $\mathbf{c} \in \mathcal{C}^m \setminus \mathcal{C}_{\mathbf{c}_0}^m$. It is thus possible to pick $j \notin \mathcal{J}_{\mathbf{c}_0}$ such that $c_j = 0$. Since, by assumption, $\boldsymbol{\theta} \notin \mathcal{M}(\Upsilon_j)$, we then have that, for any $\varepsilon > 0$,

$$\mathbb{P}_{\boldsymbol{\theta}}^{(n)}[\mathbb{I}[\boldsymbol{\phi} = \mathbf{c}] \|\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta})\| > \varepsilon] \leq \mathbb{P}_{\boldsymbol{\theta}}^{(n)}[\boldsymbol{\phi} = \mathbf{c}] \leq \mathbb{P}_{\boldsymbol{\theta}}^{(n)}[\phi_j = 0] = \mathbb{P}_{\boldsymbol{\theta}}^{(n)}[Q_j^{(n)} \leq \chi_{d-r_j, 1-\alpha}^2] \rightarrow 0.$$

This shows that $\mathbb{I}[\boldsymbol{\phi} = \mathbf{c}] \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta})$ is $o_{\mathbb{P}}(1)$ for any $\mathbf{c} \in \mathcal{C}^m \setminus \mathcal{C}_{\mathbf{c}_0}^m$. Fix now $\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m$, so that $c_j = 1$ for any $j \notin \mathcal{J}_{\mathbf{c}_0}$. Then, for any $\varepsilon > 0$,

$$\begin{aligned}
& \mathbb{P}_{\boldsymbol{\theta}}^{(n)}[\|\{\mathbb{I}[\boldsymbol{\phi} = \mathbf{c}] - \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}]\} \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta})\| > \varepsilon] \\
&= \mathbb{P}_{\boldsymbol{\theta}}^{(n)}[[\boldsymbol{\phi} \neq \mathbf{c}] \cap [\phi_j = c_j \ \forall j \in \mathcal{J}_{\mathbf{c}_0}]] \leq \mathbb{P}_{\boldsymbol{\theta}}^{(n)}[\phi_j \neq c_j (= 1) \text{ for at least one } j \notin \mathcal{J}_{\mathbf{c}_0}] \\
&= \sum_{j \notin \mathcal{J}_{\mathbf{c}_0}} \mathbb{P}_{\boldsymbol{\theta}}^{(n)}[\phi_j \neq 1] = \sum_{j \notin \mathcal{J}_{\mathbf{c}_0}} \mathbb{P}_{\boldsymbol{\theta}}^{(n)}[Q_j^{(n)} \leq \chi_{d-r_j, 1-\alpha}^2] \rightarrow 0,
\end{aligned}$$

since, by assumption, $\boldsymbol{\theta} \notin \mathcal{M}(\Upsilon_j)$ for any $j \notin \mathcal{J}_{\mathbf{c}_0}$. The result follows. \square

Proof of Theorem 3.2. Proceeding exactly along the same lines as in the proof of Theorem 3.1, we obtain that $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}} - \boldsymbol{\theta}_n) = \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}, \mathbf{c}_0} - \boldsymbol{\theta}_n) + o_{\mathbb{P}}(1)$ as $n \rightarrow \infty$ under $\mathbb{P}_{\boldsymbol{\theta}}^{(n)}$, hence (from contiguity) also under $\mathbb{P}_{\boldsymbol{\theta}_n}^{(n)}$. With the notation introduced above

the statement of the theorem, this rewrites

$$\begin{aligned}
\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\text{PMTE}} - \boldsymbol{\theta}_n) &= \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}] \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta}_n) + o_{\text{P}}(1) \\
&= \sum_{r=1}^{2^s} \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}_r] \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}_r} - \boldsymbol{\theta}_n) + o_{\text{P}}(1) \\
&= \mathbf{W}'_n \mathbf{R}_n + o_{\text{P}}(1)
\end{aligned} \tag{B.35}$$

as $n \rightarrow \infty$ under $\text{P}_{\boldsymbol{\theta}_n}^{(n)}$, where $\mathbf{W}_n := \mathbf{W}(\mathbf{D}_{\mathbf{c}_0}^{(n)})$ is based on $\mathbf{D}^{(n)} := (\mathbf{D}_{j_1}^{(n)'}, \dots, \mathbf{D}_{j_s}^{(n)'})'$ (see Assumption (B)) and where we let

$$\mathbf{R}_n := \begin{pmatrix} \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}_1} - \boldsymbol{\theta}_n) \\ \vdots \\ \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}_{2^s}} - \boldsymbol{\theta}_n) \end{pmatrix}.$$

