Fisher Information in Group-Type Models

Peter Ruckdeschel

*Fraunhofer ITWM, Abt. Finanzmathematik, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
and TU Kaiserslautern, AG Statistik, FB. Mathematik, P.O.Box 3049, 67653 Kaiserslautern, Germany*

Abstract

In proofs of L_2-differentiability, Lebesgue densities of a central distribution are often assumed right from the beginning. Generalizing Huber (1981, Theorem 4.2), we show that in the class of smooth parametric group models these densities are in fact consequences of a finite Fisher information of the model, provided a suitable representation of the latter is used. The proof uses the notions of absolute continuity in k dimensions and weak differentiability.

As examples to which this theorem applies, we spell out a number of models including a correlation model and the general multivariate location and scale model.

As a consequence of this approach, we show that in the (multivariate) location scale model, finiteness of Fisher information as defined here is in fact equivalent to L_2-differentiability and to a log-likelihood expansion giving local asymptotic normality of the model.

Paralleling Huber's proofs for existence and uniqueness of a minimizer of Fisher information to our situation, we get existence of a minimizer in any weakly closed set $\mathscr{F}$ of central distributions F. If, additionally to analogue assumptions to those of Huber (1981), a certain identifiability condition for the transformation holds, we obtain uniqueness of the minimizer. This identifiability condition is satisfied in the multivariate location scale model.

Keywords: Fisher information, group models, multivariate location and scale model, correlation estimation, minimum Fisher information, absolute continuity, weak differentiability, LAN, L2 differentiability, smoothness;
2000 MSC: 62H12, 62F12, 62F35

1. Introduction

1.1. Motivation

 L_2-differentiability as introduced by LeCam and Hájek appears to be the most suitable setup in which to derive such key properties as *local asymptotic normality* (LAN) in local asymptotic parametric statistics. In order to show this L_2-differentiability however, Lebesgue densities of a central distribution are frequently assumed right from the beginning. In this paper, we generalize Huber (1981, Theorem 4.2) from one-dimensional location to a large class of parametric models, where these Lebesgue densities are in fact a consequence of a finite Fisher information of the

model, provided a suitable definition of the latter is used. This definition may then serve—again as in Huber (1981)—as starting point for minimizing Fisher information along suitable neighborhoods of the model.

The framework in which this generalization holds covers smooth parametric group models as to be found in Bickel et al. (1998), but is valid even in a somewhat more general setting: The idea is to link transformations in the parameter space to transformations in the observation space.

The new definition of Fisher information then simply amounts to transferring differentiation in the parameter space to differentiation—in a weak sense—in the observation space. This is actually done much in a Sobolev spirit, working with generalized derivatives.

1.2. Organization of the Paper

After an introduction to the setup of smooth parametric group models, in section 2, we list the smoothness requirements for the transformations and some notation needed for our theorem. Before stating this theorem, in section 3 we first give a number of examples to which this theorem applies, the most general of which is the multivariate location and scale model from Example 3.7. Section 4 provides the main result, Theorem 4.4. In section 5, we spell out the resulting Fisher information in the examples of section 3. As announced in the motivation, in section 6, culminating in Proposition 6.2, we show that in the (multivariate) location-scale model finiteness of Fisher information is equivalent to L_2-differentiability as well as to a LAN property. Finally, in section 7 we generalize Huber's proofs for existence and uniqueness of a minimizer of Fisher information to our situation. The proofs are gathered in appendix section Appendix B; The proof of Theorem 4.4 makes use of the notions of absolute continuity in k dimensions and of weak differentiability. Both are provided in an appendix in section Appendix A.

Remark 1.1. The one-dimensional scale model, a particular case of what is covered by this paper, has been spelt out separately, in a small joint paper with Helmut Rieder, cf. Ruckdeschel and Rieder (2010).

2. Setup

2.1. Notation

$\mathbb{B}^k$ denotes the Borel σ-algebra on $\mathbb{R}^k$, $\mathscr{M}_1(\mathscr{A})$ $[\mathscr{M}_s(\mathscr{A})]$ the set of all probability [substochastic] measures on some σ-algebra $\mathscr{A}$, and for $\mu \in \mathscr{M}_1(\mathbb{B})$, for $p \in [1, \infty]$, $L_p(\mu)$ is the set of all (equivalence classes of) $\mathscr{A}|\mathbb{B}$ measurable functions with $\mathrm{E}|X|^p < \infty$, resp. $\sup_P |X| < \infty$. $\mathbb{I}_A$ denotes the indicator function of the set A. $\mathbb{I}_k$ is the k-dimensional unit matrix, $\mathrm{vec}(A)$ is the operator casting a matrix to a vector, stacking the columns of A over each other, vech the operator casting the upper half of a quadratic matrix to a vector—including the diagonal—and $A \otimes B$ the Kronecker product of matrices, and, for $A, B \in \mathbb{R}^{k \times k}$, the symmetrized product $A \underset{s}{\otimes} B := (AB + B^\tau A^\tau)/2$.

For $l \in \mathbb{N}_0 \cup \infty$ let $\mathscr{C}^l$ be the set of all l times continuously differentiable functions, where—if necessary—we specify domain and range in the notation $\mathscr{C}^l(\text{domain}, \text{range})$. Weak convergence of measures $P_n \in \mathscr{M}_1(\bar{\mathbb{B}}^k)$ to some measure $P \in \mathscr{M}_1(\bar{\mathbb{B}}^k)$ is denoted by $P_n \xrightarrow{\text{w}} P$. Inequalities and intervals in $\mathbb{R}^k$ are denoted by the same symbols as in one dimension, meaning e.g. $l < r$ iff $l_i < r_i$, for all $i = 1, \dots, k$, and $[l, r] := \{x \in \mathbb{R}^k \mid l_i \leq x_i \leq r_i, \quad \forall i = 1, \dots, k\}$.

Let $P_\theta \in \mathscr{M}_1(\mathbb{B}^k)$. $\mathbb{R}^k$ being Polish, regular conditional distributions are available, and we may write $P_\theta(dx_1, \dots, dx_k)$ as

$$P_\theta(dx_1, \dots, dx_k) = \prod_{j=1}^{k-1} P_{\theta; j|j+1:k}(dx_j | x_{j+1}, \dots, x_k) \, P_{\theta; k}(dx_k) \tag{2.1}$$

with $P_{\theta;k}$ the marginal of X_k and $P_{\theta;j|j+1:k}$ a regular conditional distribution of X_j, given $X_{j+1} = x_{j+1}, \ldots, X_k = x_k$. In the sequel, we write $y_{i:j}$ for the vector $(y_i, \ldots, y_j)^\tau$. For a measure G on $\mathscr{M}(\mathbb{B}^k)$ and a set of indices J we write G_J to denote the joint marginal of G for coordinates $i \in J$. For $y \in \mathbb{R}^k$ define $y_{-i} := y_{1:i-1;i+1:k}$, and for $y \in \mathbb{R}^{k-1}$ and $x \in \mathbb{R}$ define the expression $(x:y)_i := (y_{1:i-1}, x, y_{i:k-1})^\tau \in \mathbb{R}^k$.

2.2. Model Definition

For a fixed central distribution F on $\mathbb{B}^k$, we consider a statistical model $\mathscr{P} \subset \mathscr{M}_1(\mathbb{B}^k)$ generated by a family $\mathscr{G}$ of diffeomorphisms $\tau : \mathbb{R}^k \to \mathbb{R}^k$ defined on the observation space. Denote the inverse of τ by $\iota = \tau^{-1}$. This family is parametrized by a p dimensional parameter θ, stemming from an open parameter set $\Theta \subset \mathbb{R}^p$, and this induces the parametric model

$$\mathscr{P} = \{P_\theta \,|\, P_\theta = \tau_\theta(F), \quad \theta \in \Theta\} \tag{2.2}$$

where $\tau_\theta(F)$ denotes the image measure under τ_θ, $F \circ \iota_\theta$.

Remark 2.1. In most examples, $\mathscr{G}$ will be a group, which is also the formulation used in Lehmann (1983, section 1.3) and Bickel et al. (1998, Ch. 4). These authors did not intend to generalize Fisher information, though, and Example 3.5 shows that for our purposes a group structure of for the set $\mathscr{G}$ is not necessary.

2.3. A Smooth Compactification of $\mathbb{R}^k$

For reasons explained in Remark 4.1, we introduce the following compactification $\bar{\mathbb{R}}^k$ of $\mathbb{R}^k$:

Definition 2.2. *Let* $\mathscr{C}^l([0,1]^k, \mathbb{R})$, $l \in \mathbb{N} \cup \infty$ *the space of all continuous real-valued functions on the domain* $[0,1]^k$ *which are differentiable* l *times / arbitrarily often in* $(0,1)^k$, *and with existing one-sided derivatives on* $\partial[0,1]^k$. *We identify this space with functions on* $\bar{\mathbb{R}}^k$, *using the isometry* ℓ

$$\ell : [-\infty; \infty]^k \to [0,1]^k, \quad [\ell((x_j))]_i = \left[e^{x_i}/(e^{x_i}+1)\right]_i \tag{2.3}$$

i.e. let

$$\mathscr{C}^l(\bar{\mathbb{R}}^k, \mathbb{R}) := \mathscr{C}^l([0,1]^k, \mathbb{R}) \circ \ell = \{\varphi \,|\, \varphi = \psi \circ \ell \quad \exists \psi \in \mathscr{C}^l([0,1]^k, \mathbb{R})\} \tag{2.4}$$

For later purposes we also note the inverse of ℓ

$$\kappa : [0,1]^k \to [-\infty; \infty], \quad \kappa(y_1, \ldots, y_k) = \left(\log(y_j/(1-y_j))\right)_{j=1,\ldots,k} \tag{2.5}$$

In the same manor, unbounded, continuous functions are defined and denoted by $\mathscr{C}^l(\bar{\mathbb{R}}^k, \bar{\mathbb{R}}^m)$.

Remark 2.3. (a)   With this definition, $\bar{\mathbb{R}}^k$ becomes a compact metric space.
 (b)   Integrations along $\bar{\mathbb{R}}^k$ are understood as lifted onto $[0,1]^k$ by ℓ, i.e. $\int_{\bar{\mathbb{R}}^k} f \, dP = \int_{[0,1]^k} f \circ \kappa \, d[\ell \circ P]$.
 (c)   The choice of ℓ resp. κ is arbitrary to some extent, but satisfactory for our needs; in fact, we only have to impose $\ell \in \mathscr{C}^\infty(\mathbb{R}^k, \mathbb{R}^k)$, $\lim_{x \to -\infty}(\ell((x:y)_i))_i = 0$, $\lim_{x \to \infty}(\ell((x:y)_i))_i = 1$, for each $y \in \mathbb{R}^{k-1}$, ℓ strictly isotone in each coordinate, $|\ell'(x)D(\tau_\theta(x))| \in L_2(F)$, or, for uniformity in $\mathscr{M}_1(\mathbb{B}^k)$, $\sup_x |\ell'(x)D(\tau_\theta(x))| < \infty$.
 (d)   For every $\varphi \in \mathscr{C}^\infty(\bar{\mathbb{R}}, \mathbb{R})$, the limits $\lim_{x \to \pm\infty} \varphi(x)$ exist and $\lim_{x \to \pm\infty} \frac{d^l}{dx^l} \varphi(x) = 0$ for $l \geq 0$, as is easily seen using the chain rule and by the fact that each summand arising in a derivative has at least a factor decaying as $\exp(-|x|)$. This also implies that there are functions $\tilde{\varphi} : \bar{R} \to \mathbb{R}$ which do not lie in $\mathscr{C}^\infty(\bar{\mathbb{R}}, \mathbb{R})$ but which are in $\mathscr{C}^\infty(\mathbb{R}, \mathbb{R})$, have existing $\lim_{x \to \pm\infty} \tilde{\varphi}(x)$, and for which $\lim_{x \to \pm\infty} \frac{d^k}{dx^k} \tilde{\varphi}(x) = 0$ for $k \geq 0$: Take $1/(x^2+1)$, which has no exponentially decaying derivatives.

3

(e)   Consequently, for all $\varphi \in \mathscr{C}^\infty(\bar{\mathbb{R}}^k, \mathbb{R})$, $\int |b\varphi'| d\lambda$ is finite for any bounded, measurable function b and $\lim_{|x|\to\infty} \varphi(x)'|x|^k = 0$ for all $k \in \mathbb{N}$, hence in particular is in $L_\infty(P)$ for every probability P on $\mathbb{B}$.

(f)   If we allow for mass of $\ell \circ P$ in $[0,1]^k \setminus (0,1)^k$ —corresponding to measures in $\mathscr{M}_s(\mathbb{B}^k)$—the class $\mathscr{C}_c^k(\mathbb{R}, \mathbb{R})$ of compactly supported functions in $\mathscr{C}^k(\mathbb{R}, \mathbb{R})$ cannot distinguish any measures $P_1 \neq P_2$ on $\bar{\mathbb{B}}^k$ coinciding on $\mathbb{B}^k$, whereas $\mathscr{C}^\infty(\mathbb{R}^k, \mathbb{R})$ is measure determining on $\bar{\mathbb{B}}^k$.

(g)   The measures P_θ arising in our model from subsection 2.2 are understood as members of $\mathscr{M}_1(\bar{\mathbb{B}}^k)$, defining $P_\theta(A) = P_\theta(A \cap \mathbb{R}^k)$ for $A \in \bar{\mathbb{B}}^k$.

2.4. Assumptions

Throughout this paper, we make the following set of assumptions concerning the transformations τ, which are needed to link differentiation w.r.t. θ to differentiation w.r.t. x:

(I) $P_{\theta_1} = P_{\theta_2}$ $\qquad \Longleftrightarrow \qquad \theta_1 = \theta_2$.

(D) $\theta \mapsto \iota_\theta(x)$ is differentiable with derivative $\partial_\theta \iota_\theta(x)$.

(Dk) If $k > 1$, $x \mapsto \iota_\theta(x)$ is twice differentiable with second derivative $\partial_{xx}^2 \iota_\theta(x)$.

