Semiparametrically efficient estimation of constrained Euclidean parameters

Susyanto, N.; Klaassen, C.A.J.

DOI
10.1214/17-EJS1308
Publication date
2017
Document Version
Final published version
Published in
Electronic Journal of Statistics

Electronic Journal of Statistics
Vol. 11 (2017) 3120–3140
ISSN: 1935-7524
DOI: 10.1214/17-EJS1308

Semiparametrically efficient estimation of constrained Euclidean parameters*

Nanang Susyanto†

Department of Mathematics, Universitas Gadjah Mada
Sekip Utara, Yogyakarta, Indonesia
e-mail: nanang_susyanto@ugm.ac.id

Korteweg-de Vries Institute for Mathematics, University of Amsterdam
P.O. Box 94248, 1090 GE Amsterdam, The Netherlands

and

Chris A. J. Klaassen

Korteweg-de Vries Institute for Mathematics, University of Amsterdam
P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
e-mail: c.a.j.klaasen@uva.nl

Abstract: Consider a quite arbitrary (semi)parametric model for i.i.d. observations with a Euclidean parameter of interest and assume that an asymptotically (semi)parametrically efficient estimator of it is given. If the parameter of interest is known to lie on a general surface (image of a continuously differentiable vector valued function), we have a submodel in which this constrained Euclidean parameter may be rewritten in terms of a lower-dimensional Euclidean parameter of interest. An estimator of this underlying parameter is constructed based on the given estimator of the original Euclidean parameter, and it is shown to be (semi)parametrically efficient. It is proved that the efficient score function for the underlying parameter is determined by the efficient score function for the original parameter and the Jacobian of the function defining the general surface, via a chain rule for score functions. Efficient estimation of the constrained Euclidean parameter itself is considered as well.

Our general estimation method is applied to location-scale, Gaussian copula and semiparametric regression models, and to parametric models.

MSC 2010 subject classifications: Primary 62F30, 62F12; secondary 62F10.

Keywords and phrases: semiparametric estimation, semiparametric submodels, efficient estimator, restricted parameter, underlying parameter, Gaussian copula.

Received September 2016.

*This research was supported by the Netherlands Organisation for Scientific Research (NWO) via the project Forensic Face Recognition, 727.011.008.

†Corresponding author.

Contents

1. Introduction

Let $X_1, \ldots, X_n$ be i.i.d. copies of X taking values in the measurable space $(\mathcal{X}, \mathcal{A})$ in a semiparametric model with Euclidean parameter $\theta \in \Theta$ where Θ is an open subset of $\mathbb{R}^k$. We denote this semiparametric model by

$$\mathcal{P} = \{P_{\theta,G} \ : \ \theta \in \Theta,\ G \in \mathcal{G}\}. \tag{1.1}$$

Typically, the nuisance parameter space $\mathcal{G}$ is a subset of a Banach or Hilbert space. This space may also be finite dimensional, thus resulting in a parametric model.

We assume an asymptotically efficient estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ is given of the parameter of interest θ, which under regularity conditions means that

$$\sqrt{n}\left(\hat{\theta}_n - \theta - \frac{1}{n}\sum_{i=1}^{n} \tilde{\ell}(X_i; \theta, G, \mathcal{P})\right) \to_{P_{\theta,G}} 0 \tag{1.2}$$

holds. Here $\tilde{\ell}(\cdot; \theta, G, \mathcal{P})$ is the efficient influence function at $P_{\theta,G}$ for estimation of θ within $\mathcal{P}$ and

$$\dot{\ell}(\cdot; \theta, G, \mathcal{P}) = \left(\int_{\mathcal{X}} \tilde{\ell}(x; \theta, G, \mathcal{P})\tilde{\ell}^T(x; \theta, G, \mathcal{P}) dP_{\theta,G}(x)\right)^{-1} \tilde{\ell}(\cdot; \theta, G, \mathcal{P}) \tag{1.3}$$

is the corresponding efficient score function at $P_{\theta,G}$ for estimation of θ within $\mathcal{P}$.

The topic of this paper is asymptotically efficient estimation when it is known that θ lies on a general surface, or equivalently, when it is known that θ is determined by a lower dimensional parameter via a continuously differentiable function, which we denote by

$$\theta = f(\nu), \quad \nu \in N. \tag{1.4}$$

Here $f : N \subset \mathbb{R}^d \to \mathbb{R}^k$ with $d < k$ is known, N is open, the Jacobian

$$\dot{f}(\nu) = \left(\frac{\partial f_i(\nu)}{\partial \nu_j}\right)_{j=1,\ldots,d}^{i=1,\ldots,k} \tag{1.5}$$

of f is assumed to be of full rank on N, and ν is the unknown d-dimensional parameter to be estimated. Thus, we focus on the (semi)parametric model

$$\mathcal{Q} = \left\{ P_{f(\nu),G} \ : \ \nu \in N, \ G \in \mathcal{G} \right\} \subset \mathcal{P}. \tag{1.6}$$

In order for ν to be identifiable we have to assume that $f(\cdot)$ is invertible; note that θ itself is identifiable as it is assumed that it can be estimated efficiently.

The first main result of this paper is that a semiparametrically efficient estimator of ν, the parameter of interest, has to be asymptotically linear with efficient score function for estimation of ν equal to

$$\dot{\ell}(\cdot; \nu, G, \mathcal{Q}) = \dot{f}^T(\nu)\dot{\ell}(\cdot; \theta, G, \mathcal{P}). \tag{1.7}$$

Such a semiparametrically efficient estimator of the parameter of interest can be defined in terms of $f(\cdot)$ and the efficient estimator $\hat{\theta}_n$ of θ; see equation (4.1) in Section 4. This is our second main result.

How (1.7) is related to the chain rule for differentiation will be explained in Section 2, which proves this chain rule for score functions within regular parametric (sub)models. The semiparametric lower bound for estimators of ν is obtained via the Hájek-LeCam Convolution Theorem for regular parametric models and without projection techniques in Section 3. In Section 4 efficient estimators within $\mathcal{Q}$ of ν and $\theta = f(\nu)$ are constructed. The generality of our results facilitates the analysis of numerous statistical models. We discuss some of such parametric and semiparametric models and related literature in Section 5. Technicalities are collected in Appendices A and B.

Several examples of estimation under constraints for *nonparametric* models have been studied in [2]. The efficient influence function for such a constrained nonparametric model is determined by projection of the efficient influence function for the unconstrained model on the tangent space of the constrained model; see e.g. Example 3.3.3 of [2]. The constraints are formulated via equations the distributions should satisfy. Some such equations for distributions can be reformulated for semiparametric models in terms of equations for the Euclidean parameter. For semiparametric models constrained by such equations the efficient influence function can also be determined by projection of the efficient influence function for the unconstrained semiparametric model, as we will show in a companion paper; see [12]. Quite many semiparametric models with constraints by equations the Euclidean parameters should satisfy, can be reparametrized as in (1.6), but not all. Likewise, not all submodels (1.6) can be phrased via equations. Simple counterexamples to prove these claims are given in the companion paper. In these counterexamples the condition that N be open, is crucial. This looks like a minor feature. However, in asymptotic statistics one typically assumes the parameter space to be open in order to avoid (interesting) pathologies at the boundary.

Therefore the topics of the present paper and its companion one do not coincide completely. We do not use projection techniques in the present paper, but base our approach directly on the concept of least favorable submodel and on the parametric version of the Hájek-LeCam convolution theorem, and not on its generalization to the semiparametric situation as given by Theorem 3.3.2 and Theorem 3.4.1 of [2] and by Theorem 25.20 of [19]. This new approach seems to be well suited to the formulation of the main Theorem 3.1 and it makes our proofs elementary and pretty straightforward.

Most, if not all, papers on estimation in constrained *parametric* models focus on constrained (or restricted) maximum likelihood estimation implemented via Lagrange multipliers. The first paper on this subject seems [1]. Another early treatise related to the theme of the present paper for the parametric case is [15]. This book studies classical and Bayesian estimation for parametric models under constraints in terms of equalities and inequalities.

The topic of the present paper should not be confused with estimation of the parameter θ when it is known to lie in a subset with nonempty interior of the original parameter space. This situation corresponds to $d = k$ with $f(\cdot)$ the identity and $\nu = \theta$ in (1.4)–(1.6). If an asymptotically efficient estimator $\hat{\theta}_n$ is given for the unconstrained model, this estimator is also asymptotically efficient within the constrained model, as N is open and hence $\hat{\theta}_n$ takes values in N with probability tending to 1. A comprehensive treatment of finite sample estimation problems with N a proper subset of Θ with the same dimension, may be found in [20].

2. The chain rule for score functions

The basic building block for the asymptotic theory of semiparametric models as presented in e.g. [2] is the concept of regular parametric model. Let $\mathcal{P}_\Theta = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$ open be a parametric model with all P_θ dominated by a σ-finite measure μ on $(\mathcal{X}, \mathcal{A})$. Denote the density of P_θ with respect to μ by $p(\theta) = p(\cdot; \theta, \mathcal{P}_\Theta)$ and the $L_2(\mu)$-norm by $\| \cdot \|_\mu$. If for each $\theta_0 \in \Theta$ there exists a k-dimensional column vector $\dot{\ell}(\theta_0, \mathcal{P}_\Theta)$ of elements of $L_2(P_{\theta_0})$, the so-called score function, such that the Fréchet differentiability

$$\begin{aligned} \| \sqrt{p(\theta)} - \sqrt{p(\theta_0)} - \tfrac{1}{2} (\theta - \theta_0)^T \dot{\ell}(\theta_0, \mathcal{P}_\Theta) \sqrt{p(\theta_0)} \|_\mu & \\ = o(|\theta - \theta_0|), \quad \theta \to \theta_0, & \end{aligned} \tag{2.1}$$

holds and the $k \times k$ Fisher information matrix

$$I(\theta_0) = \int_{\mathcal{X}} \dot{\ell}(\theta_0, \mathcal{P}_\Theta) \dot{\ell}^T(\theta_0, \mathcal{P}_\Theta) dP_{\theta_0} \tag{2.2}$$

is nonsingular, and, moreover, the map $\theta \mapsto \dot{\ell}(\theta, \mathcal{P}_\Theta)\sqrt{p(\theta)}$ from Θ to $L_2^k(\mu)$ is continuous, then $\mathcal{P}_\Theta$ is called a *regular parametric* model. Often the score function may be determined by computing the logarithmic derivative of the density with respect to θ; cf. Proposition 2.1.1 of [2]. We will call $\mathcal{P}$ from (1.1) a *regular semiparametric* model if for all $G \in \mathcal{G}$

$$\mathcal{P}_{\Theta,G} = \{P_{\theta,G} : \theta \in \Theta\} \tag{2.3}$$

is a regular parametric model.