Since $\boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta}_n) = \boldsymbol{\nu}_n^{-1}(\hat{\boldsymbol{\theta}}_{\mathbf{c}} - \boldsymbol{\theta}) - \boldsymbol{\tau}_n$ for any \mathbf{c} , it directly follows from Assumption (B) that

$$\begin{pmatrix} \mathbf{R}_n \\ \mathbf{D}_{\mathbf{c}_0}^{(n)} \end{pmatrix} = \begin{pmatrix} \mathbf{F}\mathbf{S}_{\boldsymbol{\theta}}^{(n)} - (\mathbf{1}_{2^s} \otimes \boldsymbol{\tau}_n) \\ \mathbf{G}\mathbf{S}_{\boldsymbol{\theta}}^{(n)} \end{pmatrix} + o_{\text{P}}(1) =: \mathbf{U}_n + o_{\text{P}}(1) \tag{B.36}$$

as $n \rightarrow \infty$ under $\text{P}_{\boldsymbol{\theta}}^{(n)}$, hence (still from contiguity) also under $\text{P}_{\boldsymbol{\theta}_n}^{(n)}$. Denoting (as in Assumption (A)) as Λ_n the log-likelihood ratio of $\text{P}_{\boldsymbol{\theta}_n + \boldsymbol{\nu}_n \boldsymbol{\tau}_n}^{(n)}$ with respect to $\text{P}_{\boldsymbol{\theta}_n}^{(n)}$, it directly follows from (B.36) and Assumptions (A) and (B) that, under $\text{P}_{\boldsymbol{\theta}}^{(n)}$,

$$\begin{pmatrix} \mathbf{U}_n \\ \Lambda_n \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}, \boldsymbol{\tau}}, \mathbf{V}_{\boldsymbol{\theta}, \boldsymbol{\tau}})$$

with

$$\boldsymbol{\mu}_{\boldsymbol{\theta}, \boldsymbol{\tau}} := \begin{pmatrix} -(\mathbf{1}_{2^s} \otimes \boldsymbol{\tau}) \\ \mathbf{0} \\ -\frac{1}{2} \boldsymbol{\tau}' \boldsymbol{\Gamma}_{\boldsymbol{\theta}} \boldsymbol{\tau} \end{pmatrix} \quad \text{and} \quad \mathbf{V}_{\boldsymbol{\theta}, \boldsymbol{\tau}} = \begin{pmatrix} \mathbf{F}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{F}' & \mathbf{F}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{G}' & \mathbf{F}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau} \\ \mathbf{G}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{F}' & \mathbf{G}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{G}' & \mathbf{G}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\tau} \\ \boldsymbol{\tau}'\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{F}' & \boldsymbol{\tau}'\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{G}' & \boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau} \end{pmatrix}.$$

Thus, the Le Cam third Lemma directly yields that, under $P_{\theta_n}^{(n)}$,

$$U_n \xrightarrow{\mathcal{D}} \mathcal{N} \left(\begin{pmatrix} \mathbf{F}\Omega_{\theta}\boldsymbol{\tau} - (\mathbf{1}_{2^s} \otimes \boldsymbol{\tau}) \\ \mathbf{G}\Omega_{\theta}\boldsymbol{\tau} \end{pmatrix}, \begin{pmatrix} \mathbf{F}\Sigma_{\theta}\mathbf{F}' & \mathbf{F}\Sigma_{\theta}\mathbf{G}' \\ \mathbf{G}\Sigma_{\theta}\mathbf{F}' & \mathbf{G}\Sigma_{\theta}\mathbf{G}' \end{pmatrix} \right),$$

which is the desired result. Obviously, (B.36) implies that this is also the asymptotic distribution of $(\mathbf{R}'_n, \mathbf{D}_{\mathbf{c}_0}^{(n)'})'$ under $P_{\theta_n}^{(n)}$. The result then directly follows from (B.35) and from the continuous mapping theorem. \square