(C1) If $k = 1$, $x \mapsto D$ is in $\mathscr{C}^1(\mathbb{R}, \bar{\mathbb{R}}^p)$ and $x \mapsto e^{-|x|} D \circ \tau_\theta(x)$ is in $L_2^p(F)$ with

$$D = D_\theta^{(1)}(x) = \partial_\theta \iota_\theta(x) / \partial_x \iota_\theta(x) \tag{2.6}$$

(Ck) If $k > 1$, $x \mapsto D$ is in $\mathscr{C}^1(\bar{\mathbb{R}}^k, \bar{\mathbb{R}}^{k \times p})$, $x \mapsto e^{-|x|} D \circ \tau_\theta(x)$ is in $L_2^{k \times p}(F)$ and $x \mapsto V \circ \tau_\theta(x)$ is in $L_2^p(F)$ with

$$J = (J_\theta(x))_{i,j=1,\ldots,k} = ((\partial_x \iota_\theta)^{-1})_{i,j}(x) \tag{2.7}$$

$$D = (D_\theta^{(k)}(x))_{\substack{i=1\ldots k \\ j=1\ldots p}} = [J^\tau \partial_\theta \iota_\theta]_{i,j}(x) \tag{2.8}$$

$$V = (V_\theta(x))_{j=1\ldots p} = \frac{[\sum_{i=1}^k \partial_{x_i}(|\det \partial_x \iota_\theta| D_{i,j}) - \partial_{\theta_j}|\det \partial_x \iota_\theta|]_j}{|\det \partial_x \iota_\theta|}(x) \tag{2.9}$$

Remark 2.4. Using $\frac{\partial}{\partial A_{i,j}} \det A = (A^{-1})_{j,i} \det A$ and $\frac{\partial}{\partial A_{k,l}}(A^{-1})_{i,j} - (A^{-1})_{i,k}(A^{-1})_{l,j}$ and the chain rule of differentiation, one can show

$$\begin{aligned} V_j &= [|\det \partial_x \iota_\theta|]^{-1} \left[\sum_{i=1}^k \{\partial_{x_i}(|\det \partial_x \iota_\theta| D_{i,j})\} - \partial_{\theta_j}|\det \partial_x \iota_\theta| \right] = \\ &= \sum_{i,l,r,m=1}^k (J_{l,i}J_{r,m} - J_{l,r}J_{m,i}) \partial_{x_i x_r}^2 \iota_{\theta;m} \partial_{\theta_j} \iota_{\theta;l}, \end{aligned} \tag{2.10}$$

which motivates requirement (Dk).

In the sequel we use these abbreviations:

Notation 2.5. The set $\{D = 0\}$ is denoted by K. With e_i the i-th canonical unit vector in $\mathbb{R}^k$ and some $a \in \mathbb{R}^p$ and $y \in \mathbb{R}^{k-1}$, define

$$V_a := V^\tau a, \qquad D_a := Da, \qquad D_{a;i} := e_i^\tau D_a, \qquad K_i := \{e_i^\tau D = 0\} \tag{2.11}$$

Also, for later purposes—c.f. (4.4)—we introduce the functions

$$\tilde{D} = (\tilde{D}_\theta^{(k)}(x))_{\substack{i=1\ldots k \\ j=1\ldots p}} = [\partial_\theta \iota_\theta \circ \tau_\theta]_{i,j}(x), \quad \tilde{V} = (\tilde{V}_\theta(x))_{j=1\ldots p} = V_\theta \circ \tau_\theta \tag{2.12}$$

$$\tilde{V}_a := \tilde{V}^\tau a, \qquad \tilde{D}_a := \tilde{D}a, \qquad \tilde{D}_{a;i} := e_i^\tau \tilde{D}a, \qquad \tilde{K}_i := \{e_i^\tau \tilde{D} = 0\} \tag{2.13}$$

Finally, if $F \ll \lambda^k$, we write f_θ for $f \circ \iota_\theta$, with f a λ^k density of F.

4

We also introduce the following decomposition of P_θ:

$$P_\theta := P_\theta^{(0)} + \bar{P}_\theta^{(0)}, \qquad \bar{P}_\theta^{(0)}(\cdot) := P_\theta(\cdot \cap K). \tag{2.14}$$

3. Examples

For the following seven popular examples we spell out the transformations $\tau_\theta(x)$ and the respective parameter space and verify the assumptions from the preceding section.

Example 3.1 (one-dim. location). $\tau_\theta(x) := x + \theta$, $\theta \in \Theta_1 = \mathbb{R}$, $p = k = 1$.
For each $\theta \in \Theta_1$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D) and (C1) are satisfied—$\partial_\theta l_\theta = -1$, $\partial_x l_\theta = 1$, $D(x) = -1$, $K = \emptyset$—any observation x is informative for this problem.

Example 3.2 (k-dim. location, $k > 1$). $\tau_\theta(x) := x + \theta$, $\theta \in \Theta_2 = \mathbb{R}^k$ $p = k$.
For each $\theta \in \Theta_2$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D), (Dk), and (Ck) are satisfied— $\partial_{xx}^2 l_\theta = 0$, $\partial_\theta l_\theta = -\partial_x l_\theta = D = -\mathbb{I}_k$, $V = 0$, $K = \emptyset$—any observation x carries information for this problem.

Example 3.3 (one-dim. scale). $\tau_\theta(x) := \theta x$, $\theta \in \Theta_3 = \mathbb{R}_{>0}$, $p = k = 1$.
For each $\theta \in \Theta_3$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D) and (C1) are satisfied—$\partial_\theta l_\theta = -x/\theta^2$, $\partial_x l_\theta = 1/\theta$, $D(x) = -x/\theta$. Thus $K = \{0\}$, hence the point $x = 0$ is not informative for this problem, and any $x \neq 0$ is.

Example 3.4 (one-dim. loc. and scale). $\tau_\theta(x) := \theta_2 x + \theta_1$, $\theta \in \Theta_4 = \Theta_1 \times \Theta_3$, $k = 1$, $p = 2$.
For each $\theta \in \Theta_4$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D) and (C1) are satisfied— consider $\partial_\theta l_\theta = -(\frac{1}{\theta_2}; \frac{x - \theta_1}{\theta_2^2})^\tau$, $\partial_x l_\theta = \frac{1}{\theta_2}$, $D(x) = -(1; \frac{x - \theta_1}{\theta_2})$, $K = \emptyset$—any observation x carries information for this problem.

Example 3.5 (correlation, $k = 2$; $p = 1$). To $\sigma_1, \sigma_2 > 0$ known let $\theta \in \Theta_5 = (-1; 1)$

$$\tau_\theta : \mathbb{R}^2 \to \mathbb{R}^2, \quad x \mapsto \tau_\theta(x) := J_\theta x, \qquad J_\theta = \begin{pmatrix} \sigma_1(1 - \theta^2)^{\frac{1}{2}} & \theta\sigma_1 \\ 0 & \sigma_2 \end{pmatrix} x; \tag{3.1}$$

In contrast to all other examples considered here, this family does not form a group; this may easily be seen, as J_θ^{-1} does not admit a representation according to (3.1). For each $\theta \in \Theta_5$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D), (Dk), and (Ck) are clearly satisfied— $\partial_{xx}^2 l_\theta = 0$, $V = 0$, $\partial_\theta l_\theta = (1 - \theta^2)^{-\frac{3}{2}}(x_1 \theta \sigma_1^{-1} - x_2 \sigma_2^{-1}, 0)^\tau$,

$$\partial_x l_\theta = (1 - \theta^2)^{-\frac{1}{2}} \begin{pmatrix} \sigma_1^{-1} & -\theta\sigma_2^{-1} \\ 0 & (1 - \theta^2)/\sigma_2 \end{pmatrix},$$

$D = ([\theta x_1 - \frac{\sigma_1}{\sigma_2} x_2]/(1 - \theta^2), 0)^\tau$. As $K = \{x \in \mathbb{R}^2 \mid \exists \rho \in \mathbb{R} : x = \rho(\sigma_1; \theta\sigma_2)^\tau\}$, $P(K) < 1$ holds, as long as $\mathrm{supp}(P_\theta)$ is not contained in the line $\{\rho(1; \theta\sigma_2/\sigma_1)^\tau, \rho \in \mathbb{R}\}$ or equivalently, as long as $\mathrm{supp}(F) \not\subset \{\rho((1 - \theta^2)^{\frac{1}{2}}; \theta)^\tau, \rho \in \mathbb{R}\}$.

Example 3.6 (k-dim. scale, $k > 1$). $\tau_\theta(x) := \theta x$, defined for $\Theta_6 = \{S \in \mathbb{R}^{k \times k} \mid S = S^\tau \succ 0\}$, $p = \binom{k+1}{2}$. The symmetry restriction is imposed on $\mathbb{R}^{k \times k}$, allowing only for symmetric variations in the parameter.

5

Again, for each $\theta \in \Theta_6$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D), (Dk), and (Ck) are satisfied—$\partial_{xx}^2 l_\theta = 0$, $\partial_{x_i} l_{\theta;l} = (\theta^{-1})_{l,i}$,

$$\partial_{\theta_{i_1 i_2}} l_{\theta;l} = -\frac{1}{2}[(\theta^{-1})_{l,i_1}(\theta^{-1}x)_{i_2} + (\theta^{-1})_{l,i_2}(\theta^{-1}x)_{i_1}],$$

$$D = D_{i;j_1,j_2} = \frac{1}{2}[(\mathbb{I} \otimes \theta^{-1}x)_{i,j_1,j_2} + (\mathbb{I} \otimes \theta^{-1}x)_{i,j_2,j_1}], \qquad V = 0.$$

For each symmetric matrix $a \in \mathrm{GL}(k)$, we have $D(x)a = \theta^{-1}a\theta^{-1}x$; $K = \{0\}$—any observation $x \neq 0$ carries information for this problem.

Example 3.7 (k-dim. location and scale, $k > 1$). $\tau_\theta(x) := \theta_2 x + \theta_1$, for $\theta \in \Theta_7 = \mathbb{R}^k \times \Theta_6$, $p = k + \binom{k+1}{2} = k(k+3)/2$.
For each $\theta \in \Theta_7$, $\tau_\theta(\cdot)$ is a diffeomorphism; assumptions (I), (D) (Dk), and (Ck) are satisfied—$\partial_{xx}^2 l_\theta = 0$, $V = 0$, $\partial_x l_\theta = \theta_2^{-1}$; splitting off the indices for the parametric dimensions into the location part [a single index] and the scale part [a double index], we get

$$\partial_{\theta_i} l_{\theta;l} = -(\theta_2^{-1})_{i;l},$$

$$\partial_{\theta_{i_1 i_2}} l_{\theta;l} = -\frac{1}{2}[(\theta_2^{-1})_{l,i_1}(\theta_2^{-1}(x-\theta_1))_{i_2} + (\theta_2^{-1})_{l,i_2}(\theta_2^{-1}(x-\theta_1))_{i_1}];$$

$$D_{i,l} = -\mathbb{I}_{i;l},$$

$$D_{i_1 i_2,l} = -\frac{1}{2}[(\mathbb{I} \otimes \theta_2^{-1}(x-\theta_1))_{l,i_1 i_2} + (\mathbb{I} \otimes \theta_2^{-1}(x-\theta_1))_{l,i_2 i_1}].$$

Just as in Example 3.2, any observation x carries information for this problem.

4. Main Theorem

In Huber (1981, Definition 4.1 and Theorem 4.2), we find a result on the Fisher information in the one dimensional location case which is central for the famous minimax M estimator result of Huber (1964). The idea is to express Fisher information as a supremum, i.e.

$$\mathscr{I}(F) := \sup\left\{ \frac{(\int \varphi' \, dF)^2}{\int \varphi^2 \, dF} \,\bigg|\, 0 \overset{[F]}{\neq} \varphi \in \mathscr{D}_1 \right\}. \tag{4.1}$$

With this definition, Huber (1981, Thm 4.2) achieves a representation of Fisher information without assuming densities of the central distribution: $\mathscr{I}(F)$ is finite iff F is a.c. with a.c. Lebesgue density f such that $\int (f'/f)^2 f \, dx < \infty$, which in this case is just $\mathscr{I}(F)$.

Remark 4.1. (a) The proof in Huber (1981) is credited to T. Liggett and is based on Sobolev-type ideas; we take these up to generalize the result to more general models and higher dimensions.
(b) The set $\mathscr{D}_1$ in (4.1) plays the rôle of a set of test functions as in the theory generalized functions, compare Rudin (1991, Ch. 6). In the cited reference, Huber uses $\mathscr{D}_1 = \mathscr{C}_c^1(\mathbb{R},\mathbb{R})$, the subset of compactly supported functions in $\mathscr{C}^1(\mathbb{R},\mathbb{R})$. In the proof later, we will need that the sets

$$\mathscr{D}_{D;i,j} := \{D_{a;j}\,\partial_{x_i}\phi \,|\, \phi \in \mathscr{D}_k,\, a \in \mathbb{R}^p\}$$

are dense in $L_2(P_\theta^{(j)})$. Contrary to the one-dimensional location case, for $\mathscr{D}_k = \mathscr{C}_c^1(\mathbb{R}^k,\mathbb{R})$ and general $D_{a;j}$, we did not succeed to prove this; nor can we work with $\mathscr{D}_k = \mathscr{C}_{c1}(\mathbb{R}^k,\mathbb{R})$, the set of continuously

6

differentiable functions with compactly supported derivatives, as used for the one dimensional scale model in Ruckdeschel and Rieder (2010, Lem. A.1): The crucial approximation of the constant function 1 by functions $\phi \in \mathscr{C}_{c1}(\mathbb{R}^k, \mathbb{R})$, with $|\phi| \leq 1$, $|D_{a;j}\,\partial_{x_i}\phi| \leq 1$, and $|D_{a;j}\,\partial_{x_i}\phi| \to 0$ pointwise, fails for functions $D_{a;j}$ growing faster than $|x|$ for large $|x|$. Hence, instead we use the larger set $\mathscr{C}^\infty(\bar{\mathbb{R}}^k, \mathbb{R})$ from Definition 2.2.