Fix $\theta_0 \in \Theta$ and $G_0 \in \mathcal{G}$, and write $P_{\theta_0,G_0} = P_0$. Let $\psi : \Theta \to \mathcal{G}$ with $\psi(\theta_0) = G_0$ be such that

$$\mathcal{P}_\psi = \left\{ P_{\theta,\psi(\theta)} \,:\, \theta \in \Theta \right\} \tag{2.4}$$

is a regular parametric submodel of $\mathcal{P}$ with score function $\dot{\ell}(\theta_0, \mathcal{P}_\psi)$ at θ_0 and Fisher information matrix $I(\theta_0, \mathcal{P}_\psi)$, say. Let the density of $P_{\theta,\psi(\theta)}$ with respect to μ be denoted by $q(\theta)$. Since $\mathcal{P}_\psi$ is a regular parametric model the score function $\dot{\ell}(\theta_0, \mathcal{P}_\psi)$ for θ at θ_0 within $\mathcal{P}_\psi$ satisfies (cf. (2.1))

$$\begin{aligned} \| \sqrt{q(\theta)} - \sqrt{q(\theta_0)} - \tfrac{1}{2} (\theta - \theta_0)^T \, \dot{\ell}(\theta_0, \mathcal{P}_\psi) \sqrt{q(\theta_0)} \, \|_\mu \\ = o(|\theta - \theta_0|), \quad \theta \to \theta_0. \end{aligned} \tag{2.5}$$

Considering now the (semi)parametric submodel $\mathcal{Q}$ from(1.6) we fix ν_0 and write $f(\nu_0) = \theta_0$ and $f(\nu) = \theta$. Within $\mathcal{Q}$ the Fréchet differentiability (2.5) yields

$$\begin{aligned} \| \sqrt{q(f(\nu))} - \sqrt{q(f(\nu_0))} - \tfrac{1}{2} (f(\nu) - f(\nu_0))^T \, \dot{\ell}(f(\nu_0), \mathcal{P}_\psi) \sqrt{q(f(\nu_0))} \, \|_\mu \\ = o(|f(\nu) - f(\nu_0)|), \quad f(\nu) \to f(\nu_0), \end{aligned} \tag{2.6}$$

and hence

$$\begin{aligned} \| \sqrt{q(f(\nu))} - \sqrt{q(f(\nu_0))} - \tfrac{1}{2} (\nu - \nu_0)^T \dot{f}^T(\nu_0) \dot{\ell}(\theta_0, \mathcal{P}_\psi) \sqrt{q(f(\nu_0))} \, \|_\mu \\ = o(|\nu - \nu_0|), \quad \nu \to \nu_0, \end{aligned} \tag{2.7}$$

in view of the differentiability of $f(\cdot)$. Since $\dot{f}(\cdot)$ is continuous, this means that

$$\mathcal{Q}_\psi = \left\{ P_{f(\nu),\psi(f(\nu))} \,:\, \nu \in N \right\} \tag{2.8}$$

is a regular parametric submodel of $\mathcal{Q}$ with score function

$$\dot{\ell}(\nu_0, \mathcal{Q}_\psi) = \dot{f}^T(\nu_0) \dot{\ell}(\theta_0, \mathcal{P}_\psi) \tag{2.9}$$

for ν at P_0 and Fisher information matrix

$$\dot{f}^T(\nu_0) I(\theta_0, \mathcal{P}_\psi) \dot{f}(\nu_0) = \dot{f}^T(\nu_0) \int_{\mathcal{X}} \dot{\ell}(\theta_0, \mathcal{P}_\psi) \dot{\ell}^T(\theta_0, \mathcal{P}_\psi) dP_0 \; \dot{f}(\nu_0). \tag{2.10}$$

We have proved

Proposition 2.1. *Let $\mathcal{P}$ as in (1.1) be a regular semiparametric model and let $\mathcal{Q}$ as in (1.6) be a regular semiparametric submodel with $f : N \subset \mathbb{R}^d \to \mathbb{R}^k$, $d < k$, N open, a continuously differentiable function and the Jacobian $\dot{f}(\cdot)$ of full rank defined as in (1.4) and (1.5). If there exists a regular parametric submodel $\mathcal{P}_\psi$ of $\mathcal{P}$ with score function $\dot{\ell}(\theta_0, \mathcal{P}_\psi)$ for θ at $\theta_0 = f(\nu_0)$, then there exists a regular parametric submodel $\mathcal{Q}_\psi$ of $\mathcal{Q}$ with score function $\dot{\ell}(\nu_0, \mathcal{Q}_\psi)$ for ν at ν_0 satisfying (2.9).*

This Proposition is also valid for parametric models, as may be seen by choosing $\mathcal{G}$ finite dimensional or even degenerate. The basic version of the chain rule for score functions is for such a parametric model $\mathcal{P}_\Theta$. We have chosen the more elaborate formulation of Proposition 2.1 since we are going to apply the chain rule for such parametric submodels $\mathcal{P}_\psi$ of semiparametric models $\mathcal{P}$.

3. Convolution theorem and main result

An estimator $\hat{\theta}_n$ of θ within the regular semiparametric model $\mathcal{P}$ is called (locally) regular at $P_0 = P_{\theta_0,G_0}$ if it is (locally) regular at P_0 within $\mathcal{P}_\psi$ for all regular parametric submodels $\mathcal{P}_\psi$ of $\mathcal{P}$ containing P_{Θ,G_0}. According to the Hájek-LeCam Convolution Theorem for regular parametric models (see e.g. Section 2.3, in particular the Note on page 27, of [2]) this implies that for such a regular estimator $\hat{\theta}_n$ of θ within $\mathcal{P}$ the normed estimation error $\sqrt{n}(\hat{\theta}_n - \theta_0)$ has a limit distribution under P_0 that is the convolution of a normal distribution with mean 0 and covariance matrix $I^{-1}(\theta_0, \mathcal{P}_\psi)$ and another distribution, for any regular parametric submodel $\mathcal{P}_\psi$ containing P_0. If there exists $\psi = \psi_0$ such that this last distribution is degenerate at 0, we call $\hat{\theta}_n$ (locally) efficient at P_0 and $\mathcal{P}_{\psi_0}$ a least favorable parametric submodel for estimation of θ within $\mathcal{P}$ at P_0. Then the Hájek-LeCam Convolution Theorem also implies that $\hat{\theta}_n$ is asymptotically linear in the efficient influence function $\tilde{\ell}(\theta_0, G_0, \mathcal{P}) = \tilde{\ell}(\cdot; \theta_0, G_0, \mathcal{P})$ satisfying

$$\tilde{\ell}(\theta_0, G_0, \mathcal{P}) = \tilde{\ell}(\theta_0, \mathcal{P}_{\psi_0}) = I^{-1}(\theta_0, \mathcal{P}_{\psi_0})\dot{\ell}(\theta_0, \mathcal{P}_{\psi_0}), \tag{3.1}$$

which means

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0 - \frac{1}{n}\sum_{i=1}^{n} \tilde{\ell}(X_i; \theta_0, G_0, \mathcal{P})\right) \to_{P_0} 0. \tag{3.2}$$

The argument above can be extended to the more general situation that there exists a least favorable sequence of parametric submodels indexed by ψ_j , $j = 1, 2, \dots$, such that the corresponding score functions $\dot{\ell}(\theta_0, \mathcal{P}_{\psi_j})$ for θ at θ_0 within model $\mathcal{P}_{\psi_j}$ converge in $L_2^k(P_0)$ to $\dot{\ell}(\theta_0, G_0, \mathcal{P}) = \dot{\ell}(\cdot; \theta_0, G_0, \mathcal{P})$, say. A regular estimator $\hat{\theta}_n$ of θ within $\mathcal{P}$ is called efficient then, if it is asymptotically linear as in (3.2) with efficient influence function $\tilde{\ell}(\theta_0, G_0, \mathcal{P}) = \tilde{\ell}(\cdot; \theta_0, G_0, \mathcal{P})$ satisfying

$$\begin{aligned}\tilde{\ell}(\theta_0, G_0, \mathcal{P}) &= \left(\int_{\mathcal{X}} \dot{\ell}(\theta_0, G_0, \mathcal{P})\dot{\ell}^T(\theta_0, G_0, \mathcal{P})dP_0\right)^{-1} \dot{\ell}(\theta_0, G_0, \mathcal{P}) \\ &= I^{-1}(\theta_0, G_0, \mathcal{P})\dot{\ell}(\theta_0, G_0, \mathcal{P}).\end{aligned} \tag{3.3}$$

Indeed, by the Convolution Theorem for regular parametric models the convergence

$$\begin{pmatrix} \sqrt{n}\left(\hat{\theta}_n - \theta_0 - \frac{1}{n}\sum\limits_{i=1}^{n} \tilde{\ell}(X_i; \theta_0, \mathcal{P}_{\psi_j})\right) \\ \frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} \tilde{\ell}(X_i; \theta_0, \mathcal{P}_{\psi_j}) \end{pmatrix} \to_{P_0} \begin{pmatrix} R_{\mathcal{P},j} \\ Z_{\mathcal{P},j} \end{pmatrix} \tag{3.4}$$

holds with the k-vectors $R_{\mathcal{P},j}$ and $Z_{\mathcal{P},j}$ independent and $Z_{\mathcal{P},j}$ normal with mean 0 and covariance matrix $I^{-1}(\theta_0, \mathcal{P}_{\psi_j})$. Taking limits as $j \to \infty$ we see by tightness arguments and by the convergence of $\dot{\ell}(\theta_0, \mathcal{P}_{\psi_j})$ to $\dot{\ell}(\theta_0, G_0, \mathcal{P})$ in