Proof of Theorem 3.3. Since

$$\mathbf{W}'\mathbf{F} = \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}] \Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta},$$

the quantities $\boldsymbol{\mu}_{\text{PMTE}}$ and $\boldsymbol{\Gamma}_{\text{PMTE}}$ in (3.11)–(3.12) are respectively given by

$$\begin{aligned} \boldsymbol{\mu}_{\text{PMTE}} &= \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}] (\Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta} \Omega_{\theta} \boldsymbol{\tau} - \boldsymbol{\tau} + \Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta} \Sigma_{\theta} \mathbf{G}' (\mathbf{G} \Sigma_{\theta} \mathbf{G}')^{-1} (\mathbf{D} - \mathbf{G} \Omega_{\theta} \boldsymbol{\tau})) \\ &= \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}] (\Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta} \Omega_{\theta} \boldsymbol{\tau} - \boldsymbol{\tau} + \Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta} \mathbf{A}_{\theta} (\mathbf{D} - \mathbf{G} \Omega_{\theta} \boldsymbol{\tau})) \\ &= \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}] \{ (\Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta} \mathbf{E}_{\theta, \boldsymbol{\tau}} - \boldsymbol{\tau}) + \Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta} \mathbf{A}_{\theta} \mathbf{D} \} \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\Gamma}_{\text{PMTE}} &= \sum_{\mathbf{c}, \mathbf{c}' \in \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}] \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}'] \Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta} \Sigma_{\theta}^{1/2} \{ \mathbf{I}_p - \mathbf{P}(\Sigma_{\theta}^{1/2} \mathbf{G}') \} \Sigma_{\theta}^{1/2} (\Upsilon_{\mathbf{c}'} \mathbf{B}_{\mathbf{c}', \theta})' \\ &= \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}] \Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta} \Sigma_{\theta}^{1/2} (\mathbf{I}_p - \mathbf{P}(\Sigma_{\theta}^{1/2} \mathbf{G}')) \Sigma_{\theta}^{1/2} (\Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta})', \end{aligned}$$

where we used the fact that $\mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}] \mathbb{I}[\boldsymbol{\phi} =_{\mathbf{c}_0} \mathbf{c}'] = 0$ for $\mathbf{c} \neq \mathbf{c}'$. We therefore have that

$$\mathbb{E}[\boldsymbol{\Gamma}_{\text{PMTE}}] = \sum_{\mathbf{c} \in \mathcal{C}_{\mathbf{c}_0}^m} \lambda_{\mathbf{c}, \boldsymbol{\tau}} \Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta} \Sigma_{\theta}^{1/2} \{ \mathbf{I}_p - \mathbf{P}(\Sigma_{\theta}^{1/2} \mathbf{G}') \} \Sigma_{\theta}^{1/2} (\Upsilon_{\mathbf{c}} \mathbf{B}_{\mathbf{c}, \theta})'$$

and that

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\mu}_{\text{PMTE}}\boldsymbol{\mu}'_{\text{PMTE}}] &= \sum_{\mathbf{c} \in \mathcal{C}_{c_0}^m} \lambda_{\mathbf{c},\tau} (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta},\tau} - \boldsymbol{\tau}) (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta},\tau} - \boldsymbol{\tau})' \\
&\quad + \boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{A}_{\boldsymbol{\theta}} \lambda_{\mathbf{c},\tau} (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta},\tau} - \boldsymbol{\tau})' + (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta},\tau} - \boldsymbol{\tau}) (\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{A}_{\boldsymbol{\theta}} \lambda_{\mathbf{c},\tau})' \\
&\quad + \boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}} \mathbf{A}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}_{\mathbf{c},\tau} \mathbf{A}'_{\boldsymbol{\theta}} \mathbf{B}'_{\mathbf{c},\boldsymbol{\theta}} \boldsymbol{\Upsilon}'_{\mathbf{c}},
\end{aligned}$$

so that the result follows from the identity $\text{AMSE}_{\boldsymbol{\theta},\tau}(\hat{\boldsymbol{\theta}}_{\text{PMTE}}) = \mathbb{E}[\boldsymbol{\Gamma}_{\text{PTE}}] + \mathbb{E}[\boldsymbol{\mu}_{\text{PMTE}}\boldsymbol{\mu}'_{\text{PMTE}}]$.

□

We now turn to the proof of Proposition 4.1, which requires the following preliminary result (see page 32 in Saleh, 2006).