Definition 4.2. *In model $\mathscr{P}$ from (2.2), assume assume (I) and (D). Let $a \in \mathbb{R}^p$, $|a| = 1$.*

$k = 1$**:** *Assume (C1). Let $\mathscr{D}_1 = \mathscr{C}(\bar{\mathbb{R}}^1, \mathbb{R})$, D from (2.6). Then for $\theta \in \Theta$ we define*

$$\mathscr{I}_\theta(F;a) := \sup\left\{ \frac{\left(\int[\varphi' D_a]\, dP_\theta\right)^2}{\int \varphi^2\, dP_\theta} \;\middle|\; 0 \overset{[P_\theta]}{\neq} \varphi \in \mathscr{D}_1 \right\}, \tag{4.2}$$

$k > 1$**:** *Assume (Dk) and (Ck). Let $\mathscr{D}_k = \mathscr{C}(\bar{\mathbb{R}}^k, \mathbb{R})$, D and V from (2.8) and (2.9). Then for $\theta \in \Theta$ we define*

$$\mathscr{I}_\theta(F;a) := \sup\left\{ \frac{\left(\int[\nabla\varphi^\tau D_a + \varphi V_a]\, dP_\theta\right)^2}{\int \varphi^2\, dP_\theta} \;\middle|\; 0 \overset{[P_\theta]}{\neq} \varphi \in \mathscr{D}_k \right\}. \tag{4.3}$$

Remark 4.3. (a) As τ_θ, resp. ι_θ map $\mathscr{D}_k$ onto itself, we may use the identification $\psi = \varphi \circ \tau_\theta$ to see that by the transformation formula

$$\mathscr{I}_\theta(F;a) := \sup\left\{ \frac{\left(\int[\nabla\psi^\tau \tilde{D}_a + \psi \circ \iota_\theta \tilde{V}_a]\, dF\right)^2}{\int \psi^2\, dF} \;\middle|\; 0 \overset{[P_\theta]}{\neq} \psi \in \mathscr{D}_k \right\}. \tag{4.4}$$

(b) In particular, the transformation formula $\int \rho(x) P_\theta(dx) = \int \rho \circ \tau_\theta\, dF$, entails that except for the correlation model of Example 3.5, finiteness of the Fisher information for one $\theta \in \Theta$ implies finiteness for every $\theta \in \Theta$: Indeed, considering $D_\theta^{(k)} \circ \tau_\theta$ in all these models, we see that in every case, $D_\theta^{(k)} \circ \tau_\theta = D_{\mathrm{id}}^{(k)}$, where we write id referring to the parameter-value θ yielding $\iota_\theta = \mathrm{id}$, while at the same time $V = 0$. So in fact we could define *the* Fisher information of F for one reference parameter, and its finiteness then entails finiteness in the whole parametric model.

(c) In general, finiteness will however depend on the actual parameter value, which is why we define Fisher information at F with reference to θ, notationally transparent as $\mathscr{I}_\theta(F;a)$.

With Definition 4.2 we generalize Huber (1981, Thm. 4.2) to

Theorem 4.4. *In model $\mathscr{P}$ from (2.2) assume that for some fixed $\theta \in \Theta$, (I), and, if $k = 1$, (D), and (C1), resp., if $k > 1$, (Dk) and (Ck) hold. Then (the sets of) statements (i) and (ii) are equivalent:*

(i) $\sup_{a:|a|=1} \mathscr{I}_\theta(F;a) < \infty$

(ii) (a) *F admits a λ^k density f on $\iota_\theta(K^c)$.*
   (b) *For every $a \in \mathbb{R}^p$, and $i = 1, \ldots, k$*

$$\lim_{|x| \to \infty} [f_\theta \,|\det \partial_x \iota_\theta|\, D_{a;i}]((x:y)_i) = 0$$

(c) *For every $a \in \mathbb{R}^p$, and $i = 1, \ldots, k$ $f_\theta\,|\det \partial_x \iota_\theta|\, D_{a;i}$ is a.c. in k dimensions in the sense of Definition A.3.*

7

(d) *For every $a \in \mathbb{R}^p$ and $1 \le i \le k$, $\left[\frac{\partial_{x_i}(|\det \partial_x \iota_\theta| D_{a;i})}{|\det \partial_x \iota_\theta|} + \frac{D_{a;i}\partial_{x_i}f_\theta}{f_\theta}\right] \in L_2(P_\theta)$.*

If (i) resp. (ii) holds, $\mathscr{I}_\theta(F;a) = a^\tau \mathscr{I}_\theta(F)a$ with

$$\mathscr{I}_\theta = \int \Lambda_\theta \Lambda_\theta^\tau \, dP_\theta, \qquad \Lambda_\theta = (f'/f) \circ \iota_\theta \, \partial_\theta \iota_\theta + \frac{\partial_\theta |\det \partial_x \iota_\theta|}{|\det \partial_x \iota_\theta|}, \tag{4.5}$$

respectively

$$\Lambda_\theta = \partial_\theta p_\theta / p_\theta \qquad \text{with } p_\theta = f_\theta |\det \partial_x \iota_\theta|. \tag{4.6}$$

Remark 4.5. (a) Theorem 4.4 also covers model 3.1; however, it uses $\mathscr{D}_1 = \mathscr{C}^\infty(\bar{\mathbb{R}}, \mathbb{R})$ instead of $\mathscr{C}_c^1(\mathbb{R}, \mathbb{R})$, hence, as $\mathscr{C}_c^1(\mathbb{R}, \mathbb{R}) \subset \mathscr{C}^\infty(\bar{\mathbb{R}}, \mathbb{R})$, finiteness of Fisher information in Huber's definition formally is weaker than ours, so formally our implication (ii) $\Longrightarrow$ (i) is harder, (i) $\Longrightarrow$ (ii) easier than his.

(b) As a consequence of using $\mathscr{D}_1 = \mathscr{C}^\infty(\bar{\mathbb{R}}, \mathbb{R})$, we need (ii)(b), which does not show up in the corresponding Theorems Huber (1981)(one-dim. location).

(c) In Theorem 4.4, F may have λ^k singular parts on $\iota_\theta K$. But if so, then by Corollary B.3 necessarily, $\Lambda_\theta = 0$ there. This means that these parts do not contribute any information.

(d) Closedness of a.c. functions under products (Dudley, R.M., 2002, 7.2 Prob.4) entails that under the assumptions of Theorem 4.4, whenever the map $x \mapsto [D_{a;i}p_\theta]((x:y)_i)$ is a.c. on some interval $[c, d]$ where $D_{a;i} \ne 0$, so is p_θ.

5. Fisher information in Examples

In this section we specify the terms Λ_θ and $\mathscr{I}_\theta(F;a)$, as well as the quadratic form in a, $\mathscr{I}_\theta(F) = \mathscr{I}_\theta$, for Examples 3.1 to 3.7. In the sequel, $\Lambda_f(x) := -\partial_x f/f$

Example 5.1 (one-dim. location). $\Lambda_\theta(x) := \Lambda_f(x - \theta)$, $\mathscr{I}_\theta = \mathscr{I}_0 = \int \Lambda_f^2 dF$. The supremal definition of $\mathscr{I}(F)$ is (4.1), but with $\mathscr{D}_1 = \mathscr{C}^\infty(\bar{\mathbb{R}}, \mathbb{R})$.

Example 5.2 (k-dim. location, $k > 1$). $\Lambda_\theta(x) := \Lambda_f(x - \theta)$, $\mathscr{I}_\theta = \mathscr{I}_0 = \int \Lambda_f \Lambda_f^\tau dF$., $\mathscr{I}_\theta(F;a) = a^\tau \mathscr{I}_0 a$. The supremal definition of $\mathscr{I}(F)$ is

$$\mathscr{I}_0(F;a) := \sup \left\{ \frac{(\int \nabla \varphi^\tau a \, dF)^2}{\int \varphi^2 \, dF} \,\middle|\, 0 \overset{[F]}{\ne} \varphi \in \mathscr{D}_k \right\} \tag{5.1}$$

Example 5.3 (one-dim. scale). $\Lambda_\theta(x) := \frac{1}{\theta}[(x/\theta)\Lambda_f(x/\theta) + 1]$, $\mathscr{I}_\theta = \frac{1}{\theta^2}\mathscr{I}_1 = \frac{1}{\theta^2}\int(x\Lambda_f - 1)^2 dF = \frac{1}{\theta^2}(\int x^2 \Lambda_f^2 dF - 1)$. The supremal definition of $\mathscr{I}(F)$ is

$$\mathscr{I}_1(F) := \sup \left\{ \frac{(\int x\varphi'(x)F(dx))^2}{\int \varphi^2 \, dF} \,\middle|\, 0 \overset{[F]}{\ne} \varphi \in \mathscr{D}_1 \right\} \tag{5.2}$$

Example 5.4 (one-dim. location and scale).

$$\Lambda_\theta(x) := \frac{1}{\theta_2}\left(\Lambda_f(\frac{x - \theta_1}{\theta_2}), \ (\frac{x - \theta_1}{\theta_2})\Lambda_f(\frac{x - \theta_1}{\theta_2}) + 1 \right)^\tau,$$

$$\mathscr{I}_\theta(x) := \frac{1}{\theta_2^2}\mathscr{I}_{0;1}(x) = \frac{1}{\theta_2^2}\left(\begin{array}{cc} \int \Lambda_f^2 dF & \int x\Lambda_f^2 dF \\ \int x\Lambda_f^2 dF & \int(x\Lambda_f - 1)^2 dF \end{array} \right),$$

and $\mathscr{I}_\theta(F;a) = a^\tau \mathscr{I}_0 a/\theta_2$. With $a = (a_l, a_s)^\tau$, the supremal definition of $\mathscr{I}(F)$ is

$$\mathscr{I}_{e_2}(F;a) := \sup \left\{ \frac{(\int(a_l + a_s x)\varphi'(x)F(dx))^2}{\int \varphi^2 \, dF} \,\middle|\, 0 \overset{[F]}{\ne} \varphi \in \mathscr{D}_1 \right\} \tag{5.3}$$

8

Example 5.5 (correlation, $k = 2$; $p = 1$).

$$\sigma_2\sigma_1\sqrt{1-\theta^2}\,P_\theta(dx_1,dx_2) = f\Big(\frac{x_1/\sigma_1-\theta x_2/\sigma_2}{\sqrt{1-\theta^2}},\frac{x_2}{\sigma_2}\Big)\lambda^2(dx_1,dx_2) \qquad (5.4)$$

or with $f = f_{1|2}f_2$

$$\sigma_1\sqrt{1-\theta^2}\,p_{\theta;1|2}(x_1,x_2) = f_{1|2}\Big(\frac{x_1/\sigma_1-\theta x_2/\sigma_2}{\sqrt{1-\theta^2}}\Big),\qquad \sigma_2 p_{\theta;2}(x_1,x_2) = f_2\Big(\frac{x_2}{\sigma_2}\Big) \qquad (5.5)$$

and

$$(1-\theta^2)\Lambda_\theta(x_1,x_2) = \frac{f'_{1|2}}{f_{1|2}}\Big(\frac{x_1/\sigma_1-\theta x_2/\sigma_2}{\sqrt{1-\theta^2}}\Big)\frac{\theta x_1/\sigma_1-x_2/\sigma_2}{\sqrt{1-\theta^2}}+\theta,\quad \mathscr{I}_\theta = \int \Lambda_\theta^2\,dP_\theta \qquad (5.6)$$

The supremal definition of $\mathscr{I}(F)$ is

$$\mathscr{I}_\theta(F) := \sup\left\{\frac{\Big(\int[\theta x_1 - \sqrt{1-\theta^2}x_2](\partial_{x_1}\varphi)(x_1,x_2)\,F(dx_1,dx_2)\Big)^2}{(1-\theta^2)^2\int\varphi^2\,dF}\;\Big|\;\;0\overset{[F]}{\neq}\varphi\in\mathscr{D}_2\right\} \qquad (5.7)$$

Example 5.6 (k-dim. scale, $k > 1$). We give both vech expressions and matrix expressions, using symmetrized Kronecker products. We start with unsymmetrized versions.

$$\begin{aligned}
\Lambda_\theta^0(x) &= \theta^{-1}\Lambda_{\mathbb{I}_k}(\theta^{-1}x)), & \Lambda_{\mathbb{I}_k}^0(x) &= \Lambda_f(x)x^\tau - \mathbb{I}_k,\\
\Lambda_\theta(x) &= \frac{1}{2}[\Lambda_\theta^0(x)+\Lambda_\theta^0(x)^\tau], & \Lambda_\theta^v(x) &= \text{vech}[\Lambda_\theta(x)],
\end{aligned}$$

This can also be written as $\Lambda_\theta^v(x) = \text{vech}[\theta^{-1}\underset{s}{\otimes}\Lambda_{\mathbb{I}_k}(x)]$. In matrix notation this yields

$$\mathscr{I}_\theta = ((\theta^{-1}\otimes\theta^{-1})\underset{s}{\otimes}[\int(\Lambda_f X^\tau - \mathbb{I}_k)^{\otimes 2}\,dF],$$

in vector notation $\mathscr{I}_\theta = \int\Lambda_\theta^v(\Lambda_\theta^v)^\tau\,dF$. Working with $a = a^\tau \in \mathbb{R}^{k\times k}$, we get

$$\text{vech}(a)^\tau\Lambda_\theta^v(x) = \Lambda_f(\theta^{-1}x)^\tau\theta^{-1}a\theta^{-1}x - \text{tr}(\theta^{-1}a),$$

$$\mathscr{I}_\theta(F,a) = \int(\Lambda_f(y)^\tau\theta^{-1}ay - \text{tr}(\theta^{-1}a))^2\,F(dy)$$

For symmetric a, the supremal definition of $\mathscr{I}(F)$ is

$$\mathscr{I}_\theta(F,a) := \left\{\sup\frac{(\int\nabla\varphi(x)^\tau\theta^{-1}ax\,F(dx))^2}{\int\varphi^2\,dF}\;\Big|\;\;0\overset{[F]}{\neq}\varphi\in\mathscr{D}_k\right\} \qquad (5.8)$$