$L_2^k(P_0)$, that also

$$\begin{pmatrix} \sqrt{n}\left(\hat{\theta}_n - \theta_0 - \frac{1}{n}\sum_{i=1}^n \tilde{\ell}(X_i;\theta_0,G_0,\mathcal{P})\right) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\ell}(X_i;\theta_0,G_0,\mathcal{P}) \end{pmatrix} \to_{P_0} \begin{pmatrix} R_{\mathcal{P}} \\ Z_{\mathcal{P}} \end{pmatrix} \tag{3.5}$$

holds with $R_{\mathcal{P}}$ and $Z_{\mathcal{P}}$ independent k-vectors and $Z_{\mathcal{P}}$ normally distributed with mean 0 and covariance matrix $I^{-1}(\theta_0, G_0, \mathcal{P})$. To be more precise, consider the difference between the left hand sides of (3.5) and (3.4). The second half of this vector of differences equals minus the first half. Both halves converge in distribution by the central limit theorem to a normal distribution with mean 0 and covariance matrix

$$E\left(\left[\tilde{\ell}(X_1;\theta_0,G_0,\mathcal{P}) - \tilde{\ell}(X_1;\theta_0,\mathcal{P}_{\psi_j})\right]\left[\tilde{\ell}(X_1;\theta_0,G_0,\mathcal{P}) - \tilde{\ell}(X_1;\theta_0,\mathcal{P}_{\psi_j})\right]^T\right) \tag{3.6}$$

as $n \to \infty$. This implies that the vector of differences is tight and, as the left hand side of (3.4) is tight, that the left hand side of (3.5) is tight as well. Let $(R_{\mathcal{P}}^T, Z_{\mathcal{P}}^T)^T$ be a limit point of the left hand side of (3.5). As (3.6) converges to 0 for $j \to \infty$, this limit point is also the limit in distribution of $(R_{\mathcal{P},j}^T, Z_{\mathcal{P},j}^T)^T$. Consequently, all limit points $(R_{\mathcal{P}}^T, Z_{\mathcal{P}}^T)^T$ have the same distribution. By the independence of $R_{\mathcal{P},j}$ and $Z_{\mathcal{P},j}$ for all j we obtain

$$\begin{aligned} E\left(\exp\left\{is^T R_{\mathcal{P}} + it^T Z_{\mathcal{P}}\right\}\right) &= \lim_{j\to\infty} E\left(\exp\left\{is^T R_{\mathcal{P},j} + it^T Z_{\mathcal{P},j}\right\}\right) \\ &= \lim_{j\to\infty} E\left(\exp\left\{is^T R_{\mathcal{P},j}\right\}\right) E\left(\exp\left\{is^T Z_{\mathcal{P},j}\right\}\right) \\ &= E\left(\exp\left\{is^T R_{\mathcal{P}}\right\}\right) E\left(\exp\left\{it^T Z_{\mathcal{P}}\right\}\right) \end{aligned} \tag{3.7}$$

and hence (3.5) with $R_{\mathcal{P}}$ and $Z_{\mathcal{P}}$ independent. This independence turns (3.5) into a convolution theorem.

If $R_{\mathcal{P}}$ is degenerate at 0, then $\hat{\theta}_n$ is locally asymptotically efficient at P_0 within $\mathcal{P}$ and the sequence of regular parametric submodels $\mathcal{P}_{\psi_j}$ is least favorable indeed.

Now, let us assume such a least favorable sequence and efficient estimator $\hat{\theta}_n$ exist at $P_0 = P_{\theta_0,G_0}$ with $\theta_0 = f(\nu_0)$ and $f(\cdot)$ from (1.4) and (1.5) continuously differentiable. By the chain rule for score functions from Proposition 2.1 the score function $\dot{\ell}(\nu_0, \mathcal{Q}_{\psi_j})$ for ν at ν_0 within $\mathcal{Q}_{\psi_j}$ satisfies

$$\dot{\ell}(\nu_0, \mathcal{Q}_{\psi_j}) = \dot{f}^T(\nu_0)\dot{\ell}(\theta_0, \mathcal{P}_{\psi_j}) \tag{3.8}$$

and hence the corresponding influence function $\tilde{\ell}(\nu_0, \mathcal{Q}_{\psi_j})$ satisfies

$$\tilde{\ell}(\nu_0, \mathcal{Q}_{\psi_j}) = \left(\dot{f}^T(\nu_0) I(\theta_0, \mathcal{P}_{\psi_j}) \dot{f}(\nu_0)\right)^{-1} \dot{f}^T(\nu_0)\dot{\ell}(\theta_0, \mathcal{P}_{\psi_j}). \tag{3.9}$$

Let $\hat{\nu}_n$ be a locally regular estimator of ν at P_0 within the regular semiparametric model $\mathcal{Q}$. By the Convolution Theorem for regular parametric models the convergence

$$\begin{pmatrix} \sqrt{n}\left(\hat{\nu}_n - \nu_0 - \frac{1}{n}\sum_{i=1}^n \tilde{\ell}(X_i;\nu_0,\mathcal{Q}_{\psi_j})\right) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\ell}(X_i;\nu_0,\mathcal{Q}_{\psi_j}) \end{pmatrix} \rightarrow_{P_0} \begin{pmatrix} R_{\mathcal{Q},j} \\ Z_{\mathcal{Q},j} \end{pmatrix} \tag{3.10}$$

holds with the k-vectors $R_{\mathcal{Q},j}$ and $Z_{\mathcal{Q},j}$ independent and $Z_{\mathcal{Q},j}$ normal with mean 0 and covariance matrix $I^{-1}(\theta_0,\mathcal{Q}_{\psi_j})$. By the convergence of $\dot{\ell}(\theta_0,\mathcal{P}_{\psi_j})$ to $\dot{\ell}(\theta_0,G_0,\mathcal{P})$ in $L_2^k(P_0)$, the influence functions from (3.9) converge in $L_2^d(P_0)$ to

$$\begin{aligned} \tilde{\ell}(\nu_0,G_0,\mathcal{Q}) &= \left(\dot{f}^T(\nu_0)I(\theta_0,G_0,\mathcal{P})\dot{f}(\nu_0)\right)^{-1}\dot{f}^T(\nu_0)\dot{\ell}(\theta_0,G_0,\mathcal{P}) \\ &= \left(\dot{f}^T(\nu_0)I(\theta_0,G_0,\mathcal{P})\dot{f}(\nu_0)\right)^{-1}\dot{f}^T(\nu_0)I(\theta_0,G_0,\mathcal{P})\tilde{\ell}(\theta_0,G_0,\mathcal{P}) \end{aligned} \tag{3.11}$$

and the argument leading to (3.5) yields the convolution theorem

$$\begin{pmatrix} \sqrt{n}\left(\hat{\nu}_n - \nu_0 - \frac{1}{n}\sum_{i=1}^n \tilde{\ell}(X_i;\nu_0,G_0,\mathcal{Q})\right) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\ell}(X_i;\nu_0,G_0,\mathcal{Q}) \end{pmatrix} \rightarrow_{P_0} \begin{pmatrix} R_{\mathcal{Q}} \\ Z_{\mathcal{Q}} \end{pmatrix} \tag{3.12}$$

with $R_{\mathcal{Q}}$ and $Z_{\mathcal{Q}}$ independent. Note that $Z_{\mathcal{Q}}$ has a normal distribution with mean 0 and covariance matrix

$$I^{-1}(\nu_0,G_0,\mathcal{Q}) = \left(\dot{f}^T(\nu_0)I(\theta_0,G_0,\mathcal{P})\dot{f}(\nu_0)\right)^{-1}. \tag{3.13}$$

Under an additional condition on $f(\cdot)$ we shall construct an estimator $\hat{\nu}_n$ of ν based on $\hat{\theta}_n$ for which $R_{\mathcal{Q}}$ is degenerate. This construction of $\hat{\nu}_n$ will be given in the next section together with a proof of its efficiency, and this will complete the proof of our main result formulated as follows.

Theorem 3.1. *Let $\mathcal{P}$ from (1.1) be a regular semiparametric model with $P_0 = P_{\theta_0,G_0} \in \mathcal{P}, \theta_0 = f(\nu_0)$, and $f(\cdot)$ from (1.4) and (1.5) continuously differentiable. Furthermore, let $f(\cdot)$ have an inverse on $f(N)$ that is differentiable with a bounded Jacobian. If there exists a least favorable sequence of regular parametric submodels $\mathcal{P}_{\psi_j}$ and an asymptotically efficient estimator $\hat{\theta}_n$ of θ satisfying (3.5) with $R_{\mathcal{P}} = 0$ a.s., then there exists a least favorable sequence of regular parametric submodels $\mathcal{Q}_{\psi_j}$ of the restricted model $\mathcal{Q}$ from (1.6) and an asymptotically efficient estimator $\hat{\nu}_n$ of ν taking values in N, satisfying (3.12) with $R_{\mathcal{Q}} = 0$ a.s., and attaining the asymptotic information bound (3.13).*

Note that (3.11) and (3.12) with $R_{\mathcal{Q}} = 0$ a.s. imply the chain rule for score functions as formulated in (1.7).

4. Efficient estimator of the parameter of interest

For many specific types of (semi)parametric problems methods to construct efficient estimators have been devised. A general approach is upgrading a $\sqrt{n}$-consistent estimator as in Sections 2.5 (the parametric case) and 7.8 (the general case) of [2]. A somewhat different upgrading approach is used in the following construction.