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

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

and

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

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

Proof of Proposition 4.1. Since the constrained and unconstrained estimators in Section 4 are asymptotically efficient (in the corresponding constrained and unconstrained Gaussian models, respectively), $\boldsymbol{\Upsilon}_{\mathbf{c}} \mathbf{B}_{\mathbf{c},\boldsymbol{\theta}}$ is given by $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}$, $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2} \mathbf{P}_{\boldsymbol{\Upsilon}_{\text{scale}}} \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}$, $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2} \mathbf{P}_{\boldsymbol{\Upsilon}_{\text{shape}}} \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}$, and $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2} \mathbf{P}_{\boldsymbol{\Upsilon}_{\text{cov}}} \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}$, for $\mathbf{c} = (1, 1)$, $\mathbf{c} = (0, 1)$, $\mathbf{c} = (1, 0)$, and $\mathbf{c} = (0, 0)$, respectively. Easy computations yield

$$\mathbf{I}_d - \mathbf{P}(\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{1/2} \mathbf{G}') = \mathbf{I}_d - \mathbf{P}_{\boldsymbol{\Upsilon}_{\text{shape}}}^{\perp} - \mathbf{P}_{\boldsymbol{\Upsilon}_{\text{scale}}}^{\perp},$$

where we let $\mathbf{P}_{\mathbf{r}}^\perp := \mathbf{I}_d - \mathbf{P}(\mathbf{\Gamma}_\theta^{1/2} \mathbf{r}_n^{-1} \mathbf{\Upsilon})$. Moreover, the matrices $\mathbf{\Lambda}_{\mathbf{c}, \mathbf{0}}$ can be easily computed from Lemma B.1 (these matrices are block-diagonal with each diagonal block proportional to the identity matrix). Straightforward computations then show that (3.15) reduces to

$$\begin{aligned} \text{AMSE}_{\theta, \mathbf{0}}(\hat{\boldsymbol{\theta}}_{\text{PMTE}}) &= \mathbf{\Gamma}_\theta^{-1/2} \{ \alpha^2 \mathbf{I}_d + \mathbf{P}_{\mathbf{r}_{\text{scale}}} (\alpha(1-\alpha) \mathbf{P}_{\mathbf{r}_{\text{shape}}} + \gamma_{\text{shape}, (0,1)} \mathbf{P}_{\mathbf{r}_{\text{shape}}}^\perp) \mathbf{P}_{\mathbf{r}_{\text{scale}}} \\ &\quad + (\gamma_{\text{shape}, (1,1)} - \alpha^2) \mathbf{P}_{\mathbf{r}_{\text{shape}}}^\perp + \mathbf{P}_{\mathbf{r}_{\text{shape}}} (\alpha(1-\alpha) \mathbf{P}_{\mathbf{r}_{\text{scale}}} + \gamma_{\text{scale}, (1,0)} \mathbf{P}_{\mathbf{r}_{\text{scale}}}^\perp) \mathbf{P}_{\mathbf{r}_{\text{shape}}} \\ &\quad + (1-\alpha)^2 \mathbf{P}_{\mathbf{r}_{\text{cov}}} + (\gamma_{\text{scale}, (1,1)} - \alpha^2) \mathbf{P}_{\mathbf{r}_{\text{scale}}}^\perp \} \mathbf{\Gamma}_\theta^{-1/2}, \end{aligned}$$

with $\gamma_{\text{scale}, (1,1)} = \alpha(1 - \Psi_{k-1+2}(\chi_{k-1,1-\alpha}^2))$, $\gamma_{\text{shape}, (1,1)} = \alpha(1 - \Psi_{(k-1)b_p+2}(\chi_{(k-1)b_p,1-\alpha}^2))$, $\gamma_{\text{shape}, (0,1)} = (1-\alpha)(1 - \Psi_{(k-1)b_p+2}(\chi_{(k-1)b_p,1-\alpha}^2))$ and $\gamma_{\text{scale}, (1,0)} = (1-\alpha)(1 - \Psi_{k-1+2}(\chi_{k-1,1-\alpha}^2))$.

It then follows from (4.26) that

$$\begin{aligned} \text{AMSE}_{\theta, \mathbf{0}}^{\text{scalar}}(\hat{\boldsymbol{\theta}}_{\text{PMTE}}) &= \alpha^2(b_p + 1) + (1 - \alpha^2)(b_p + 1) + \gamma_{\text{scale}, (1,1)}(k-1) + \gamma_{\text{shape}, (1,1)}(k-1)b_p \\ &\quad + \gamma_{\text{shape}, (0,1)}(k-1)b_p + \gamma_{\text{scale}, (1,0)}(k-1) \\ &= d - (k-1)b_p \Psi_{(k-1)b_p+2}(\chi_{(k-1)b_p,1-\alpha}^2) - (k-1) \Psi_{k-1+2}(\chi_{k-1,1-\alpha}^2), \end{aligned}$$

which is the desired result. \square