Example 5.7 (k-dim. location and scale, $k > 1$). Partitioning Λ into a location block (l) and a scale block (s), we get

$$\begin{aligned}
\Lambda_{l,\theta_1,\theta_2}(x) &= \theta_2^{-1}\Lambda_{l,0,\mathbb{I}_k}(\theta_2^{-1}(x-\theta_1)), & \Lambda_{l,0,\mathbb{I}_k}(x) &= \Lambda_f(x)\\
\Lambda_{s,\theta_1,\theta_2}(x) &= \theta_2^{-1}\Lambda_{s,0,\mathbb{I}_k}(\theta_2^{-1}(x-\theta_1)), & \Lambda_{s,0,\mathbb{I}_k}^0(x) &= \Lambda_f(x)x^\tau - \mathbb{I}_k\\
\Lambda_{s,0,\mathbb{I}_k}(x) &= (\Lambda_{s,0,\mathbb{I}_k}^0(x)+\Lambda_{s,0,\mathbb{I}_k}^0(x)^\tau)/2, & \Lambda_{s,0,\mathbb{I}_k}^v(x) &= \text{vech}(\Lambda_{s,0,\mathbb{I}_k}(x))
\end{aligned}$$

$$\mathscr{I}_\theta = \begin{pmatrix} \mathscr{I}_{1,1,\theta} & \mathscr{I}_{1,s,\theta} \\ \mathscr{I}_{1,s,\theta}^\tau & \mathscr{I}_{s,s,\theta} \end{pmatrix}$$

with

$$\mathscr{I}_{1,1,\theta} = \theta_2^{-1}\big[\int \Lambda_f \Lambda_f^\tau \, dF\big]\theta_2^{-1},$$

$$\mathscr{I}_{1,s,\theta} = \theta_2^{-1}\big[\int \Lambda_f \operatorname{vech}[\theta_2^{-1} \underset{s}{\otimes} (\Lambda_f X^\tau - \mathbb{I}_k)]^\tau \, dF\big]$$

$$\mathscr{I}_{s,s,\theta} = \int \operatorname{vech}[\theta_2^{-1} \underset{s}{\otimes} (\Lambda_f X^\tau - \mathbb{I}_k)] \operatorname{vech}[\theta_2^{-1} \underset{s}{\otimes} (\Lambda_f X^\tau - \mathbb{I}_k)]^\tau \, dF$$

Working with $a = (a_l^\tau; \operatorname{vech}(a_s)^\tau)^\tau$, $a_l \in \mathbb{R}^k$ $a_s = a_s^\tau \in \mathbb{R}^{k \times k}$, we get

$$\operatorname{vech}(a)^\tau \Lambda_\theta^v(x) = a_l^\tau \theta_2^{-1} \Lambda_f(\theta_2^{-1}(x-\theta_1)) + \Lambda_f(\theta_2^{-1}(x-\theta_1))^\tau \theta_2^{-1} a_s \theta_2^{-1}(x-\theta_1) - \operatorname{tr}(\theta_2^{-1}a_s),$$

$$\mathscr{I}_\theta(F,a) = \int (a_l^\tau \theta_2^{-1} \Lambda_f(y) + \Lambda_f(y)^\tau \theta_2^{-1} a_s y - \operatorname{tr}(\theta_2^{-1}a_s))^2 \, F(dy)$$

The supremal definition of $\mathscr{I}(F)$ is

$$\mathscr{I}_\theta(F,a) := \sup\left\{ \frac{\left(\int \nabla \varphi(x)^\tau \theta_2^{-1}[a_s x + a_l] F(dx)\right)^2}{\int \varphi^2 \, dF} \;\middle|\; 0 \overset{[F]}{\neq} \varphi \in \mathscr{D}_k \right\} \tag{5.9}$$

To keep the order of the examples as in section 3, we place a remark here, concerning Example 3.5

Remark 5.8. The fact that we are dealing with a one dimensional parameter seems to indicate that it should be possible to treat the problem using only one dimensional densities. Factorizations (5.5) and (5.6) seem to point into the same direction, as they seem to suggest that working with

$$\sigma_1 \sqrt{1-\theta^2} P_\theta(dx_1, dx_2) = f_{1|2}\Big(\frac{x_1/\sigma_1 - \theta x_2/\sigma_2}{\sqrt{1-\theta^2}}, \frac{x_2}{\sigma_2}\Big) \lambda(dx_1) F_2(\sigma_2^{-1}dx_2) \tag{5.10}$$

instead of (5.4), we could allow for any second marginal F_2—possibly even $F_2 \perp \lambda$—and just focus on the conditional densities for each fixed x_2 section.

Theorem 4.4, however, excludes that possibility for finite Fisher information. To be fair, one has to admit that anyway, not every F with $P_\theta = \tau_\theta F$ could be allowed for (5.10), but only exactly those achieving this representation. But even then it is of rather marginal interest, as may be seen in the following example: Consider $Y_1 \sim \mathscr{N}(0,1)$, $Y_2 \sim \pm 1$ with $P(Y_2 = 1) = P(Y_2 = -1) = 1/2$, Y_1, Y_2 independent and $F := \mathscr{L}(Y_1, Y_2)$. Then for any $\theta \in \Theta_5$, $X = \tau_\theta(Y) = ((1-\theta^2)^{\frac{1}{2}}Y_1 + \theta Y_2, Y_2)$, and recovering θ from observations of X amounts to estimating $E[X_1|X_2 = x_2]$ for $x_2 = \pm 1$—a task falling into the usual $O_P(n^{-\frac{1}{2}})$-type of statistical decision problems; if on the other hand, we take $F = \mathscr{L}(Y_1, (1-\alpha^2)^{\frac{1}{2}}Y_1 + \alpha Y_2,)$, for any $0 < |\alpha| < 1$, then, for $\theta \neq -(2-\alpha^2)^{-\frac{1}{2}}$, $\mathscr{L}(X)$ is concentrated on two lines $X_2 = a_i + \beta X_1$, $i = 1,2$ with $\beta = (1-\alpha^2)^{\frac{1}{2}}/[(1-\theta^2)^{\frac{1}{2}} + \theta(1-\alpha^2)^{\frac{1}{2}}]$. But as we assume F to be known, knowledge of β is just as good as knowledge of θ. Having fixed an observation $X^{(0)}$, β may be recovered exactly, as soon as we have found two further observations $X^{(1)}$ and $X^{(2)}$ both lying on the same line as $X^{(0)}$, which will happen in finite time almost surely. Thus here a single observation must have infinite information on θ—which is just according to our theorem.

6. Consequences for the LAN Approach

In general finiteness of Fisher information does not imply L_2-differentiability without additional assumptions like, e.g. that for λ^k almost all x and for all $\rho \in \mathbb{R}^p$ the map $s \mapsto p_{\theta+s\rho}(x)$ is a.c. and the Fisher information $\mathscr{I}_\theta$ is continuous in θ—c.f. Le Cam (1986, 17.3 Prop.4). All examples from section 3—except for the correlation example, Example 3.5—provide more structure, though. They may all be summarized in the (multivariate) location scale model of Example 3.7. First of all, due to the invariance/dilation relations of Lebesgue measure w.r.t. affine transformations, we may limit attention to the reference parameter $(0, \mathbb{I}_k)$. Even more though, we have the following generalization of Lemmas by Hájek (1972) (one-dimensional location) and Swensen (1980, Ch.2, Sec.3) to the multivariate location case

Proposition 6.1. *Assume that in the multivariate location and scale model 3.7, Fisher information as defined in (5.8) is finite for some parameter value. Then the model is L_2-differentiable for any parameter value.*

Hence Theorem 4.4 gives a sufficient condition for these models to be L_2-differentiable and as a consequence to be LAN.

On the other hand, L_2-differentiability requires finiteness of $\mathscr{I}_\theta$, so that in the multivariate location and scale case, for all central distributions F, the model with central distribution F is L_2 differentiable iff $\sup_a \mathscr{I}_\theta(F; a) < \infty$.
In the i.i.d. setup Le Cam (1986, 17.3 Prop.2) even show that L_2-differentiability is both necessary and sufficient to get an LAN expansion of the likelihoods in form

$$\log dP^n_{\theta+h/\sqrt{n}}/dP^n_\theta = \frac{1}{\sqrt{n}}\sum_{i=1}^n h^\tau \Lambda_\theta(x_i) - \frac{1}{2}h^\tau \mathscr{I}_\theta h + o_{P^n_\theta}(n^0) \tag{6.1}$$

with some $\Lambda_\theta \in L_2(P_\theta)$ and $0 \prec \mathscr{I}_\theta = E[\Lambda_\theta \Lambda_\theta^\tau] \prec \infty$, so again in the setup of the (multivariate) location scale model of Example 3.7 finiteness of Fisher information is both necessary and sufficient to such an LAN expansion. Altogether we have

Proposition 6.2. *In models 3.1, 3.2, 3.3, 3.4, 3.6, 3.7, the following statements are equivalent*

  (i) The respective Fisher information from (4.3) is finite for any parameter value.
 (ii) Conditions (ii) of Theorem 4.4 hold for any parameter value.
(iii) The model is L_2-differentiable for any parameter value.
(iv) The model admits the LAN property (6.1) for any parameter value.

Remark 6.3. The proof uses the translation invariance and the transformation property under dilations of k-dimensional Lebesgue measure, so there is not much room for extensions beyond group models induced by subgroups of the general affine group.

7. Minimization of the Fisher information

Representations (4.2) resp. (4.3) for Fisher information allow for minimization, resp. to maximization of the trace or maxev of $\mathscr{I}_\theta$ w.r.t. the central distribution P_θ or F. In this paper, we settle the questions of (strict) convexity and lower continuity just as in Huber (1981), but replace

vague topology used in Huber (1981) by weak topology. This is done in order to establish existence and uniqueness of a minimizing $F^{(0)}$ in some suitable neighborhood of the (ideal) model. To this end define for $a \in \mathbb{R}^p$, $\varphi \in \mathscr{D}_k$, $\|\varphi \circ \tau_\theta\|_{L_2(F)} \neq 0$

$$\mathscr{I}_\theta(F;a;\varphi) := \frac{\left(\int \nabla \varphi^\tau D_a + \varphi V_a \, d[\tau_\theta F]\right)^2}{\int \varphi^2 \, d[\tau_\theta F]} \tag{7.1}$$

and

$$\bar{\mathscr{I}}_\theta(F) := \sup \mathscr{I}_\theta(F;a), \quad a \in \mathbb{R}^p, \ |a| = 1 \tag{7.2}$$

7.1. Weak Lower Semicontinuity and Convexity

To show weak lower semicontinuity and convexity, we use that for fixed $\varphi \in \mathscr{D}_k$, $\varphi \neq 0 \, [P_\theta]$, $F \mapsto \mathscr{I}_\theta(F;a;\varphi)$ is weak continuous (by definition) and convex (by Huber (1981, Lemma 4.4)). Essentially we may then use that the supremum of continuous functions is lower semicontinous and the supremum of convex functions remains convex; but some subtle additional arguments are needed as the set of φ's over which we are maximizing may vary from F to F; these can be found in Ruckdeschel and Rieder (2010, Proof to Prop. 2.1). Altogether we have shown

Proposition 7.1. *For each $a \in \mathbb{R}^p$, the mapping $F \mapsto \mathscr{I}_\theta(F;a)$ is weakly lower-semicontinuous and convex in $F \in \mathscr{M}_1(\bar{\mathbb{B}}^k)$. The same goes for $F \mapsto \bar{\mathscr{I}}_\theta(F)$.*

Remark 7.2. Using $\bar{\mathbb{R}}^k$ from Definition 2.2, we work with a compact definition space right away, which moreover is endowed with a separable metric, so any subset of probability measures on $\bar{\mathbb{B}}^k$ is tight, hence by Prokhorov's theorem weakly relatively sequentially compact.

Corollary 7.3. *In any weakly closed set $\mathscr{F} \subset \mathscr{M}_1(\bar{\mathbb{B}}^k)$, both $\bar{\mathscr{I}}_\theta$ and $\mathscr{I}_{\theta;a}$—for fixed a—attain their minimum in some $F_0 \in \mathscr{F}$.*

7.2. Strict Convexity—Uniqueness of a Minimizer

We essentially take over the assumptions of Huber (1981); we fix $\theta \in \Theta$ and consider variations in F of the following form: For $F_i \in \mathscr{M}(\bar{\mathbb{B}}^k)$ $i = 0, 1$ consider

$$\tilde{F}_t := (1-t)F_0 \circ \iota_\theta + tF_1 \circ \iota_\theta \tag{7.3}$$

We distinguish cases (a) and $(\bar{\mathscr{I}})$, i.e., of a given one-dimensional projection $a \neq 0$, and the corresponding maximal eigenvalue, respectively.

Proposition 7.4. *Under assumptions*

(a) *The set $\mathscr{F}$ of admitted central distributions F is convex.*

(b) *There is a $F_0 \in \mathscr{P}$ minimizing*

    (a) *$\mathscr{I}_\theta(F;a)$ along $\mathscr{F}$ and $\mathscr{I}_\theta(F_0;a) < \infty$.*

    $(\bar{\mathscr{I}})$ *$\bar{\mathscr{I}}_\theta(F)$ along $\mathscr{F}$ and $\bar{\mathscr{I}}_\theta(F_0) < \infty$.*

(c) *The set where the Lebesgue–density $\tilde{f}_0$ of $\tilde{F}_0$ is strictly positive is convex and contains the support of every $\tilde{F}_t$ derived from some $F_1 \in \mathscr{F}$.*

(d) (a) *$\lambda^k(\{x \mid a^\tau \partial_\theta \iota_\theta(x) = 0\}) = 0$*

    $(\bar{\mathscr{I}})$ *$\lambda^k(\{x \mid \exists a : |a| = 1 \ s.t. \ a^\tau \partial_\theta \iota_\theta(x) = 0\}) = 0$*

12

the map $F \mapsto \mathscr{I}_\theta(F;a)$ (case (a)) resp. $F \mapsto \bar{\mathscr{I}}_\theta(F)$ (case ($\bar{\mathscr{I}}$)) is strictly convex, hence there is a unique minimizer of F_0.

Remark 7.5. Assumption (d) holds for the k dimensional location scale model of Example 3.7: For symmetric a_s and the scale part θ_s it holds that $a_s^\tau \partial_{\theta_s} \iota_\theta = a_s \theta_s^{-1} x$. But for any a_s with $|a_s| = 1$, $\dim \ker a_s \theta_s^{-1} \leq k - 1$, hence $\lambda^k(\ker a_s \theta_s^{-1}) = 0$.