Theorem 4.1. *Consider the situation of Theorem 3.1, where $\hat{\theta}_n$ is an efficient estimator of θ within the model $\mathcal{P}$. If the symmetric positive definite $k\times k$-matrix $\hat{I}_n$ is a consistent estimator of $I(\theta, G, \mathcal{P})$ within $\mathcal{P}$ and $\bar{\nu}_n$ is a $\sqrt{n}$-consistent estimator of ν within $\mathcal{Q}$, then*

$$\hat{\nu}_n = \bar{\nu}_n + \left(\dot{f}^T(\bar{\nu}_n)\hat{I}_n\dot{f}(\bar{\nu}_n)\right)^{-1}\dot{f}^T(\bar{\nu}_n)\hat{I}_n\left[\hat{\theta}_n - f\left(\bar{\nu}_n\right)\right] \tag{4.1}$$

is efficient, i.e., it satisfies (3.12) with $R_{\mathcal{Q}} = 0$ a.s.

Proof. The continuity of $\dot{f}(\cdot)$ and the consistency of $\bar{\nu}_n$ and $\hat{I}_n$ imply that

$$\hat{K}_n = \left(\dot{f}^T(\bar{\nu}_n)\hat{I}_n\dot{f}(\bar{\nu}_n)\right)^{-1}\dot{f}^T(\bar{\nu}_n)\hat{I}_n \tag{4.2}$$

converges in probability under P_0 to

$$K_0 = \left(\dot{f}^T(\nu_0)I(\theta_0, G_0, \mathcal{P})\dot{f}(\nu_0)\right)^{-1}\dot{f}^T(\nu_0)I(\theta_0, G_0, \mathcal{P}). \tag{4.3}$$

This means that $\hat{K}_n$ consistently estimates K_0. In view of (4.1), (3.11), (3.3), and (3.5) with $R_{\mathcal{P}} = 0$ we obtain

$$\begin{aligned}
&\sqrt{n}\left(\hat{\nu}_n - \nu_0 - \frac{1}{n}\sum_{i=1}^n \tilde{\ell}(X_i; \nu_0, G_0, \mathcal{Q})\right) \\
&\quad = \sqrt{n}\left(\bar{\nu}_n - \nu_0 + \hat{K}_n\left[\hat{\theta}_n - f\left(\bar{\nu}_n\right)\right] - \frac{1}{n}\sum_{i=1}^n K_0\tilde{\ell}(X_i; \theta_0, G_0, \mathcal{P})\right) \\
&\quad = \sqrt{n}\left(\bar{\nu}_n - \nu_0 - \hat{K}_n\left[f\left(\bar{\nu}_n\right) - f(\nu_0)\right]\right) \\
&\qquad\qquad + \left[\hat{K}_n - K_0\right]\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\ell}(X_i; \theta_0, G_0, \mathcal{P}) + o_p(1).
\end{aligned} \tag{4.4}$$

By the consistency of $\hat{K}_n$ the second term at the right hand side of (4.4) converges to 0 in probability under P_0 in view of the central limit theorem. Because $f\left(\bar{\nu}_n\right) = f(\nu_0) + \dot{f}(\nu_0)\left(\bar{\nu}_n - \nu_0\right) + o_p\left(\bar{\nu}_n - \nu_0\right)$ holds and $K_0\dot{f}(\nu_0)$ equals the $d\times d$ identity matrix, the first part of the right hand side of (4.4) also converges to 0 in probability under P_0. $\square$

To prove Theorem 3.1 with the help of Theorem 4.1 we will construct a $\sqrt{n}$-consistent estimator $\bar{\nu}_n$ of ν and subsequently a consistent estimator $\hat{I}_n$ of $I(\theta, G, \mathcal{P})$. Let $\|\cdot\|$ be a Euclidean norm on $\mathbb{R}^k$. We choose $\bar{\nu}_n$ in such a way that

$$\| f\left(\bar{\nu}_n\right) - \hat{\theta}_n \| \leq \inf_{\nu\in N} \| f(\nu) - \hat{\theta}_n \| + \frac{1}{n} \tag{4.5}$$

holds. Of course, if the infimum is attained, we choose $\bar{\nu}_n$ as the minimizer. There are many numerical optimization techniques that will yield a $\bar{\nu}_n$ satisfying (4.5). By the triangle inequality and the $\sqrt{n}$-consistency of $\hat{\theta}_n$ we obtain

$$\begin{aligned} \| f(\bar{\nu}_n) - f(\nu_0) \| &\leq \inf_{\nu \in N} \| f(\nu) - \hat{\theta}_n \| + \frac{1}{n} + \| f(\nu_0) - \hat{\theta}_n \| \\ &\leq 2 \| \hat{\theta}_n - f(\nu_0) \| + \frac{1}{n} = O_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{4.6}$$

The assumption from Theorem 3.1 that $f(\cdot)$ has an inverse on $f(N)$ that is differentiable with a bounded Jacobian, suffices to conclude that (4.6) guarantees $\sqrt{n}$-consistency of $\bar{\nu}_n$.

In regular parametric models without nuisance parameters any consistent estimator of θ yields a consistent estimator of the Fisher information matrix $I(\theta, \mathcal{P}) = I(\theta, G, \mathcal{P})$ by substitution. Typically, in regular semiparametric models the construction of an efficient estimator is accompanied by an estimator of the efficient influence function, which can be simply transformed into a consistent estimator of the Fisher information matrix. Nevertheless, in order to formally complete the proof of Theorem 3.1 we shall construct a consistent estimator of the Fisher information matrix based on the given efficient estimation method $\hat{\theta}_n$ alone, although this estimator will probably have little practical value. In constructing this estimator we split the sample in blocks as follows. Let (k_n), (ℓ_n), and (m_n) be sequences of integers such that $k_n = \ell_n m_n, k_n/n \to \kappa, 0 < \kappa < 1$, and $\ell_n \to \infty, m_n \to \infty$ hold as $n \to \infty$. Such sequences of integers exist. For $j = 1, \dots, \ell_n$ let $\hat{\theta}_{n,j}$ be the efficient estimator of θ based on the observations $X_{(j-1)m_n+1}, \dots, X_{jm_n}$ and $\hat{\theta}_{n,0}$ be the efficient estimator of θ based on the remaining observations $X_{k_n+1}, \dots, X_n$. Consider the "empirical" characteristic function

$$\hat{\phi}_n(t) = \frac{1}{\ell_n} \sum_{j=1}^{\ell_n} \exp\left\{ it\sqrt{m_n}\left(\hat{\theta}_{n,j} - \hat{\theta}_{n,0}\right)\right\}, \ t \in \mathbb{R}^k, \tag{4.7}$$

which we rewrite as

$$\begin{aligned} \hat{\phi}_n(t) &= \exp\left\{-it\sqrt{m_n}\left(\hat{\theta}_{n,0} - \theta_0\right)\right\} \frac{1}{\ell_n} \sum_{j=1}^{\ell_n} \exp\left\{ it\sqrt{m_n}\left(\hat{\theta}_{n,j} - \theta_0\right)\right\} \\ &= \exp\left\{-it\sqrt{m_n}\left(\hat{\theta}_{n,0} - \theta_0\right)\right\} \tilde{\phi}_n(t). \end{aligned} \tag{4.8}$$

In view of $m_n/(n - k_n) \to 0$ and (3.5) with $R_{\mathcal{P}} = 0$ a.s. we see that the first factor at the right hand side of (4.8) converges to 1 as $n \to \infty$. The efficiency of $\hat{\theta}_n$ in (3.5) with $R_{\mathcal{P}} = 0$ a.s. also implies

$$\begin{aligned} E\left(\tilde{\phi}_n(t)\right) &= E\left(\exp\left\{ it\sqrt{m_n}\left(\hat{\theta}_{n,1} - \theta_0\right)\right\}\right) \\ &\to E\left(\exp\left\{ itZ_{\mathcal{P}}\right\}\right) \end{aligned} \tag{4.9}$$

as $n \to \infty$, with $Z_{\mathcal{P}}$ normally distributed with mean 0 and covariance matrix $I^{-1}(\theta_0, G_0, \mathcal{P})$. Some computation shows

$$\begin{aligned} &E\left(\left|\tilde{\phi}_n(t) - E\left(\tilde{\phi}_n(t)\right)\right|^2\right) \\ &\qquad = \frac{1}{\ell_n}\left(1 - \left|E\left(\exp\left\{it\sqrt{m_n}\left(\hat{\theta}_{n,1} - \theta_0\right)\right\}\right)\right|^2\right) \le \frac{1}{\ell_n}. \end{aligned} \tag{4.10}$$

It follows by Chebyshev's inequality that $\tilde{\phi}_n(t)$ and hence $\hat{\phi}_n(t)$ converges under $P_0 = P_{\theta_0, G_0}$ to the characteristic function of $Z_{\mathcal{P}}$ at t,

$$\hat{\phi}_n(t) \to_{P_0} E\left(\exp\{itZ_{\mathcal{P}}\}\right) = \exp\left\{-\tfrac{1}{2}t^T I^{-1}(\theta_0, G_0, \mathcal{P})t\right\}. \tag{4.11}$$

For every $t \in \mathbb{R}^k$ we obtain

$$-2\log\left(\Re\left(\hat{\phi}_n(t)\right)\right) \to_{P_0} t^T I^{-1}(\theta_0, G_0, \mathcal{P})t. \tag{4.12}$$

Choosing $k(k+1)/2$ appropriate values of t we may obtain from (4.12) an estimator of $I^{-1}(\theta_0, G_0, \mathcal{P})$ and hence of $I(\theta_0, G_0, \mathcal{P})$. Indeed, with t equal to the unit vectors u_i we obtain estimators of the diagonal elements of $I^{-1}(\theta_0, G_0, \mathcal{P})$ and an estimator of its (i, j) element is obtained via

$$\log\left(\Re\left(\hat{\phi}_n(u_i)\right)\right) + \log\left(\Re\left(\hat{\phi}_n(u_j)\right)\right) - \log\left(\Re\left(\hat{\phi}_n(u_i + u_j)\right)\right).$$

When needed, the resulting estimator of $I(\theta_0, G_0, \mathcal{P})$ can be made positive definite by changing appropriate components of it by an asymptotically negligible amount, while the symmetry is maintained.