7.3. Existence of a Maximizer of $\operatorname{tr} \mathscr{I}_\theta^{-1}(F)$

Proposition 7.6. *Let $\mathscr{F}$ be a weakly closed subset of $\mathscr{M}_1(\bar{\mathbb{B}}^k)$. Assume that for all $0 \neq a \in \mathbb{R}^p$, $\min_{F \in \mathscr{F}} \mathscr{I}_\theta(F;a) > 0$. Then the function $F \mapsto \operatorname{tr} \mathscr{I}_\theta^{-1}(F)$ is weakly upper-semicontinuous on $\mathscr{F}$, and consequentially, attains its maximum along $\mathscr{F}$ in some $F_0 \in \mathscr{F}$.*

Appendix A. Functional Analysis and Generalized Differentiability

Appendix A.1. Dense Functions

Proposition A.1. *Let μ be a σ-finite measure on $\mathbb{B}^k$. Then the set $\mathscr{C}_c^\infty(\mathbb{R}^k, \mathbb{R})$ is dense in any $L_p(\mu)$, $p \in [1, \infty)$. In particular, there is a $c_0 \in (0, \infty)$ s.t. for any $a < b \in \mathbb{R}$ and any $\delta > 0$ there is a $\varphi = \varphi_{a,b,\delta} \in \mathscr{C}_c^\infty(\mathbb{R}, [0,1])$, with $\varphi \equiv 0$ on $[a - \delta; b + \delta]^c$, $\varphi \equiv 1$ on $[a + \delta; b - \delta]$ and $|\dot{\varphi}| \leq c_0/\delta$.*

Proof : Denseness is a consequence of Lusin's Theorem, compare Rudin (1974, Thm. 3.14). To achieve the universal bound c_0, we may use functions $\hat{f}(t) = \left(\int_0^t \tilde{f}(s) \, ds \right) / \left(\int_0^1 \tilde{f}(u) \, du \right)$, for $f(t) = e^{-1/t}$, $\tilde{f}(t) = f(t) f(1-t)$. $\square$

Appendix A.2. Absolute Continuity

We recall the following characterization of absolute continuity [notation a.c.] of functions $F : \mathbb{R} \to \mathbb{R}$ that can be found in Rudin (1974, Ch. 8).

Theorem A.2. *For $F : [a,b] \to \mathbb{R}$, $a < b \in \mathbb{R}$ the following statements 1. to 3. are equivalent*

1. *F is a.c. on $[a,b]$*
2. (a) *$F'(x)$ exists $\lambda(dx)$ a.e. on $[a,b]$ and $F' \in L_1(\lambda_{|[a,b]})$.*
   (b) *$F(x) - F(a) = \int_a^x F'(s) \lambda(ds)$ for all $x \in [a,b]$.*
3. *There is some $u \in L_1(\lambda_{|(a,b)})$ s.t. for $x \in [a,b]$, $F(x)$ has the representation*

$$F(x) = F(a) + \int_a^x u(s) \, \lambda(ds)$$

We also recall that a.c. functions, are closed under products (Dudley, R.M., 2002, 7.2 Prob.4). In particular, integration by parts is available. In this paper, we call a function $F : \mathbb{R} \to \mathbb{R}$ a.c. if the equivalent statements 1. to 3. from Theorem A.2 are valid for each compact interval $[a,b] \subset \mathbb{R}$.

Appendix A.3. Absolute Continuity in Higher Dimensions

A little care has to be taken about null sets when transferring absolute continuity to higher dimensions. The next definition is drawn from Simader (2001).

Definition A.3. *A function $f : (\mathbb{R}^k, \mathbb{B}^k) \to (\mathbb{R}, \mathbb{B})$ is called* absolutely continuous (in k dimensions)*, if for every $i = 1, \ldots, k$, there is a set $N_i \in \mathbb{B}^{k-1}$ with $\lambda^{k-1}(N_i) = 0$ s.t. for $y \in N_i^c$, the function $f_{i,y} : (\mathbb{R}, \mathbb{B}) \to (\mathbb{R}, \mathbb{B})$, $x \mapsto f_{i,y}(x) = f((x:y)_i)$ is a.c. in the usual sense.*

In the proof of (ii) $\Rightarrow$ (i) in Theorem 4.4, we need the following lemma:

Lemma A.4. *Let* $f : \mathbb{R}^k \to \mathbb{R}$ *a.c. in* k *dimensions. Then for each* $i = 1 \dots, k$

$$\lambda^k(\{f = 0\}, \{\partial_{x_i} f \neq 0\}) = 0 \tag{A.1}$$

Proof : Let $g(z) = \mathrm{I}_{\{f=0\} \cap \{\partial_{x_i} f \neq 0\}}(z)$. Then $g \geq 0$ and Tonelli applies, so the section-wise defined function $h_y(x) := g((x : y)_i)$ is measurable for each $y \in \mathbb{B}^{k-1}$ and defining the possibly infinite integrals $H(y) := \int h_y(x) \lambda(dx)$ we get $\lambda^k(\{f = 0\}, \{\partial_{x_i} f \neq 0\}) = \int g \, d\lambda^k = \int H(y) \lambda^{k-1}(dy)$. But for each y the instances x where $h_y(x) = 0$, $h'_y(x) \neq 0$ are separated by open one-dim. sets where $h_y \neq 0$, as $h_y(x) = 0$, $h'_y(x) \neq 0$ implies that for some $0 < |x' - x| < \varepsilon$, $|f(x')| > |x' - x| |h'_y(x)|/2 > 0$. Hence at most there can be a countable number of such x, and thus $H(y) = 0$ for each y. $\qquad\square$

Appendix A.4. Weak Differentiability

For proving absolute continuity in Theorem 4.4 we have worked with the notion of weak differentiability; to this end we compile the following definitions and propositions again drawn from Simader (2001), which we have specialized to differentiation of order one.

Definition A.5. *Let* $u \in L_{1,\mathrm{loc}}(\lambda^k)$, $1 \leq i \leq k$. *Then* $v_i \in L_{1,\mathrm{loc}}(\lambda^k)$ *is called* weak derivative of u (with respect to x_i), *denoted by* $\tilde{\partial}_{x_i} u$, *if*

$$\int_{\mathbb{R}^k} u \, \partial_{x_i} \varphi \, d\lambda^k = - \int_{\mathbb{R}^k} v_i \varphi \, d\lambda^k \qquad \forall \varphi \in \mathscr{C}_c^\infty(\mathbb{R}^k, \mathbb{R}) \tag{A.2}$$

Remark A.6. (a) The weak derivative is unique, as for the difference $d = v_i - v'_i$ of two potential candidates, we have $\int_{\mathbb{R}^k} d\varphi \, d\lambda^k = 0$ for all $\varphi \in \mathscr{C}_c^\infty(\mathbb{R}^k, \mathbb{R})$, so by Proposition A.1, d must be 0 $[\lambda^k]$.
 (b) Weak derivatives belonging to $L_2(\lambda^k)$ give rise to the space $\mathscr{W}_{2;1} = \mathscr{W}_{2;1}(\lambda^k)$ of all functions $f : \mathbb{R}^k \to \mathbb{R}$ with weak derivatives in $L_2(\lambda^k)$ of order one endowed with the norm $\|f\|_{\mathscr{W}_{2;1}}^2 := \sum_{i=l}^k \|\partial_{x_i} f\|_{L_2(\lambda^k)}^2$ which is called *Sobolev space* of order 2 and 1 for which there is a rich theory.
 (c) The following two propositions—under the additional requirement that ∇f resp. $\tilde{\nabla} f$ be in $L_2^k(\lambda^k)$, however—may also be found in Maz'ya (1985, Thm.'s 1 and 2).

Proposition A.7. *Let* $f \in L_{1,\mathrm{loc}}(\lambda^k)$ *with a weak gradient* $\tilde{\nabla} f$. *Then there is some* $\tilde{f}$, *a.c. in* k *dimensions with usual gradient* $\nabla \tilde{f}$, *such that—up to a* λ^k-*null set—* $\tilde{f} = f$ *and* $\tilde{\nabla} f = \nabla \tilde{f}$.

Proof : Let again $\Omega_m = [-m, m]$ and consider $\chi_m \in \mathscr{D}_k$ with $0 \leq \chi_m \leq 1$, $\varphi_m \equiv 0$ on Ω_{m+1}^c, $\chi_m \equiv 1$ on Ω_m, and let $f_m = f\chi_m$. Then $f_m \in L_1(\lambda^k)$ and we have for any $\phi \in \mathscr{D}_k$

$$-\int f_m \partial_{x_i} \phi \, d\lambda^k = -\int \chi_m \phi \, \tilde{\partial}_{x_i} f \, d\lambda^k = \int f \partial_{x_i}(\chi_m \phi) \, d\lambda^k = \int f(\phi \, \partial_{x_i} \chi_m + \chi_m \partial_{x_i} \phi) \, d\lambda^k$$

so that f_m is weakly differentiable and $\tilde{\partial}_{x_i} f_m = \chi_m \tilde{\partial}_{x_i} f + f \partial_{x_i} \chi_m \in L_1(\lambda^k)$. By Fubini we obtain some $N_{m,i} \in \mathbb{B}^{k-1}$ with $\lambda^{k-1}(N_{m,i}) = 0$ such that $v_m : \mathbb{R}^{k-1} \to \mathbb{R}$ defined as

$$v_m(y) := \int_{\mathbb{R}} |\tilde{\partial}_{x_i} f_m((t : y)_i)| \lambda(dt) \quad \text{for } y \in N_{m,i}^c \quad \text{and} \quad 0 \text{ else}$$

is finite for $y \in \mathbb{R}^{k-1}$, lies in $L_1(\lambda^{k-1})$ and $\int_{\mathbb{R}^{k-1}} v_m \, d\lambda^{k-1} = \|\tilde{\partial}_{x_i} f_m\|_{L_1(\lambda^k)}$. Thus we may define to $x \in \mathbb{R}$

$$F_m((x : y)_i) := \int_{-\infty}^x \tilde{\partial}_{x_i} f_m((t : y)_i) \lambda(dt) \quad \text{for } y \in N_{m,i}^c \quad \text{and} \quad 0 \text{ else} \tag{A.3}$$

Apparently, $F_m \in L_{1,\mathrm{loc}}(\lambda^k)$ and for $y \in N_{m,i}^c$, $x \mapsto F_m((x : y)_i)$ is a.c. Let $\phi \in \mathscr{D}_k$; then Fubini yields

$$I \quad := \quad \int_{\mathbb{R}^k} \phi F_m \, d\lambda^k = \int_{N_{m,i}^c} \int_{\mathbb{R}} \tilde{\partial}_{x_i} f_m((t : y)_i) \int_{\mathbb{R}} \mathrm{I}_{\{x \geq t\}} \phi((x : y)_i) \lambda(dx) \lambda(dt) \lambda^{k-1}(dy)$$

14

So far we do not know if the inner integral on the RHS is in $L_1(\lambda^k)$, so another localization argument is needed. To this end let $\psi \in \mathscr{D}_k$, $\psi \equiv 1$ on Ω_{m+1}, $\psi \equiv 0$ on Ω_{m+2}^c; then as $f_m, \tilde{\partial}_{x_i} f_m \equiv 0$ on Ω_{m+1}^c, we have $f_m \equiv f_m \psi$, $\tilde{\partial}_{x_i} f_m \equiv \psi \tilde{\partial}_{x_i} f_m$, and $f_m \partial_{x_i} \psi \equiv 0$. For with $u = (t{:}y)_i$ define the function

$$\varphi(u) := \psi(u) \int I_{\{x \geq t\}} \phi((x{:}y)_i) \, \lambda(dx),$$

which clearly lies in $\mathscr{D}_k$. Fubini and the definition of weak differentiability entail

$$I \;=\; \int_{N_{m,i}^c} \int_{\mathbb{R}} \tilde{\partial}_{x_i} f_m((t{:}y)_i) \varphi((t{:}y)_i) \, \lambda(dt) \, \lambda^{k-1}(dy) = \int_{\mathbb{R}^k} \varphi \tilde{\partial}_{x_i} f_m \, d\lambda^k = -\int_{\mathbb{R}^k} f_m \partial_{x_i} \varphi \, d\lambda^k$$

But, $\partial_{x_i} \varphi(u) = \partial_{x_i} \psi(u) \int_{\mathbb{R}} I_{\{x \geq t\}}(x) \phi((x{:}y)_i) \, \lambda(dx) - \psi(u)\phi(u)$, as $f_m \partial_{x_i}\psi \equiv 0$, $f_m \equiv f_m \psi$ we get

$$I = \int_{\mathbb{R}^k} \phi F_m \, d\lambda^k = \int_{\mathbb{R}^k} f_m \psi \phi \, d\lambda^k = \int_{\mathbb{R}^k} f_m \phi \, d\lambda^k, \tag{A.4}$$

Because ϕ was arbitrary in $\mathscr{D}_k$, $F_m = f_m \;[\lambda^k]$, and by letting $m \to \infty$ we may extend this to $\mathbb{R}^k$. Fubini then provides a λ^{k-1}-null set S_i s.t. for $y \in S_i^c$ the projection set $S_i^{(y)} := \{x \in \mathbb{R} : (x{:}y)_i \in S_i\}$ has λ-measure 0. Let $N_i := \bigcup_m N_{m,i}$; then $\lambda^{k-1}(N_i) = 0$, and for $y \in N_i^c$ the functions $x \mapsto F_m((x{:}y)_i)$ are a.c., hence continuous in particular. For $y \in (N_i \cup S_i)^c$, $x \in (S_i^{(y)})^c$ and $k \in \mathbb{N}$, even $F_m((x{:}y)_i) = F_{m+k}((x{:}y)_i)$, and hence by continuity, for all $y \in (N_i \cup S_i)^c$, $F_m((x{:}y)_i) = F_{m+1}((x{:}y)_i)$ for all x. Hence, writing again $u = (t{:}y)_i$, this gives a unique function $\tilde{f}_i \in L_{1,\mathrm{loc}}(\lambda^k)$ defined as

$$\tilde{f}_i(u) := \begin{cases} \lim_m F_m(u) & \text{for } u \in \mathbb{R}^k, \quad y \in (N_i \cup S_i)^c \\ 0 & \text{else} \end{cases} \tag{A.5}$$

s.t. that $\tilde{f}_i$ is a.c. w.r.t. x_i in the sense that there is λ^{k-1}-null set $\tilde{N}_i$ s.t. for $y \in \tilde{N}_i^c$ the function $x \mapsto \tilde{f}_i((x{:}y)_i)$ is a.c. By construction, $\lambda^k(\{\tilde{f}_i \neq f\} \cup \{\tilde{\partial} f \neq \partial \tilde{f}_i\}) = 0$. Applying this argument for each $i = 1, \dots, k$, we see that there is a function $\tilde{f}$ which is a.c. in k dimensions, s.t. $\lambda^k(\{\tilde{f} \neq f\} \cup \{\tilde{\nabla} f \neq \nabla \tilde{f}\}) = 0$. $\quad\square$

Proposition A.8. *Let $f \in L_{1,\mathrm{loc}}(\lambda^k)$ be a.c. in k dimensions. If its classical partial derivatives $\partial_{x_i} f$, are extended by 0 on those lines where absolute continuity fails, and the so extended gradient belongs to $L_{1,\mathrm{loc}}^k(\lambda^k)$, then there is a weak gradient of f and the extended gradient can be taken as a version of the weak gradient.*

Proof: As f is a.c. in k dimensions there exist $N_i \in \mathbb{B}^{k-1}$ such that for $y \in N_i^c$ the functions $x \mapsto f((x{:}y)_i)$ are a.c. Let $\phi \in \mathscr{D}_k$ and $y \in N_i^c$. Then $x \mapsto \phi((x{:}y)_i) \in \mathscr{D}_1$ and thus by integration by parts, for $y \in N_i^c$, we have

$$\int_{\mathbb{R}} f((x{:}y)_i) \, \partial_{x_i} \phi((x{:}y)_i) \, \lambda(dx) = -\int_{\mathbb{R}} \phi((x{:}y)_i) \, \partial_{x_i} f((x{:}y)_i) \, \lambda(dx) \tag{A.6}$$

Obviously, extending $f \partial_{x_i} \phi$, $\phi \partial_{x_i} f$ by 0 on $y \in N_i$, these two functions belong to $L_1(\lambda^k)$. Fubini thus yields a set $\tilde{N}_i \in \mathbb{B}^{k-1}$, $\lambda^{k-1}(\tilde{N}_i) = 0$, s.t. for $y \in \tilde{N}_i^c$, $x \mapsto [f \partial_{x_i} \phi]((x{:}y)_i)$, $x \mapsto [\phi \partial_{x_i} f]((x{:}y)_i)$ belong to $L_1(\lambda)$. Hence by Fubini $\int_{\mathbb{R}^k} f \partial_{x_i} \phi \, d\lambda^k = -\int_{\mathbb{R}^k} \phi \partial_{x_i} f \, d\lambda^k$. As $\partial_{x_i} f \in L_{1,\mathrm{loc}}(\lambda^k)$ by definition of absolute continuity in k dimensions, this possibly extended $\partial_{x_i} f$ is a weak derivative of f. $\quad\square$

Remark A.9. Having this "almost" coinciding of weak differentiability and absolute continuity in k dimensions in mind, we drop the notational difference of weak and classical derivatives.

Appendix B. Proofs

Appendix B.1. Preparations

Before proving Theorem 4.4, some preparations are needed.

We want to parallel the proof given in Huber (1981) credited to T. Liggett: The idea is to define for given $a \in \mathbb{R}^p$ linear functionals $\tilde{T}_{a;i}$ on the dense subset $\mathscr{C}^\infty(\bar{\mathbb{R}}^k, \mathbb{R})$ of $L_2(P_\theta)$ as

$$\tilde{T}_{a;i} : \mathscr{C}^\infty(\bar{\mathbb{R}}^k, \mathbb{R}) \to \mathbb{R}, \qquad \tilde{T}_{a;i}(\varphi) := \int D_{a;i} \, \partial_{x_i} \varphi \, dP_\theta. \tag{B.1}$$

Remark B.1. As also true for the one-dimensional location model treated in Huber (1981) and in the one-dimensional scale model in Ruckdeschel and Rieder (2010), it is not clear à priori whether this is a sound definition, i.e., whether $\tilde{T}_{a;i}$ respect equivalence classes of functions in $L_2(P_\theta)$:

As by (Dk) resp. (Dl), $D_{a;i}$ is continuously differentiable, it is bounded on compacts, hence $D_{a;i} \partial_{x_i} \varphi \in L_1(P_\theta)$ for any $\varphi \in \mathscr{C}^\infty(\bar{\mathbb{R}}^k, \mathbb{R})$; but even then, it is still not clear whether (B.1) makes a definition: Take $x^{(0)} \in \mathbb{R}^k$ so that $x_i^{(0)} D_{a;i}(x^{(0)}) \neq 0$ and P_θ Dirac measure for $\{x^{(0)}\}$.

Then obviously, $\varphi_0(x) = (x^{(0)})^\tau (x - x^{(0)}) = 0 \; P_\theta(dx)$-a.e., but it also holds, $\partial_{x_i} \varphi_0(x) D_{a;i}(x) = x_i^{(0)} D_{a;i}(x^{(0)})$ $P_\theta(dx)$-a.e., with the consequence that, although $\varphi_0 = 0 \, [P_\theta]$, $\tilde{T}_{a;i}(\varphi_0) \neq \tilde{T}_{a;i}(0) = 0$. Of course, φ_0 must be modified away from $x^{(0)}$ to some $\tilde{\varphi}$ s.t. $\tilde{\varphi}$ belongs to $\mathscr{C}^\infty(\bar{\mathbb{R}}^k, \mathbb{R})$. Luckily enough, this case cannot occur under condition (i) of Theorem 4.4, as then

$$\varphi = 0 \, [P_\theta] \qquad \Longrightarrow \qquad \tilde{T}_{a;i}(\varphi) = \int \partial_{x_i} \varphi \, D_{a;i} \, dP_\theta = 0 \tag{B.2}$$

which may be proved just along the lines of the first paragraph of Ruckdeschel and Rieder (2010, Proof to Thm. 2.2). Due to linearity of differentiation, evaluated member-wise in an P_θ-equivalence class, this shows that $\tilde{T}$ respects P_θ-equivalence classes.

Next we need a lemma showing denseness of certain sets in suitable L_2's. To do so we define for $i, j = 1, \dots, k$

$$\mathscr{D}_{D;i,j} := \{ D_{a;i} \, \partial_{x_i} \phi \mid \phi \in \mathscr{D}_k, \, a \in \mathbb{R}^p \} \tag{B.3}$$

and recalling that $K_j := \{ e_j^\tau D = 0 \}$ we introduce the decompositions corresponding to (B.4)

$$P_\theta := P_\theta^{(j)} + \bar{P}_\theta^{(j)}, \qquad \bar{P}_\theta^{(j)}(\cdot) := P_\theta(\cdot \cap K_j). \tag{B.4}$$

Lemma B.2. $\mathscr{D}_{D;i,j}$ *is dense in* $L_2(Q)$ *for any* σ*-finite measure on* $\mathbb{B}^k \cap K_j^c$. *In particular it is measure-determining for* $\mathbb{B}^k \cap K_j^c$.

Proof: Approximating $f \in L_2(Q)$ in $L_2(Q)$ by $f_n := f \mathrm{I}_{\Omega_n}$ with $\Omega_n = [-n, n]^k$, we may restrict ourselves to Ω_N for N sufficiently large. Thus we have to show that for each interval $J \subset K_j^c$ and each $\varepsilon > 0$, there is a $\hat{\phi}$ in $\mathscr{D}_{D;i,j}$ with $\| \mathrm{I}_J - \hat{\phi} \|_{L_2(P_\theta)} \leq \varepsilon$.

To this end fix $a \in \mathbb{R}^p$; as $D_{a;j}$ is continuous, the set K_j^c is open, hence is the countable union of k dimensional intervals $J_m := (l^{(m)}; r^{(m)})$, $m \in \mathbb{N}$, with $l^{(m)} < r^{(m)}$ and $|D_{a;j}| > 0$ on J_m.

So it suffices to show that any indicator to an interval $I = [\tilde{l}; \tilde{r}]$, $I \subset J_m$ with endpoints s.t. $Q(\partial I) = 0$ may be approximated in $L_2(Q)$ by functions in $\mathscr{D}_{D;i,j}$. But, for given $\varepsilon > 0$, Proposition A.1 provides an element $\varphi_0 \in \mathscr{C}_c^\infty(\mathbb{R}^k, \mathbb{R})$ such that $\| \varphi_0 - \mathrm{I}_I \|_Q < \varepsilon$. By construction its anti-derivative $\psi_0((x : y)_i) := \int \mathrm{I}_{(\tilde{l}_i^{(m)} \leq z \leq x)} \varphi_0((z : y)_i)) / D_{a;j}((z : y)_i)) \lambda(dz)$ lies in $\mathscr{C}^\infty(\bar{\mathbb{R}}^k, \mathbb{R})$ hence φ_0 in $\mathscr{D}_{D;i,j}$. In particular we may approximate the Q measure for k-dimensional intervals disjoint to K_j, which determines Q. $\square$

Appendix B.2. Proof of the Main Theorem

(ii) $\Rightarrow$ (i) of Theorem 4.4

In order to avoid specializing the case $k = 1$, define $V := 0$ there.

Fix $a \in \mathbb{R}^p$, $|a| = 1$. f being a density, $\lambda(\{f = 0, \partial_{x_i} f \neq 0\}) = 0$ for each i and we may write

$$\int (\partial_{x_i}\varphi) D_{a;i}\, dP_\theta = \int_{K^c} (\partial_{x_i}\varphi) D_{a;i}\, dP_\theta = \int_{\iota(K^c)} [(\partial_{x_i}\varphi) D_{a;i}] \circ \tau_\theta\, dF \stackrel{(ii)(a)}{=} \int_{\iota(K^c)} [(\partial_{x_i}\varphi) D_{a;i}] \circ \tau_\theta\, f\, d\lambda^k =$$

$$= \int_{K^c} (\partial_{x_i}\varphi) D_{a;i} f_\theta\, |\det \partial_x \iota_\theta|\, d\lambda^k \stackrel{(*)}{=} -\int_{K^c} \varphi\, f_\theta\, \partial_{x_i}(D_{a;i}\, |\det \partial_x \iota_\theta|) + \varphi D_{a;i}\, |\det \partial_x \iota_\theta|\, \partial_{x_i} f_\theta\, d\lambda^k =$$

$$= -\int_{K^c} \varphi p_\theta \left[\frac{\partial_{x_i}(|\det \partial_x \iota_\theta| D_{a;i})}{|\det \partial_x \iota_\theta|} + \frac{D_{a;i} \partial_{x_i} f_\theta}{f_\theta}\right] d\lambda^k \stackrel{(**)}{=} -\int \varphi p_\theta \left[\frac{\partial_{x_i}(|\det \partial_x \iota_\theta| D_{a;i})}{|\det \partial_x \iota_\theta|} + \frac{D_{a;i} \partial_{x_i} f_\theta}{f_\theta}\right] d\lambda^k$$

In equation $(*)$ we use that by (ii)(c), on K^c, f is a.c. λ^k a.e. so that integration by parts integration by parts is available without having to care about border values due to (ii)(b). By (ii)(d) the resulting integrand on the RHS of $(*)$ is in $L_2(P_\theta)$. In equation $(**)$, we used the fact that in each expression considered above, there appears at least one $D_{a;i}$ or a derivative $\partial_{x_i} D_{a;i}$; Lemma A.4 applies and hence

$$\lambda^k(\{D_{a;i} = 0\}, \{\partial_{x_i} D_{a;i} \neq 0\}) = 0$$

Representations (4.5) **and** (4.6)**:**

Writing out $\partial_x f_\theta = (\partial_x \iota_\theta)(\partial_x f) \circ \iota_\theta$, we see that

$$D^\tau_{a;\cdot} \partial_x f_\theta = a^\tau(\partial_\theta \iota_\theta) J(\partial_x \iota_\theta)(\partial_x f) \circ \iota_\theta = a^\tau(\partial_\theta \iota_\theta)(\partial_x f) \circ \iota_\theta = a^\tau \partial_\theta f_\theta \tag{B.5}$$

Thus we get

$$\frac{\sum_i \partial_{x_i}[D_{a;i}\, |\det \partial_x \iota_\theta|]}{|\det \partial_x \iota_\theta|} + \frac{D^\tau_{a;\cdot} \partial_x f_\theta}{f_\theta} - V_a \stackrel{\text{Def. } V}{=} \frac{a^\tau \partial_\theta\, |\det \partial_x \iota_\theta|}{|\det \partial_x \iota_\theta|} + \frac{D^\tau_{a;\cdot} \partial_x f_\theta}{f_\theta} \stackrel{\text{(B.5)}}{=} \frac{a^\tau \partial_\theta\, |\det \partial_x \iota_\theta|}{|\det \partial_x \iota_\theta|} + \frac{a^\tau \partial_\theta f_\theta}{f_\theta} = a^\tau \Lambda_\theta,$$

so $\Lambda_\theta \in L_2^p(P_\theta)$ by (ii)(d) and hence

$$\left(\int \nabla \varphi^\tau D_a + \varphi V_a\, dP_\theta\right)^2 = \left(\int \varphi \frac{a^\tau \partial_\theta p_\theta}{p_\theta}\, dP_\theta\right)^2 \leq \int [a^\tau \Lambda_\theta]^2\, dP_\theta \int \varphi^2\, dP_\theta,$$

which shows that $\mathscr{I}(F; a) \leq \int (a^\tau \Lambda_\theta)^2\, dP_\theta$. The upper bound may be approximated by a sequence $\varphi_n \in \mathscr{D}_k$ tending to $a^\tau \Lambda_\theta$ in $L_2(P_\theta)$ entailing (4.5) and (4.6).

(i) $\Rightarrow$ (ii) in Theorem 4.4

We will give a proof largely paralleling Huber (1981), although we may skip some of his arguments.

Well defined operators and Riesz-Fréchet: We consider the linear functionals $\tilde{T}_{a;i}$ from (B.1), defined on the dense subset $\mathscr{C}^\infty(\bar{\mathbb{R}}^k, \mathbb{R})$ of $L_2(P_\theta)$, which are well defined due to (B.2). In particular $\tilde{T}_{a;i}$ are bounded linear operators with squared operator norms bounded by $\mathscr{I}_\theta(F; a)$, hence can be extended by continuity to continuous linear operators $T_{a;i} : L_2(P_\theta) \to \mathbb{R}$ with the same operator norms. Thus Riesz Fréchet applies, yielding generating elements $g_{a;i} \in L_2(P_\theta)$ s.t.

$$T_{a;i}(\varphi) = -\int g_{a;i}\varphi\, dP_\theta \qquad \forall \varphi \in L_2(P_\theta) \qquad \text{and } \|g_{a;i}\|_{L_2(P_\theta)}^2 = \|\tilde{T}_{a;i}\| \tag{B.6}$$

We conclude inductively for $i = 1, \ldots, k$.