Under a mild uniform integrability condition it has been shown in [11], that existence of an efficient estimator $\hat{\theta}_n$ of θ in $\mathcal{P}$ implies the existence of a consistent and $\sqrt{n}$-unbiased estimator of the efficient influence function $\tilde{\ell}(\cdot; \theta, G, \mathcal{P})$. Basing this estimator on one half of the sample and taking the average of this estimated efficient influence function at the observations from the other half of the sample, we could have constructed another estimator of the efficient Fisher information. However, this estimator would have been more involved, and, moreover, it needs this extra uniformity condition.

With the help of Theorem 4.1, the estimator $\bar{\nu}_n$ of ν from (4.5), and the construction via (4.12) of an estimator $\hat{I}_n$ of the efficient Fisher information we have completed our construction of an efficient estimator $\hat{\nu}_n$ as in (4.1) of ν. This estimator can be turned into an efficient estimator of $\theta = f(\nu)$ within the model $\mathcal{Q}$ from (1.6) by

$$\tilde{\theta}_n = f(\hat{\nu}_n) \tag{4.13}$$

with efficient influence function

$$\begin{aligned} \tilde{\ell}(\theta_0, G_0, \mathcal{Q}) &= \dot{f}(\nu_0)\tilde{\ell}(\nu_0, G_0, \mathcal{Q}) \\ &= \dot{f}(\nu_0)\left(\dot{f}^T(\nu_0) I(\theta_0, G_0, \mathcal{P})\dot{f}(\nu_0)\right)^{-1}\dot{f}^T(\nu_0)\dot{\ell}(\theta_0, G_0, \mathcal{P}) \end{aligned} \tag{4.14}$$

$$= \dot{f}(\nu_0)\left(\dot{f}^T(\nu_0)I(\theta_0,G_0,\mathcal{P})\dot{f}(\nu_0)\right)^{-1}\dot{f}^T(\nu_0)I(\theta_0,G_0,\mathcal{P})\tilde{\ell}(\theta_0,G_0,\mathcal{P})$$

and asymptotic information bound

$$I^{-1}(\theta_0,G_0,\mathcal{Q}) = \dot{f}(\nu_0)\left(\dot{f}^T(\nu_0)I(\theta_0,G_0,\mathcal{P})\dot{f}(\nu_0)\right)^{-1}\dot{f}^T(\nu_0). \tag{4.15}$$

Indeed, according to Section 2.3 of [2], $\tilde{\theta}_n$ is efficient for estimation of θ under the additional information $\theta = f(\nu)$.

Remark 4.1. If $f(\cdot)$ is a linear function, i.e., $\theta = L\nu + \alpha$ holds with the $k \times d$-matrix L of maximum rank d, then

$$\bar{\nu}_n = (L^T L)^{-1} L^T (\hat{\theta}_n - \alpha) \tag{4.16}$$

attains the infimum at the right hand side of (4.5). So, the estimator (4.1) becomes

$$\hat{\nu}_n = \left(L^T \hat{I}_n L\right)^{-1} L^T \hat{I}_n \left[\hat{\theta}_n - \alpha\right] \tag{4.17}$$

with efficient influence function (3.11) and asymptotic information bound (3.13) with $\dot{f}(\nu_0) = L$, and the estimator from (4.13)

$$\tilde{\theta}_n = L\left(L^T \hat{I}_n L\right)^{-1} L^T \hat{I}_n \left[\hat{\theta}_n - \alpha\right] + \alpha. \tag{4.18}$$

Note that $\tilde{\theta}_n$ is the projection of $\hat{\theta}_n$ on the flat $\{\theta \in \mathbb{R}^k : \theta = L\nu + \alpha, \nu \in \mathbb{R}^d\}$ under the inner product determined by $\hat{I}_n$ (cf. Appendix A) and that the covariance matrix of its limit distribution equals the asymptotic information bound

$$I^{-1}(\theta_0,G_0,\mathcal{Q}) = L\left(L^T I(\theta_0,G_0,\mathcal{P}) L\right)^{-1} L^T. \tag{4.19}$$

Another way to describe this submodel $\mathcal{Q}$ with $\theta = L\nu + \alpha$ is by linear restrictions

$$\mathcal{Q} = \{P_{L\nu+\alpha} : \nu \in N, G \in \mathcal{G}\} = \left\{P_{\theta,G} : R^T\theta = \beta, \theta \in \Theta, G \in \mathcal{G}\right\}, \tag{4.20}$$

where $R^T\alpha = \beta$ holds and the $k \times d$-matrix L and the $k \times (k-d)$-matrix R are matching such that the columns of L are orthogonal to those of R and the $k \times k$-matrix $(L\,R)$ is of rank k. Note that the open subset N of $\mathbb{R}^d$ determines the open subset Θ of $\mathbb{R}^k$ and vice versa. See [4], [18], [14], and [10] for some examples of estimation under linear restrictions.

In terms of the restrictions described by R and β the efficient estimator $\tilde{\theta}_n$ of θ from(4.18) within the submodel $\mathcal{Q}$ can be rewritten as

$$\tilde{\theta}_n = \hat{\theta}_n - \hat{I}_n^{-1} R\left(R^T \hat{I}_n^{-1} R\right)^{-1}\left(R^T \hat{\theta}_n - \beta\right), \tag{4.21}$$

with asymptotic information bound

$$L(L^T I L)^{-1} L^T = I^{-1} - I^{-1}R(R^T I^{-1} R)^{-1} R^T I^{-1},\ I = I(\theta_0,G_0,\mathcal{P}), \tag{4.22}$$

as will be proved in Appendix A.

5. Examples

In this section we present five examples, which illustrate our construction of (semi)parametrically efficient estimators. We shall discuss location-scale, Gaussian copula, and semiparametric regression models, and parametric models under linear restrictions.

Example 5.1. Multivariate normal with common mean

Let $\mathcal{G}$ be the collection of nonsingular $k \times k$-covariance matrices and let the parametric starting model be the collection of nondegenerate normal distributions with mean vector θ and covariance matrix Σ,

$$\mathcal{P} = \left\{ P_{\theta,\Sigma} \,:\, \theta \in \mathbb{R}^k,\ \Sigma \in \mathcal{G} \right\}. \tag{5.1}$$

Efficient estimators of θ and Σ are the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ and the sample covariance matrix $\hat{\Sigma}_n = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)^T$, respectively. Note that $\bar{X}_n$ attains the finite sample Cramér-Rao bound and the asymptotic information bound with $I(\theta, \Sigma, \mathcal{P}) = \Sigma^{-1}$.

The parametric submodel we consider is

$$\mathcal{Q} = \left\{ P_{\mathbf{1}_k \mu, \Sigma} \,:\, \mu \in \mathbb{R},\ \Sigma \in \mathcal{G} \right\}. \tag{5.2}$$

In view of (4.17) and (3.13)

$$\hat{\mu}_n = \left(\mathbf{1}_k^T \hat{\mathbf{\Sigma}}_n^{-1} \mathbf{1}_k \right)^{-1} \mathbf{1}_k^T \hat{\mathbf{\Sigma}}_n^{-1} \bar{\mathbf{X}}_\mathbf{n} \tag{5.3}$$

is an efficient estimator of μ within $\mathcal{Q}$ that attains the asymptotic lower bound $\left(\mathbf{1}_k^T \Sigma^{-1} \mathbf{1}_k \right)^{-1}$. In case the covariance matrix Σ is diagonal with its variances denoted by $\sigma_1^2, \ldots, \sigma_k^2$, we are dealing with the Graybill-Deal model as presented on page 88 of [20]. With $\bar{X}_{i,n} = \frac{1}{n} \sum_{j=1}^n X_{j,i}$, $S_{i,n}^2 = \frac{1}{n} \sum_{j=1}^n (X_{j,i} - \bar{X}_{i,n})^2$, and $\hat{\Sigma}_n = \operatorname{diag}(S_{1,n}^2, \ldots, S_{k,n}^2)$ we obtain the Graybill-Deal estimator

$$\hat{\mu}_n = \frac{\sum_{i=1}^k \bar{X}_{i,n} / S_{i,n}^2}{\sum_{i=1}^k 1 / S_{i,n}^2} \tag{5.4}$$

with asymptotic lower bound $\left(\mathbf{1}_k^T \Sigma^{-1} \mathbf{1}_k \right)^{-1} = 1 / \sum_{i=1}^k 1/\sigma_i^2$.

Example 5.2. Coefficient of variation known

Let $g(\cdot)$ be an absolutely continuous density on $(\mathbb{R}, \mathcal{B})$ with mean 0, variance 1, and derivative $g'(\cdot)$, such that $\int [1 + x^2](g'/g(x))^2 g(x)dx$ is finite. Consider the location-scale family corresponding to $g(\cdot)$. Let there be given efficient estimators $\bar{\mu}_n$ and $\bar{\sigma}_n$ of μ and σ, respectively, based on $X_1, \ldots, X_n$, which are i.i.d. with density $\sigma^{-1} g((\cdot - \mu)/\sigma)$. By I_{ij} we denote the element in the $i$the row and jth column of the matrix $I = \sigma^2 I(\theta, G, \mathcal{P})$, where the Fisher information matrix $I(\theta, G, \mathcal{P})$ is as defined in (3.3) with $\theta = (\mu, \sigma)^T$. Some computation shows $I_{11} = \int (g'/g)^2 g, I_{12} = I_{21} = \int x (g'/g(x))^2 g(x)dx$, and $I_{22} = \int [xg'/g(x) + 1]^2 g(x)dx$ exist and are finite; cf. Section I.2.3 of [6].