$\boxed{i = 1}$ **Using Fubini:** We have for $\varphi \in \mathscr{D}_k$

$$T_{a;1}(\varphi) = \int D_{a;1}\, \partial_{x_1} \varphi\, dP_\theta \tag{B.7}$$

17

On the other hand by the fundamental theorem of calculus,

$$T_{a;1}(\varphi) = -\int \varphi g_{a;1}\, dP_\theta = \int_{\mathbb{R}^k}\int_{\mathbb{R}} \mathrm{I}_{\{x_1 \geq y_1\}}\, \partial_{x_1}\varphi(x_1,y_{2:k})\lambda(dx_1)\, g_{a;1}(y)\, P_\theta(dy).$$

Now for each compact A, the integrand $\tilde{h}_1(x_1,y;A) = \mathrm{I}_A(x_1)\mathrm{I}_{\{x_1 \geq y_1\}} g_{a;1}(y)$ is in $L_1(\lambda(dx_1) \otimes P_\theta(dy))$. Fubini for Markov kernels thus yields a $\lambda \otimes P_{\theta;2:k}$-null set N_1 such that for $(x_1,y_{2:k}) \in N_1^c$, $x_1 \in A$, the function $h_1(y_1) := \tilde{h}_1(x_1,y;A)$ belongs to $L_1\big(P_{\theta;1|2:k}(dy_1|y_{2:k})\big)$. We now define for $D_{a;1} \neq 0$ the function $p_{1|2:k}^{(a;1;A)}$ as

$$[D_{a;1}p_{1|2:k}^{(a;1;A)}](x_1,y_{2:k}) := \begin{cases} \int h_1(y_1) P_{\theta;1|2:k}(dy_1|y_{2:k}) & \text{for} \quad (x_1,y_{2:k}) \in N_1^c,\, x_1 \in A \\ 0 & \text{else} \end{cases} \tag{B.8}$$

where obviously the dependence in A is such that for another compact $A' \supset A$, $p_{1|2:k}^{(a;1;A)} = p_{1|2:k}^{(a;1;A')}\mathrm{I}_A(x_1)$. Hence for arbitrary x_1 take A such that $x_1 \in A$ and eliminate the index A in the superscript where it is clear from the context. We also note that by Cauchy-Schwarz, for $P_{\theta;2:k}(dy_{2:k})$-a.e. $y_{2:k}$,

$$|[D_{a;1}p_{1|2:k}^{(a;1)}](x_1,y_{2:k})|^2 \leq \int g_{a;1}(y)^2 P_{\theta;1|2:k}(dy_1|y_{2:k}) < \infty \tag{B.9}$$

Getting rid of the dependence on a**:** To understand, how $p_{1|2:k}^{(a;1)}$ is related to $p_{1|2:k}^{(a';1)}$ for $a \neq a' \in \mathbb{R}^p$, we consider again (B.1), (B.8): Both sides of the latter must be of form $\tilde{W}^\tau a$ for some $\mathbb{R}^p$ valued $\tilde{W}$ independent of a; in particular

$$g_{a;1} = w_1^\tau a \tag{B.10}$$

for some $w_1 \in L_2^p(P_\theta)$. Hence

$$p_{1|2:k}^{(a;1)} = p_{1|2:k}^{(a';1)} \qquad \text{on } \{D_{a;1}^{(k)} \neq 0\} \cap \{D_{a';1}^{(k)} \neq 0\} \tag{B.11}$$

and, as we only need p orthogonal values of a to specify w_1, we arrive at a maximally extended $p_{1|2:k}^{(1)}$ defined on $K_1^c = \{D_{\cdot;1} \neq 0\}$. Also, for $P_{\theta;2:k}(dy_{2:k})$-a.e. $y_{2:k}$,

$$|[D_{a;1}p_{1|2:k}^{(1)}](x_1,y_{2:k})|^2 \leq |a|^2 \int |w_1(y)|^2 P_{\theta;1|2:k}(dy_1|y_{2:k}) \tag{B.12}$$

$p_{1|2}^{(1)}$ **is a density:** Plugging in this maximal definition, we get for $\varphi \in \mathscr{D}_k$, using $A = \mathrm{supp}(\varphi)$,

$$T_{a;1}(\varphi) \quad = \quad \int D_{a;1}\,\partial_{x_1}\varphi\, dP_\theta = \int [D_{a;1}\,\partial_{x_1}\varphi\, p_{1|2:k}^{(1)}](x_1,y_{2:k})\lambda(dx_1)P_{\theta;2:k}(dy_{2:k}). \tag{B.13}$$

where integrability of the integrands follows from Remark 2.3(e) and (C1)/(Ck), and for the right one from (B.12), which also entails that $P_{\theta;2:k}(dy_{2:k})$-a.s., $x_1 \mapsto D_{a;1}p_{1|2:k}^{(1)}$ is the λ-density of a σ-finite signed measure. Hence, we have shown that $P_\theta(dx_1,dy_{2:k})$ and $p_{1|2:k}(x_1,y_{2:k})\lambda(dx_1)P_{\theta;2:k}(dy_{2:k})$ when restricted to K_1^c define the same functional on the set $\mathscr{D}_{D;1,1}$, which is measure-determining for $\mathbb{B}^k \cap K_1^c$ due to Lemma B.2.

Therefore, the restriction to compacts A can be dropped entirely, and we may work with $A = \mathbb{R}$. Using Fubini once again, we see that on K_1^c, there is a $P_{\theta;2:k}(dy_{2:k})$-null set $\tilde{N}_1$, s.t. for fixed $y_{2:k} \in \tilde{N}_1^c$, the function $p_{1|2:k}^{(1)}(x_1,y_{2:k})$ is a Lebesgue density of the regular conditional distribution $P_{\theta;1|2:k}^{(0)}(dx_1|y_{2:k})$, hence non negative and in $L_1(\lambda)$.

Replacing K_1 by K: Similarly as for the dependence on a, we may extend the definition of $p_1^{(1)}(x_1,y_{2:k})$ to the set K^c: Any $\partial_{x_i}\varphi$ for $\varphi \in \mathscr{D}_k$ may also be interpreted as $\partial_{x_j}\tilde{\varphi}$ for some $\tilde{\varphi} \in \mathscr{D}_k$. More specifically,

$\tilde{\varphi} = \varphi \circ \pi_{i,j}$ with $\pi_{i,j}$ the permutation of coordinates i and j. Thus introducing for $1 \leq i, l \leq k$ operators $\tilde{T}_{a;i,j} : \mathscr{D}_k \to \mathbb{R}$, $\varphi \mapsto \int \partial_{x_i} \varphi D_{a;j} dP_\theta$, we amply see their boundedness in operator norm by $\|T_{a;j}\|$, hence extending them to $L_2(P_\theta)$ as before, giving operators $T_{a;i,j}$, we also get generating elements $g_{a;i,j} \in L_2(P_\theta)$ by Riesz-Fréchet and eventually, using denseness of $\mathscr{D}_{D;i,j}$ in $L_2(P_\theta^{(j)})$, we obtain correspondingly defined $p_{1|2:k}^{(j)}$ for $j = 1, \ldots k$. Now $p_{1|2:k}^{(j)}$ being Lebesgue densities of $P_{\theta;1|2:k}^{(0)}(dx_1|y_{2:k})$, there is a $P_{\theta;2:K}$-null set—for simplicity again $\tilde{N}_1$ —such that for $y \in \tilde{N}_1^c$, for each pair $j_1 \neq j_2$,

$$p_{1|2:k}^{(j_1)}((x,y)_1) = p_{1|2:k}^{(j_2)}((x,y)_1) \quad [\lambda(dx)] \qquad \text{on } K_{j_1}^c \cap K_{j_2}^c, \tag{B.14}$$

so we may indeed speak of a maximally extended $p_{1|2:k}$ defined on K^c.

$\boxed{i-1 \to i}$ Assume we have already shown that there is a $P_{\theta;i:k}$-null set $\tilde{N}_{i-1}$ such that for $y_{i:k} \in \tilde{N}_{i-1}^c$, $P_\theta^{(0)}$ admits some conditional density,

$$p_{1:i-1|i:k}(x_{1:i-1}, y_{i:k}) \lambda^{i-1}(dx_{1:i-1}) = P_{\theta;1:i-1|i:k}^{(0)}(dx_{1:i-1}|y_{i:k}). \tag{B.15}$$

Arguing just as for $i = 1$, we get

$$T_{a;i}(\varphi) \;=\; \int D_{a;i}\, \partial_{x_i} \varphi\, dP_\theta = \int \mathrm{I}_{\{D_{a;i} \neq 0\}}\, D_{a;i}\, \partial_{x_i} \varphi\, dP_\theta = - \int \varphi g_{a;i}\, dP_\theta.$$

Thus using the induction assumptions we proceed as before, i.e.; define $\tilde{h}_i(x_i, y; A)$ for some compact A, a the section-wise defined function $h_i(y_i) := \tilde{h}_i(x_i, y; A)$, the function $p_{1:i|i+1:k}^{(a;i;A)}$ which extends to $\mathbb{R}$ giving $p_{1:i|i+1:k}^{(a;i)}$, and where the dependence on a may be dropped, giving $p_{1:i|i+1:k}^{(i)}$. As this defines the same functional on the set $\mathscr{D}_{D;i,i}$ as the Markov kernel $P_{1:i|i+1:k}$, by Lemma B.2, $p_{1:i|i+1:k}^{(i)}$ is a conditional density defined on K_i^c. Using the coordinate permutation argument to drop the dependence on K_i, we obtain $p_{1:i|i+1:k}$ defined on K^c.

Hence the induction is complete, and we have shown that $P_\theta^{(0)}$ admits a λ^k density p_θ which we denote by

$$p_\theta(x) := p_\theta(x_{1:k}) := p_{\theta;1:k}(x_{1:k}) := p_{1:k}(x_{1:k}).$$

Showing $g_{a;i} = 0\ [\bar{P}_\theta^{(0)}]$ **:** Writing (B.8) and its analogue for general i for any fixed $a \in \mathbb{R}^p$ with $P_\theta^{(0)}$ and $\bar{P}_\theta^{(0)}$, we see that by Fubini, for y outside a $P_{\theta;-i}$-null set,

$$[D_{a;i} p_{1:i|i+1:k}]((x:y)_i) := \int_{-\infty}^{x} g_{a;i}((z:y)_i)\, p_\theta((z:y)_i) \lambda(dz) + \gamma_{a;i}((x:y)_i) \tag{B.16}$$

with

$$\gamma_{a;i}((x:y)_i) := \int_{-\infty}^{x} g_{a;i}((z:y)_i)\, \bar{P}_{\theta;i|-i}^{(0)}(dz|y) \tag{B.17}$$

We next show that for fixed $a \in \mathbb{R}^p$ and fixed y outside a $P_{\theta;-i}$-null set, the value of $\gamma_{a;i} \equiv 0$:
To this end we show that for any Borel subset B of K or equivalently for any proper or improper interval $I = [l, r] \subset K$,

$$\int_I g_{a;i}((z:y)_i)\, \bar{P}_{\theta;i|-i}^{(0)}(dz|y) = 0$$

Of course, $\int_I dP_{\theta;i|-i}^{(0)}(dz|y) = 0$, $[P_{\theta;-i}]$. Consider $\phi_n \in \mathscr{D}_k$ with $0 \leq \phi_n \leq 1$, $\phi_n \equiv 1$ on I and $\phi_n \equiv 0$ for $\{x \in \mathbb{R}^k \mid \mathrm{dist}(x, I) > 1/n\}$, and $|\partial_{x_i} \phi_n| \leq 6n$. The last bound is chosen according to the bound $|\dot{\varphi}| \leq 2c_0 \delta$ from Proposition A.1. Then $\phi_n \to \mathrm{I}_I$ pointwise, hence by dominated convergence and Cauchy-Schwartz we get

$$\int_I \phi_n^2\, dP_{\theta;i|-i}^{(0)} = \mathrm{o}(n^0), \qquad \left| \int g_{a;i} \phi_n\, dP_{\theta;i|-i}^{(0)} \right| = \mathrm{o}(n^0).$$

On the other hand let $C := \max\{|\partial_{x_i} D_{a;i}| \,\big|\, x \in \mathrm{supp}(\phi_1)\}$. Then as $D_{a;i} = 0$ on I, and because $\mathrm{supp}(\partial_{x_i}\phi_n) \subset \{x \in \mathbb{R}^k \,|\, 0 < \mathrm{dist}(x, I) \leq 1/n\}$, we have for $x \in \mathrm{supp}(\phi_1)$ that $|D_{a;i}(x)| \leq C|x| \leq C/n$, and hence

$$|\partial_{x_i}\phi_n D_{a;i}| \leq 6CI_{\{\mathrm{supp}(\phi_n) \cap I^c\}}.$$

Thus, due to the shrinking of $\{\mathrm{supp}(\phi_n) \cap I^c\}$, for $x \notin I$, $[\partial_{x_i}\phi_n D_{a;i}](x) \to 0$. Furthermore $[\partial_{x_i}\phi_n D_{a;i}](x) = 0$ on I, as $D_{a;i}(x) = 0$ on $I \subset K$ by definition, hence also $[\partial_{x_i}\phi_n D_{a;i}] \to 0$ pointwise and with dominated convergence

$$\int D_{a;i}\,\partial_{x_i}\phi_n\,dP_{\theta;i|-i} = \mathrm{o}(n^0).$$

So we have

$$\mathrm{o}(n^0) + \int_I g_{a;i}(z)\,\bar{P}^{(0)}_{i|-i}(dz) \quad = \quad \int_I g_{a;i}\phi_n\,d\bar{P}^{(0)}_{i|-i} = -\int_I \partial_{x_i}\phi_n D_{a;i}\,dP_{i|-i} - \int_I g_{a;i}\phi_n\,dP^{(0)}_{i|-i} = \mathrm{o}(n^0),$$

which implies $\gamma_{a;i} \equiv 0$ and hence, integrating by $\bar{P}^{(0)}_{\theta;-i}$ over any $A \in \mathbb{B}^{k-1}$, $g_{a;i} = 0\ [\bar{P}^{(0)}_\theta]$. Similarly, we obtain

$$g_{a;i,j} = 0 \qquad [\bar{P}^{(0)}_\theta]. \tag{B.18}$$

This also entails that

$$\int \partial_{x_i}\varphi\,D_{a;j}dP_\theta \quad = \quad \int \partial_{x_i}\varphi\,D_{a;j}p_\theta\,d\lambda^k = -\int \varphi g_{a;i,j}dP_\theta \overset{(B.18)}{=} -\int \varphi g_{a;i,j}p_\theta\,d\lambda^k. \tag{B.19}$$

Application of Proposition A.7: From (B.19), we get $\int \partial_{x_i}\varphi\,D_{a;j}p_\theta\,d\lambda^k = -\int \varphi g_{a;i,j}p_\theta\,d\lambda^k$ for all $\varphi \in \mathscr{D}_k$. By Definition A.5, $g_{a;i,j}p_\theta$ thus is the weak derivative of $D_{a;j}p_\theta$ w.r.t. x_i. By Proposition A.7 there is a modification of $D_{a;j}p_\theta$ on a λ^k-null set such that this modification—for simplicity again denoted by $D_{a;j}p_\theta$—is a.c. in k dimensions. Hence, for λ^k a.e. x, $D_{a;j}p_\theta$ is differentiable w.r.t. x_i in the classical sense with a derivative coinciding with $g_{a;i,j}p_\theta$ up to a λ^k-null set.

As $D_{a;j}$ is continuously differentiable, p_θ is differentiable on K^c_j for λ^k a.e. x, and using again all the different $D_{a;j}$, $j = 1, \ldots, k$, the same is even true on K^c.

Proof of (ii)(a)–(d): Defining for $\theta \in \Theta$

$$f^{(\theta)} := (p_\theta/|\det\partial_x\iota_\theta|)\circ\tau_\theta, \tag{B.20}$$

and recalling that $P_\theta = F\circ\iota_\theta$, we see that by the Lebesgue transformation formula, $f^{(\theta)}$ must be a density of F, hence the index θ may be dropped, and (ii)(a) follows. Once again by the transformation formula,

$$p_\theta = |\det\partial_x\iota_\theta|(f\circ\iota_\theta) = |\det\partial_x\iota_\theta|f_\theta. \tag{B.21}$$

and thus (ii)(c) holds. For (ii)(b) we consider κ defined analogously as for $k = 1$ as inverse to ℓ from (2.3): We lift (4.3) to $[0,1]^k$, giving

$$\left(\int_{[0,1]^k} \psi'\kappa q_\theta\,d\lambda\right)^2 \leq \mathscr{I}_\theta(F)\int_{[0,1]^k}\psi^2\,d[\ell(P_\theta)] \qquad \forall\psi \in \mathscr{C}^\infty([0,1]^k, \mathbb{R}),$$

where $q_\theta = p_\theta\circ\kappa$, and we have to show that $[\kappa q_\theta](u) = 0$ for $u \in \partial([0,1]^k)$. We only show $u_1 = 1$, all other cases follow similarly. Let $\psi_n \in \mathscr{D}_k$, $\psi_n \to I_{\{1\}\times[0,1]^{k-1}}$ in $L_2(\ell(P_\theta))$ and pointwise. Then by Fubini and by integration by parts

$$\int_{[0,1]^{k-1}}\int_0^1 [\partial_{x_1}(\psi_n)\kappa q_\theta]((x:y)_1)\,\lambda(dx)\,\lambda^{k-1}(dy) =$$

$$= \int_{[0,1]^{k-1}}\left[\psi_n\kappa q_\theta\big|_0^1 - \int_0^1 g\circ\kappa\,\psi_n\,[\ell(P_\theta)]_{1|2:k}(dx|y)\right][\ell(P_\theta)]_{2:k}(dy)$$

But $\int_{[0,1]^k} \psi_n^2 \, d[\ell(P_\theta)] \to 0$ entails by Fubini, $\int_0^1 \psi_n^2 \, d[\ell(P_\theta)]_{1|2:k} \to 0$ $[\ell(P_\theta)]_{2:k}(dy)$ a.e. and by Cauchy-Schwartz that also $\int_0^1 g \circ \kappa \psi_n \, d[\ell(P_\theta)]_{1|2:k} \to 0$ and hence

$$\left([\psi_n \kappa q_\theta]((1:y)_1) + \mathrm{o}(n^0) \right)^2 \le \mathrm{o}(n^0) \qquad \left[[\ell(P_\theta)]_{2:k}(dy) \right]$$

and due to continuity of $[\psi_n \kappa q_\theta]$, (ii)(b) follows. For (ii)(d) we proceed as in part (ii) $\Rightarrow$ (i)

$$(\sum_i g_{a;i} - V_a)p_\theta = (\sum_i \partial_{x_i}[D_{a;i} p_\theta]) - V_a p_\theta \stackrel{(B.21)}{=} \sum_i \partial_{x_i}[D_{a;i}|\det \partial_x \iota_\theta| f_\theta] - V_a p_\theta =$$

$$= \quad p_\theta \left(\frac{\sum_i \partial_{x_i}[D_{a;i}|\det \partial_x \iota_\theta|]}{|\det \partial_x \iota_\theta|} + \frac{\sum_i D_{a;i} \partial_{x_i} f_\theta}{f_\theta} - V_a \right) = p_\theta \left(\frac{a^\tau \partial_{\theta_j}|\det \partial_x \iota_\theta|}{|\det \partial_x \iota_\theta|} + \frac{\sum_i D_{a;i} \partial_{x_i} f_\theta}{f_\theta} \right) \quad \text{(B.22)}$$

$$= \quad p_\theta \left(\frac{a^\tau \partial_{\theta_j}|\det \partial_x \iota_\theta|}{|\det \partial_x \iota_\theta|} + \frac{a^\tau \partial_{\theta_j} f_\theta}{f_\theta} \right) = p_\theta \frac{a^\tau \partial_{\theta_j} p_\theta}{p_\theta} = p_\theta a^\tau \Lambda_\theta. \quad \text{(B.23)}$$

Now (ii)(c) follows from (B.22) and the fact that V_a and all $g_{a;i}$ are in $L_2(P_\theta)$, and assertions (4.5) and (4.6) from (B.23). $\qquad \square$

The next corollary shows that K is uninformative for our problem in the sense that $\bar{P}_\theta^{(0)}$-a.e. $\Lambda_\theta = 0$.

Corollary B.3. *Under the assumptions of Theorem 4.4, setting*

$$\Lambda_\theta := -V + \sum_i w_i \quad \text{(B.24)}$$

with w_i from (B.10) (for $i = 1$) and respectively defined otherwise, it holds that

$$\Lambda_\theta = 0 \qquad [\bar{P}_\theta^{(0)}] \quad \text{(B.25)}$$

Proof: (B.24) is defined according to (B.23) on K^c, and as (B.18) entails $\sum_i w_i = 0 \, [\bar{P}_\theta^{(0)}]$, the assertion is a direct consequence of $x \in K \iff D(x) = 0 \iff \partial_\theta \iota_\theta(x) = 0 \stackrel{(2.10)}{\Rightarrow} V(x) = 0$. $\qquad \square$

Appendix B.3. Proofs of Sections 6

For the proof of Proposition 6.1 we need two lemmas:

Lemma B.4. *The multivariate location model 3.2 is L_2-differentiable iff it is "partially" in each coordinate separately, i.e.;*

$$\int \left(\sqrt{f}(x_1, \dots, x_j + h, \dots, x_k) - \sqrt{f}(x)(1 - \tfrac{1}{2}\Lambda_{f,j}(x)) \right)^2 \lambda^k(dx) = \mathrm{o}(h) \quad \text{(B.26)}$$

Proof to Lemma B.4: Garel and Hallin (1995, Lemma 2.1) $\qquad \square$

Lemma B.5. *The multivariate location model 3.6 is L_2-differentiable iff it is "partially" in each coordinate separately, i.e.; for each $i, j = 1, \dots, k$ and each $A = A^\tau \in \mathbb{R}^{k \times k}$*

$$\int \left(\sqrt{\det}(\mathbb{I}_k + h\delta_{i,j} A)\sqrt{f}((\mathbb{I}_k + h\delta_{i,j}A)x) - \sqrt{f}(x)(1 + \tfrac{1}{2}\Lambda_{\mathbb{I}_k}(x)) \right)^2 \lambda^k(dx) = \mathrm{o}(h^2) \quad \text{(B.27)}$$

where $\delta_{i,j}$ is the matrix in $\mathbb{R}^{k \times k}$ with but 0 entries except at position i, j.

Proof to Lemma B.5: With obvious translation we may parallel Garel and Hallin (1995, Lemma 2.1). A proof is given in Ruckdeschel (2001, Lemma B.3.3). $\qquad \square$

Proof to Proposition 6.1: Putting together Lemmas B.4 and B.5, we have reduced the problem to the respective questions in the one dimensional location resp. scale model, which is proven in Hájek (1972) (one-dimensional location) and Swensen (1980, Ch.2, Sec.3) (one-dimensional scale); Ruckdeschel and Rieder (2010, Prop. 3.1) in addition shows that in the pure scale case, we may allow for mass in 0. $\qquad \square$

Appendix B.4. Proofs of Section 7

Proof to Proposition 7.4: For fixed $0 \neq a \in \mathbb{R}^p$, the proof goes through word by word as in Huber (1981), simply replacing f'_t by $a^\tau \partial_\theta \tilde{f}_t$ and f_t by $\tilde{f}_t$: By a monotone convergence argument it is shown that we may differentiate twice under the integral sign, giving

$$\frac{d^2}{dt^2} \mathscr{I}_\theta(F_t; a) = \int 2 \left(\frac{a^\tau \partial_\theta \tilde{f}_1}{\tilde{f}_1} - \frac{a^\tau \partial_\theta \tilde{f}_0}{\tilde{f}_0} \right)^2 \frac{\tilde{f}_0^2 \tilde{f}_1^2}{\tilde{f}_t^3} \, d\lambda^k.$$

So we conclude that $a^\tau \partial_\theta \log \tilde{f}_0 = a^\tau \partial_\theta \log \tilde{f}_1$ $\lambda^k(dx)$ a.e., i.e.,

$$a^\tau \partial_\theta \iota_\theta \frac{\nabla f_0}{f_0} \circ \iota_\theta(x) + a^\tau \partial_\theta \log |\det \partial_x \iota_\theta(x)| = a^\tau \partial_\theta \iota_\theta \frac{\nabla f_1}{f_1} \circ \iota_\theta(x) + a^\tau \partial_\theta \log |\det \partial_x \iota_\theta(x)|,$$

where due to (d) up to a λ^k-null set $\frac{\nabla f_0}{f_0} \circ \iota_\theta = \frac{\nabla f_1}{f_1} \circ \iota_\theta$, and hence up to a λ^k-null set $\nabla \log \tilde{f}_0 = \nabla \log \tilde{f}_1$. Integrating this out w.r.t x_i, we get by (c) that $\tilde{f}_0(x) = c_i(x_{-i}) \tilde{f}_1(x)$ for $\lambda^{(k-1)}$ almost all x_{-i}. Varying i, we see that for some $c > 0$, $c_i(x_{-i}) = c$ for all $i = 1, \dots, k$ and for λ^k almost all x, and hence

$$\mathscr{I}_\theta(F_1; a) = \int \left(\frac{a^\tau \partial_\theta \tilde{f}_1}{\tilde{f}_1}\right)^2 \tilde{f}_1 \, d\lambda^k = \int \left(\frac{a^\tau \partial_\theta \tilde{f}_0}{\tilde{f}_0}\right)^2 c \tilde{f}_0 \, d\lambda^k = c \mathscr{I}_\theta(F_0; a)$$

and $c = 1$. As this holds for any $0 \neq a \in \mathbb{R}^p$, the assertion for $\bar{\mathscr{I}}_\theta(F)$ follows. $\qquad \square$

Proof to Proposition 7.6: As by Proposition 7.1, for any $a \in \mathbb{R}^p$ the mapping $F \mapsto \mathscr{I}_\theta(F; a)$ is weakly lower-semicontinuous, the same goes for the following, recursively defined mappings: Let $a_1 \in \mathbb{R}^p$, $|a_1| = 1$ realize

$$\bar{\mathscr{I}}_{\theta;1}(F) := \bar{\mathscr{I}}_\theta(F) = \max \mathscr{I}_\theta(F; a), \quad a \in \mathbb{R}^p, |a| = 1$$

and for $i = 2, \dots, k$, assuming a_j already defined for $j = 1, \dots, i-1$, let $a_i \in \mathbb{R}^p$, $|a_i| = 1$ realize

$$\bar{\mathscr{I}}_{\theta;i}(F) := \max \mathscr{I}_\theta(F; a), \quad a \in \mathbb{R}^p, |a| = 1, a \perp \{a_j\}_{j<i}.$$

Then each of the $\bar{\mathscr{I}}_{\theta;i}(F)$, $i = 1, \dots, k$ is weakly lower-semicontinuous by the same argument as $\bar{\mathscr{I}}_\theta(F)$ and is strictly positive by assumption. Hence for each $i = 1, \dots, k$, the mapping $F \mapsto 1/\bar{\mathscr{I}}_{\theta;i}(F)$ is weakly upper-semicontinuous, and so is the sum $\sum_i 1/\bar{\mathscr{I}}_{\theta;i}(F)$. But this sum is just the trace of $[\mathscr{I}_\theta(F)]^{-1}$. The corresponding statement as to the attainment of the maximum is shown just as Corollary 7.3 $\qquad \square$

Acknowledgement

References

Bickel P.J., Klaassen C.A., Ritov Y. and Wellner J.A. (1998): *Efficient and adaptive estimation for semiparametric models*. Springer.

Dudley, R.M. (2002): *Real Analysis and Probability*. Cambridge University Press, 2nd edition.

Garel B. and Hallin M. (1995): Local asymptotic normality of multivariate ARMA processes with a linear trend. *Ann. Inst. Stat. Math.*, **47**(3): 551–579.

Hájek J. (1972): Local asymptotic minimax and admissibility in estimation. In: Le Cam L.M., Neyman J. and Scott E.L. (Eds.) *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability.*, Vol. 1: Theory of statistics., p. 175–194. University of California Press.

Huber P.J. (1964): Robust estimation of a location parameter. *Ann. Math. Stat.*, **35**: 73–101.

—— (1981): *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley.

Le Cam L. (1986): *Asymptotic methods in statistical decision theory*. Springer.

Lehmann E. (1983): *Theory of point estimation*. John Wiley & Sons.

Maz'ya V.G. (1985): *Sobolev spaces*. Springer. Transl. from the Russian by T. O. Shaposhnikova.

Ruckdeschel P. (2001): *Ansätze zur Robustifizierung des Kalman–Filters*, Vol. 64. Bayreuther Mathematische Schriften, Bayreuth.

Ruckdeschel P. and Rieder H. (2010): Fisher information of scale. Preprint. `arXiv:1005.0983`

Rudin W. (1974): *Real and complex analysis*. McGraw-Hill, 2nd edition.

—— (1991): *Functional analysis*. McGraw-Hill, 2nd edition.

Simader C.G. (2001): On absolute continuity. Unpublished manuscript, Universität Bayreuth, Germany.

Swensen A.R. (1980): *Asymptotic inference for a class of stochastic processes*. Dissertation, University of California, Berkely, CA.