We consider the submodel with the coefficient of variation known to be equal to a given constant $c = \sigma/\mu$ and with $\nu = \mu$ the parameter of interest. According to Theorem 4.1 the estimator $\hat{\nu}_n = \hat{\mu}_n$ of μ from (4.1) with $\bar{\nu}_n = \bar{\mu}_n$ and $\hat{\theta}_n = (\bar{\mu}_n, \bar{\sigma}_n)^T$ is efficient and some computation shows

$$\hat{\mu}_n = \left(I_{11} + 2cI_{12} + c^2 I_{22}\right)^{-1} \left[\left(I_{11} + cI_{12}\right)\bar{\mu}_n + \left(I_{12} + cI_{22}\right)\bar{\sigma}_n\right]. \tag{5.5}$$

In case the density $g(\cdot)$ is symmetric around 0, the Fisher information matrix is diagonal and $\hat{\mu}_n$ from (5.5) becomes

$$\hat{\mu}_n = \left(I_{11} + c^2 I_{22}\right)^{-1} \left[I_{11}\bar{\mu}_n + cI_{22}\bar{\sigma}_n\right]. \tag{5.6}$$

In the normal case with $g(\cdot)$ the standard normal density $\hat{\mu}_n$ reduces to

$$\hat{\mu}_n = (1 + c^2)^{-1} \left[\bar{\mu}_n + 2c\bar{\sigma}_n\right] \tag{5.7}$$

with $\bar{\mu}_n$ and $\bar{\sigma}_n$ equal to e.g. the sample mean and the sample standard deviation, respectively; cf. [8], [5], and [9].

Example 5.3. Gaussian copula models

Let

$$\mathbf{X}_1 = (X_{1,1}, \ldots, X_{1,m})^T, \ldots, \mathbf{X}_n = (X_{n,1}, \ldots, X_{n,m})^T$$

be i.i.d. copies of $\mathbf{X} = (X_1, \ldots, X_m)^T$. For $i = 1, \ldots, m$, the marginal distribution function of X_i is continuous and will be denoted by F_i. It is assumed that $(\Phi^{-1}(F_1(X_1)), \ldots, \Phi^{-1}(F_m(X_m)))^T$ has an m-dimensional normal distribution with mean 0 and positive definite correlation matrix $C(\theta)$, where Φ denotes the one-dimensional standard normal distribution function. Here the parameter of interest θ is the vector in $\mathbb{R}^{m(m-1)/2}$ that summarizes all correlation coefficients ρ_{rs}, $1 \le r < s \le m$. We will set this general Gaussian copula model as our semiparametric starting model $\mathcal{P}$, i.e.,

$$\mathcal{P} = \{P_{\theta,G} \ : \ \theta = (\rho_{12}, \ldots, \rho_{(m-1)m})^T \ , G = (F_1(\cdot), \ldots, F_m(\cdot)) \in \mathcal{G}\}. \tag{5.8}$$

The unknown continuous marginal distributions are the nuisance parameters collected as $G \in \mathcal{G}$.

Theorem 3.1 of [13] shows that the normal scores rank correlation coefficient is semiparametrically efficient in $\mathcal{P}$ for the 2-dimensional case with normal marginals with unknown variances constituting a least favorable parametric submodel. As [7] explains at the end of its Section 1 and in its Section 4, its Theorem 4.1 proves that normal marginals with unknown, possibly unequal variances constitute a least favorable parametric submodel, also for the general m-dimensional case. Since the maximum likelihood estimators are efficient for the parameters of a multivariate normal distribution, the sample correlation coefficients are efficient for estimation of the correlation coefficients based on multivariate normal observations. But each sample correlation coefficient and hence its efficient influence function involve only two components of the multivariate normal observations. Apparently, the other components of the multivariate normal observations carry no information about the value of the respective correlation coefficient. Effectively, for each correlation coefficient we are in the 2-dimensional case and invoking again Theorem 3.1 of [13] we see that also in the general m-dimensional case the normal scores rank correlation coefficients are semiparametrically efficient. They are defined as

$$\hat{\rho}_{rs}^{(n)} = \frac{\frac{1}{n}\sum_{j=1}^{n} \Phi^{-1}\left(\frac{n}{n+1}\mathbb{F}_r^{(n)}(X_{j,r})\right)\Phi^{-1}\left(\frac{n}{n+1}\mathbb{F}_s^{(n)}(X_{j,s})\right)}{\frac{1}{n}\sum_{j=1}^{n}\left[\Phi^{-1}\left(\frac{j}{n+1}\right)\right]^2} \tag{5.9}$$

with $\mathbb{F}_r^{(n)}$ and $\mathbb{F}_s^{(n)}$ being the marginal empirical distributions of F_r and F_s, respectively, $1 \leq r < s \leq m$. The Van der Waerden or normal scores rank correlation coefficient $\hat{\rho}_{rs}^{(n)}$ from (5.9) is a semiparametrically efficient estimator of ρ_{rs} with efficient influence function

$$\begin{aligned}\tilde{\ell}_{\rho_{rs}}(X_r, X_s) &= \Phi^{-1}(F_r(X_r))\,\Phi^{-1}(F_s(X_s)) \\ &\quad - \tfrac{1}{2}\rho_{rs}\left\{\left[\Phi^{-1}(F_r(X_r))\right]^2 + \left[\Phi^{-1}(F_s(X_s))\right]^2\right\}.\end{aligned} \tag{5.10}$$

This means that

$$\hat{\theta}_n = (\hat{\rho}_{12}^{(n)}, \ldots, \hat{\rho}_{(m-1)m}^{(n)})^T \tag{5.11}$$

efficiently estimates θ with efficient influence function

$$\tilde{\ell}(\mathbf{X}; \theta, G, \mathcal{P}) = (\tilde{\ell}_{\rho_{12}}(X_1, X_2), \ldots, \tilde{\ell}_{\rho_{(m-1)m}}(X_{m-1}, X_m))^T. \tag{5.12}$$

Subexample 5.3.1. Exchangeable Gaussian copula

The exchangeable m-variate Gaussian copula model

$$\mathcal{Q} = \{P_{\mathbf{1}_k\rho, G} \,:\, \rho \in (-1/(m-1), 1),\ G \in \mathcal{G}\} \subset \mathcal{P} \tag{5.13}$$

is a submodel of the Gaussian copula model $\mathcal{P}$ with a one-dimensional parameter of interest $\nu = \rho$. In this submodel all correlation coefficients have the same value ρ. So, $\theta = \mathbf{1}_k\rho$ with $\mathbf{1}_k$ indicating the vector of ones of dimension $k = m(m-1)/2$. In order to construct an efficient estimator of ρ within $\mathcal{Q}$ along the lines of Section 4, in particular Remark 4.1, we first apply (4.16) with $\alpha = 0$ and $L = \mathbf{1}_k$ to obtain the (natural) $\sqrt{n}$-consistent estimator

$$\bar{\rho}_n = \bar{\nu}_n = \frac{1}{k}\sum_{r=1}^{m-1}\sum_{s=r+1}^{m} \hat{\rho}_{rs}^{(n)}. \tag{5.14}$$

For $\theta = \mathbf{1}_k\rho$ we get by simple but tedious calculations (see Appendix B)

$$E\tilde{\ell}_{\rho_{rs}}\tilde{\ell}_{\rho_{tu}} = \begin{cases} (1-\rho^2)^2 & \text{if} \quad |\{r,s\} \cap \{t,u\}| = 2, \\ \frac{1}{2}(1-\rho)^2\rho(2+3\rho) & \text{if} \quad |\{r,s\} \cap \{t,u\}| = 1, \\ 2(1-\rho)^2\rho^2 & \text{if} \quad |\{r,s\} \cap \{t,u\}| = 0. \end{cases} \tag{5.15}$$

It makes sense to estimate $I(\mathbf{1}_k, G, \mathcal{P})$ by substituting $\bar{\rho}_n$ for ρ in (5.15), to compute the inverse of the resulting matrix, and to choose this matrix as the estimator $\hat{I}_n$. To this end, we note that for every pair $\{r, s\}$, $1 \le r \ne s \le m$, there are $2(m-2)$ pairs of $\{t, u\}$'s having one element in common and there are $\frac{1}{2}(m-2)(m-3)$ pairs of $\{t, u\}$'s having no elements in common. Hence, the sum of the components of each column vector of $I^{-1}(\mathbf{1}_k\rho, G, \mathcal{P})$ is $(1-\rho)^2(1+(m-1)\rho)^2$. Each matrix with the components of each column vector adding to 1 has the property that the sum of all row vectors equals the vector with all components equal to 1, and hence the components of each column vector of its inverse also add up to 1. This implies

$$\mathbf{1}_k^T \hat{I}_n = (1-\bar{\rho}_n)^{-2} \left(1 + (m-1)\bar{\rho}_n\right)^{-2} \mathbf{1}_k^T$$

and hence by (4.17)

$$\hat{\rho}_n = \left(\mathbf{1}_k^T \hat{I}_n \mathbf{1}_k\right)^{-1} \mathbf{1}_k^T \hat{I}_n \hat{\theta}_n = \frac{1}{k} \mathbf{1}_k^T \hat{\theta}_n = \binom{m}{2}^{-1} \sum_{r=1}^{m-1} \sum_{s=r+1}^{m} \hat{\rho}_{rs}^{(n)} = \bar{\rho}_n \tag{5.16}$$

attains the asymptotic information bound (cf. (3.13))

$$\left(\mathbf{1}_k^T I\left(\mathbf{1}_k\rho, G, \mathcal{P}\right) \mathbf{1}_k\right)^{-1} = \binom{m}{2}^{-1} (1-\rho)^2 (1+(m-1)\rho)^2. \tag{5.17}$$

[7] proved the efficiency of the pseudo-likelihood estimator for ρ in dimension $m = 4$. [17] extended this result to general m and presented the efficient lower bounds for $m = 3$ and $m = 4$ in its Example 5.3. However, its maximum pseudo-likelihood estimator is not as explicit as our (5.16).

Subexample 5.3.2. Four-dimensional circular Gaussian copula

A particular, one-dimensional parameter type of four-dimensional circular Gaussian copula model has been studied in [7] and [17]. It is defined by its correlation matrix

$$\begin{pmatrix} 1 & \rho & \rho^2 & \rho \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho & \rho^2 & \rho & 1 \end{pmatrix}. \tag{5.18}$$

Our semiparametric starting model $\mathcal{P}$ is the same as in (5.8) with $m = 4$, but with the components of θ rearranged as follows

$$\theta = (\rho_{12}\ ,\ \rho_{14}\ ,\ \rho_{23}\ ,\ \rho_{34}\ ,\ \rho_{13}\ ,\ \rho_{24})^T.$$

Now, with $f(\rho) = (\rho\ ,\ \rho\ ,\ \rho\ ,\ \rho\ ,\ \rho^2\ ,\ \rho^2)^T$ the present circular Gaussian submodel $\mathcal{Q}$ may be written as

$$\mathcal{Q} = \{P_{f(\rho),G}\ :\ \rho \in (-\tfrac{1}{3}, 1)\,,\ G \in \mathcal{G}\}.$$

In order to construct an efficient estimator of ρ within $\mathcal{Q}$ along the lines of Theorem 4.1, we propose as a $\sqrt{n}$-consistent estimator of ρ

$$\bar{\rho}_n = \tfrac{2}{3}\bar{\rho}_{n,1} + \tfrac{1}{3}\,\mathrm{sign}\,(\bar{\rho}_{n,1})\,\bar{\rho}_{n,2},$$

$$\bar{\rho}_{n,1} = \tfrac{1}{4}\left(\hat{\rho}_{12}^{(n)} + \hat{\rho}_{14}^{(n)} + \hat{\rho}_{23}^{(n)} + \hat{\rho}_{34}^{(n)}\right),\ \bar{\rho}_{n,2} = \tfrac{1}{2}\left(\sqrt{\hat{\rho}_{13}^{(n)}} + \sqrt{\hat{\rho}_{24}^{(n)}}\right) \tag{5.19}$$

As in (5.15) we get by simple but tedious calculations (see Appendix B)

$$I^{-1}(f(\rho), G, \mathcal{P}) = \tfrac{1}{2}\left(1-\rho^2\right)^2 \tag{5.20}$$

$$\begin{pmatrix} 2 & \rho^2 & \rho^2 & 2\rho^2 & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) \\ \rho^2 & 2 & 2\rho^2 & \rho^2 & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) \\ \rho^2 & 2\rho^2 & 2 & \rho^2 & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) \\ 2\rho^2 & \rho^2 & \rho^2 & 2 & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) \\ \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & 2\left(1+\rho^2\right)^2 & 4\rho^2 \\ \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & 4\rho^2 & 2\left(1+\rho^2\right)^2 \end{pmatrix},$$

which has inverse

$$I(f(\rho), G, \mathcal{P}) = \tfrac{1}{2}\left(1-\rho^2\right)^{-4} \tag{5.21}$$

$$\begin{pmatrix} \rho^4+2 & 3\rho^2 & 3\rho^2 & \rho^4+2\rho^2 & -\left(\rho^3+2\rho\right) & -\left(\rho^3+2\rho\right) \\ 3\rho^2 & \rho^4+2 & \rho^4+2\rho^2 & 3\rho^2 & -\left(\rho^3+2\rho\right) & -\left(\rho^3+2\rho\right) \\ 3\rho^2 & \rho^4+2\rho^2 & \rho^4+2 & 3\rho^2 & -\left(\rho^3+2\rho\right) & -\left(\rho^3+2\rho\right) \\ \rho^4+2\rho^2 & 3\rho^2 & 3\rho^2 & \rho^4+2 & -\left(\rho^3+2\rho\right) & -\left(\rho^3+2\rho\right) \\ -\left(\rho^3+2\rho\right) & -\left(\rho^3+2\rho\right) & -\left(\rho^3+2\rho\right) & -\left(\rho^3+2\rho\right) & 2\frac{\rho^6+\rho^4+1}{\rho^4+1} & 2\frac{\rho^6+2\rho^2}{\rho^4+1} \\ -\left(\rho^3+2\rho\right) & -\left(\rho^3+2\rho\right) & -\left(\rho^3+2\rho\right) & -\left(\rho^3+2\rho\right) & 2\frac{\rho^6+2\rho^2}{\rho^4+1} & 2\frac{\rho^6+\rho^4+1}{\rho^4+1} \end{pmatrix}.$$

Substituting $\bar{\rho}_n$ into (5.21) we obtain a $\sqrt{n}$-consistent estimator of $I(f(\rho), G, \mathcal{P})$. In view of $\dot{f}(\rho) = (1,1,1,1,2\rho,2\rho)^T$ we have

$$\dot{f}^T(\rho) I(f(\rho), G, \mathcal{P}) = \left(1-\rho^2\right)^{-3}\left(1+\rho^2, 1+\rho^2, 1+\rho^2, 1+\rho^2, -2\rho, -2\rho\right).$$

Consequently the asymptotic lower bound for estimation of ρ within $\mathcal{Q}$ equals

$$\left[\dot{f}(\rho)^T I(f(\rho), G, \mathcal{P})\dot{f}(\rho)\right]^{-1} = \tfrac{1}{4}\left(1-\rho^2\right)^2. \tag{5.22}$$

Substituting $\bar{\rho}_n$ for ρ we obtain as the efficient estimator from Theorem 4.1

$$\hat{\rho}_n = \bar{\rho}_n + \frac{1+\bar{\rho}_n^2}{1-\bar{\rho}_n^2}\left(\bar{\rho}_{n,1} - \bar{\rho}_n\right) - \frac{\bar{\rho}_n}{1-\bar{\rho}_n^2}\left(\tfrac{1}{2}\left(\hat{\rho}_{13}^{(n)} + \hat{\rho}_{24}^{(n)}\right) - \bar{\rho}_n^2\right). \tag{5.23}$$

[7] has shown that the pseudo maximum likelihood estimator is not efficient in this case. [17] has established the asymptotic lower bound (5.22) and has constructed an alternative, efficient, one-step updating estimator suggesting the pseudo maximum likelihood estimator as the preliminary estimator.

Example 5.4. Partial spline linear regression

Here the observations are realizations of i.i.d. copies of the random vector $X = (Y, Z^T, U^T)^T$ with Y, Z, and U 1-dimensional, k-dimensional, and p-dimensional random vectors with the structure

$$Y = \theta^T Z + \psi(U) + \varepsilon, \tag{5.24}$$

where the measurement error ε is independent of Z and U, has mean 0, finite variance, and finite Fisher information for location, and where $\psi(\cdot)$ is a real valued function on $\mathbb{R}^p$. [16] calls this partially linear additive regression, [2] mentions it as partial spline regression, whereas [3] is talking about the partial smoothing spline model. Under the regularity conditions of its Theorem 8.1 [16] presents an efficient estimator of θ and a consistent estimator of $I(\theta, G, \mathcal{P})$. Consequently our Theorem 4.1 may be applied directly in order to obtain an efficient estimator of ν in appropriate submodels with $\theta = f(\nu)$ without our construction of an estimator of $I(\theta, G, \mathcal{P})$ via characteristic functions. Note that for submodels with θ restricted to a linear subspace, $\theta = L\nu$ say, our approach is not needed, since the reparametrization $Y = \nu^T L^T Z + \psi(U) + \varepsilon$ brings the estimation problem back to its original (5.24).

Example 5.5. Restricted maximum likelihood estimator

As mentioned in the Introduction, most papers on estimation in constrained *parametric* models focus on constrained (or restricted) maximum likelihood estimation implemented via Lagrange multipliers; cf. [1]. Maximum likelihood estimation of the generalized linear model under linear restrictions on the parameters is done in [14] via an iterative procedure using a penalty function. [10] introduces the restricted EM algorithm for maximum likelihood estimation under linear restrictions. Our approach as described in Remark 4.1 with $\hat{\theta}_n$ a(n unrestricted) maximum likelihood estimator avoids such iterative procedures.

Appendix A: Proof of bound subject to linear restriction

In this appendix proofs will be presented of (4.21) and (4.22).

Since $\hat{I}_n$ has been chosen to be symmetric and positive definite, $x^T \hat{I}_n y$, $x, y \in \mathbb{R}^k$, is an inner product on $\mathbb{R}^k$. Define the $k \times k$-matrices $\Pi_{n,L}$ and $\Pi_{n,R}$ by

$$\begin{aligned} \Pi_{n,L} &= L \left(L^T \hat{I}_n L \right)^{-1} L^T \hat{I}_n, \\ \Pi_{n,R} &= \hat{I}_n^{-1} R \left(R^T \hat{I}_n^{-1} R \right)^{-1} R^T. \end{aligned} \tag{A.1}$$

With the above inner product these matrices are projection matrices on the linear subspaces spanned by the columns of L and $\hat{I}_n^{-1} R$, respectively. Indeed, $\Pi_{n,L}\Pi_{n,L} = \Pi_{n,L}$, $\Pi_{n,R}\Pi_{n,R} = \Pi_{n,R}$, $(x - \Pi_{n,L}x)^T \hat{I}_n \Pi_{n,L} x = 0$, $x \in \mathbb{R}^k$, $(y - \Pi_{n,R} y)^T \hat{I}_n \Pi_{n,R} y = 0$, $y \in \mathbb{R}^k$, $\Pi_{n,L} L x = Lx$, $x \in \mathbb{R}^d$, and $\Pi_{n,R} \hat{I}_n^{-1} R y = \hat{I}_n^{-1} R y$, $y \in \mathbb{R}^{k-d}$ hold. The linear subspaces spanned by the columns of L and $\hat{I}_n^{-1} R$ have dimensions d and $k - d$, respectively, since the matrices (L, R) and $\hat{I}_n$ are nonsingular. Moreover, these linear subspaces are orthogonal in view of $L^T \hat{I}_n \hat{I}_n^{-1} R = L^T R = 0$. This implies

$$\Pi_{n,L} x + \Pi_{n,R} x = x, \quad x \in \mathbb{R}^k. \tag{A.2}$$

Combining (A.1), (A.2), and (4.18) we obtain (4.21) and, by the consistency of $\hat{I}_n$, (4.22).

Appendix B: Computation of bound in Example 5.3

We will present the computational details for (5.15) and (5.20). Since our computations will be based on fourth moments of multivariate normal random variables, we consider

$$Z = \begin{pmatrix} Z_a \\ Z_b \\ Z_c \\ Z_d \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{ab} & \rho_{ac} & \rho_{ad} \\ \rho_{ba} & 1 & \rho_{bc} & \rho_{bd} \\ \rho_{ca} & \rho_{cb} & 1 & \rho_{cd} \\ \rho_{da} & \rho_{db} & \rho_{dc} & 1 \end{pmatrix} \right).$$

The following fourth moments of Z can be obtained by straightforward computations:

- $E(Z_a^4) = 3$
- $E(Z_a^3 Z_b) = 3\rho_{ab}$
- $E(Z_a^2 Z_b^2) = 1 + 2\rho_{ab}^2$
- $E(Z_a^2 Z_b Z_c) = \rho_{bc} + 2\rho_{ab}\rho_{ac}$
- $E(Z_a Z_b Z_c Z_d) = \rho_{ab}\rho_{cd} + \rho_{ac}\rho_{bd} + \rho_{ad}\rho_{bc}$.

For every $i, j = 1, \ldots, \binom{n}{2}$ let M_{ij} be the element in the i-th row and j-th column of the efficient lower bound $I^{-1}(\theta, G, \mathcal{P})$. Because of $\theta_i = \rho_{ab}$, $\theta_j = \rho_{cd}$ for some a, b, c, and d, we have

$$M_{ij} = E\left(Z_a Z_b - \tfrac{1}{2}\rho_{ab}\left[Z_a^2 + Z_b^2\right]\right)\left(Z_c Z_d - \tfrac{1}{2}\rho_{cd}\left[Z_c^2 + Z_d^2\right]\right).$$

We have three cases:

- $|\{a, b\} \cap \{c, d\}| = 2$

$$\begin{aligned} M_{ii} &= E\left(Z_a Z_b - \tfrac{1}{2}\rho_{ab}\left[Z_a^2 + Z_b^2\right]\right)^2 \\ &= E\left(Z_a^2 Z_b^2\right) - \rho_{ab} E\left(Z_a^3 Z_b + Z_b^3 Z_a\right) + \tfrac{1}{4}\rho_{ab}^2 E\left(Z_a^4 + 2Z_a^2 Z_b^2 + Z_b^4\right) \\ &= \left(1 + 2\rho_{ab}^2\right) - \rho_{ab}\left(3\rho_{ab} + 3\rho_{ab}\right) + \tfrac{1}{4}\rho_{ab}^2\left(3 + 2\left[1 + 2\rho_{ab}^2\right] + 3\right) \\ &= \left(1 - \rho_{ab}^2\right)^2 \end{aligned}$$

- $|\{a, b\} \cap \{c, d\}| = 1$ (without loss of generality assume $d = a$)

$$\begin{aligned} M_{ij} &= E\left(Z_a Z_b - \tfrac{1}{2}\rho_{ab}\left[Z_a^2 + Z_b^2\right]\right)\left(Z_a Z_c - \tfrac{1}{2}\rho_{ac}\left[Z_a^2 + Z_c^2\right]\right) \\ &= E\left(Z_a^2 Z_b Z_c\right) - \tfrac{1}{2}\rho_{ab} E\left(Z_a^3 Z_c + Z_b^2 Z_a Z_c\right) \\ &\quad -\tfrac{1}{2}\rho_{ac} E\left(Z_a^3 Z_b + Z_c^2 Z_a Z_b\right) \\ &\quad +\tfrac{1}{4}\rho_{ab}\rho_{ac} E\left(Z_a^4 + Z_a^2 Z_b^2 + Z_a^2 Z_c^2 + Z_b^2 Z_c^2\right) \\ &= \left(\rho_{bc} + 2\rho_{ab}\rho_{ac}\right) - \tfrac{1}{2}\rho_{ab}\left(3\rho_{ac} + \left[\rho_{ac} + 2\rho_{ab}\rho_{bc}\right]\right) \\ &\quad -\tfrac{1}{2}\rho_{ac}\left(3\rho_{ab} + \left[\rho_{ab} + 2\rho_{ac}\rho_{bc}\right]\right) \\ &\quad +\tfrac{1}{4}\rho_{ab}\rho_{ac}\left(3 + \left[1 + 2\rho_{ab}^2\right] + \left[1 + 2\rho_{ac}^2\right] + \left[1 + 2\rho_{bc}^2\right]\right) \\ &= \tfrac{1}{2}\left(1 - \rho_{ab}^2 - \rho_{ac}^2\right)\left(2\rho_{bc} - \rho_{ab}\rho_{ac}\right) + \tfrac{1}{2}\rho_{ab}\rho_{ac}\rho_{bc}^2 \end{aligned}$$

- $|\{a,b\} \cap \{c,d\}| = 0$

$$
\begin{aligned}
M_{ij} &= E\left(Z_aZ_b - \tfrac{1}{2}\rho_{ab}\left[Z_a^2 + Z_b^2\right]\right)\left(Z_cZ_d - \tfrac{1}{2}\rho_{cd}\left[Z_c^2 + Z_d^2\right]\right) \\
&= E\left(Z_aZ_bZ_cZ_d\right) - \tfrac{1}{2}\rho_{ab}E\left(Z_a^2Z_cZ_d + Z_b^2Z_cZ_d\right) \\
&\quad -\tfrac{1}{2}\rho_{cd}E\left(Z_c^2Z_aZ_b + Z_d^2Z_aZ_b\right) \\
&\quad +\tfrac{1}{4}\rho_{ab}\rho_{cd}E\left(Z_a^2Z_c^2 + Z_b^2Z_c^2 + Z_a^2Z_d^2 + Z_b^2Z_d^2\right) \\
&= \rho_{ab}\rho_{cd} + \rho_{ac}\rho_{bd} + \rho_{ad}\rho_{bc} - \tfrac{1}{2}\rho_{ab}\left(\left[\rho_{cd} + 2\rho_{ac}\rho_{ad}\right] + \left[\rho_{cd} + 2\rho_{bc}\rho_{bd}\right]\right) \\
&\quad -\tfrac{1}{2}\rho_{cd}\left(\left[\rho_{ab} + 2\rho_{ac}\rho_{bc}\right] + \left[\rho_{ab} + 2\rho_{ad}\rho_{bd}\right]\right) \\
&\quad +\tfrac{1}{4}\rho_{ab}\rho_{cd}\left(\left[1 + 2\rho_{ac}^2\right] + \left[1 + 2\rho_{bc}^2\right] + \left[1 + 2\rho_{ad}^2\right] + \left[1 + 2\rho_{bd}^2\right]\right) \\
&= \rho_{ac}\rho_{bd} + \rho_{ad}\rho_{bc} - \left(\rho_{ab}\rho_{ac}\rho_{ad} + \rho_{ab}\rho_{bc}\rho_{bd} + \rho_{ac}\rho_{bc}\rho_{cd} + \rho_{ad}\rho_{bd}\rho_{cd}\right) \\
&\quad +\tfrac{1}{2}\rho_{ab}\rho_{cd}\left(\rho_{ac}^2 + \rho_{bc}^2 + \rho_{ad}^2 + \rho_{bd}^2\right)
\end{aligned}
$$

Finally, substitution of the correlation structures in Subexample 5.3.1 and Subexample 5.3.2 gives (5.15) and (5.20), respectively.

Acknowledgements

We would like to thank Raymond Veldhuis for suggesting a statistical problem in forensic face recognition, which has led to this study, and we would like to thank him, Constance van Eeden, and Jon Wellner for useful discussions.

References

[1] Aitchison, J. and Silvey, S.D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Ann Math Stat* **29**, 813–828. MR0094873

[2] Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. John Hopkins University Press, Baltimore. (1998) revised paperbound edition. Springer, New York. MR1245941

[3] Cheng, G., Zhang, H.H., and Shang, Z. (2015). Sparse and efficient estimation for partial spline models with increasing dimension. *Ann Inst Stat Math* **67**, 93–127. MR3297860

[4] Cobb, C. W. and Douglas, P. H. (1928). A theory of production. *The American Economic Review* **18**, 139–165.

[5] Gleser L.J. and Healy J.D. (1976).Estimating the mean of a normal distribution with known coefficient of variation. *J Am Stat Assoc* **71**, 977–981. MR0438563

[6] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. Academic Press, New York. MR0229351

[7] Hoff, P. D., Niu, X. and Wellner, J. A. (2014). Information bounds for Gaussian copulas. *Bernoulli* **20**, 604–622. MR3178511

[8] Khan, R.A. (1968). A note on estimating the mean of a normal distribution with known coefficient of variation. *J Am Stat Assoc.* **63**, 1039–1041.

[9] Khan, R.A. (2015). A remark on estimating the mean of a normal distribution with known coefficient of variation. *Statistics* **49**, 705–710. MR3349087
[10] Kim, D. K. and Taylor, J. M. (1995). The restricted EM algorithm for maximum likelihood estimation under linear restrictions on the parameters. *Journal of the American Statistical Association* **90**, 708–716. MR1340522
[11] Klaassen, C.A.J. (1987). Consistent estimation of the influence function of locally asymptotically linear estimates. *Annals of Statistics* **15**, 1548–1562. MR0913573
[12] Klaassen, C.A.J. and Susyanto, N. (2016). Semiparametrically Efficient Estimation of Euclidean Parameters under Equality Constraints. *arXiv:1606.07749*.
[13] Klaassen, C. A. J. and Wellner, J. A. (1997). Effcient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli* **3**, 55–77. MR1466545
[14] Nyquist, H. (1991). Restricted Estimation of Generalized Linear Models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **40**, 133–141.
[15] Rothenberg, T.J. (1973). Efficient Estimation with a Priori Information. *Cowles Foundation Monograph* **23**, Yale University Press, New Haven, CT. MR0326904
[16] Schick, A. (1993). On efficient estimation in regression models. *Annals of Statistics* **21**, 1486-1521. MR1241276
[17] Segers, J., van den Akker, R. and Werker, B. J. M. (2014). Semiparametric Gaussian copula models: geometry and efficient rank-based estimation. *The Annals of Statistics* **42**, 1911–1940. MR3262472
[18] Stone, R. et al. (1954). *The Measurement of Consumers' Expenditure and Behavior in the United Kingdom, 1920-1938*. Cambridge University Press, Cambridge. MR0060206
[19] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge. MR1652247
[20] van Eeden, C. (2006). Restricted-Parameter-Space Estimation Problems; Admissibility and Minimaxity Properties. *Lecture Notes in Statistics* **188**, Springer, New York. MR2265239