Semiparametric Regression: Hadamard Differentiability and Efficient Score Functions for some Testing Problems

Tadeusz Inglot

Polish Academy of Sciences and Wrocław University of Technology

Teresa Ledwina

Polish Academy of Sciences

Abstract. The aim of this paper is to provide formal and selfcontained derivation of efficient scores for testing some hypotheses in two widely used linear semiparametric regression models.

The derivation consists of two basic steps. The first one is a calculation of Hadamard derivatives of the underlying model densities over parameters from suitable Banach spaces. The resulting derivatives are defined via some vectors, called score vectors. The second step is systematic derivation of projections of components of the score vectors related to parameters of interest onto the space spanned by components corresponding to nuisance parameters. The efficient scores are residuals resulting from these projections.

Acknowledgements. The research was supported by the KBN grant 5 P03A 03020.

Content

INTRODUCTION

The motivation of this work is some construction of sensitive test for asserting goodness of fit of some semiparametric linear regression model. Let $Z = (X, Y)$ denote a random vector in $I \times R$, $I = [0,1]$. We would like to verify the null hypothesis asserting validity of the model $M(0)$

$$Y = \beta[v(X)]^T + \epsilon,$$

where X and ϵ are independent, $E\epsilon = 0$, $E\epsilon^2 < \infty$, $\beta \in R^q$ is a vector of unknown parameters while $v(x) = (v_1(x), ..., v_q(x))$ is a vector of known functions. The distributions of X and ϵ possess densities, but are otherwise unknown. The symbol ^{T} denotes the transposition. All vectors are row ones.

The basic steps of our construction are as follows.

- Embed the null model $M(0)$ into auxiliary model $M_k(\theta)$

  $$Y = \theta[u(X)]^T + \beta[v(X)]^T + \epsilon,$$

  where $\theta \in R^k$ is a vector of unknown parameters, $u(x) = (u_1(x), ..., u_k(x))$ is a vector of known functions.

- Given a fixed k, construct score test statistic for testing $\theta = 0$ against $\theta \neq 0$ in $M_k(\theta)$.

- Incorporate into the score statistic a dimension k fitted by some score-based selection rule.

Recall that the score statistic is simply some special norm of estimated efficient score vector. The efficient score vector is obtained as residual from projections [derived under the null hypothesis] of the scores for the parameters of interest on scores for the nuisance parameters.

The aim of this paper is to provide formal and selfcontained derivation of efficient scores for testing $\theta = 0$ in the model $M(\theta)$. The rest of the programme described above was carried out in Inglot and Ledwina (2003). To allow some useful extension we shall consider in the present paper heteroscedastic model as well. Therefore we introduce the following settings

[0] $$Y = \theta[u(X)]^T + \beta[v(X)]^T + \sigma(X)\epsilon,$$

[1] $$Y = \theta[u(X)]^T + \beta[v(X)]^T + \epsilon,$$

along with the following basic model assumptions on the model [1]

$u(x) = (u_1(x), ..., u_k(x))$, $v(x) = (v_1(x), ..., v_q(x))$, and the given functions $u_1, ...u_k, v_1, ..., v_q$ are bounded and linearly independent;

$\theta \in R^k$, $\beta \in R^q$ are unknown parameters;

< M[1]1 > X has an unknown density g with respect to the Lebesgue measure λ supported on $I = [0,1]$;

ϵ has an unknown density f with respect to the Lebesgue measure λ on R. The density f satisfies $E_f\epsilon = 0$, $\tau = E_f\epsilon^2$ and $0 < \tau < \infty$, and the respective assumptions on the model [0]

the set of assumptions $< M[1]1 >$;

< M[0]1 > $E_f|\epsilon| = 1$;

$\sigma(x)$ is an unknown positive function , bounded and bounded away from 0.

In both cases we like to test $\theta = 0$ in the presence of unknown remaining parameters. To put things more formally set

$$\kappa = \kappa_{[0]} = (\theta, \eta), \quad \eta = (\beta, \sqrt{g}, \sqrt{f}, s), \quad s = -\log \sigma$$

and

$$\kappa_{[1]} = (\theta, \eta_{[1]}), \ \eta_{[1]} = (\beta, \sqrt{g}, \sqrt{f}).$$

Introduce also

$$a = (\theta, \beta) \ \text{ and } \ w(x) = (u(x), v(x)).$$

So, the first null hypothesis is $\kappa = (0, \eta)$ while the second is $\kappa_{[1]} = (0, \eta_{[1]})$. The alternatives are unrestricted.

Set $Z = (X, Y)$. Families of model densities induced by [0] and [1] have the form

[0] $$p(z;\kappa) = g(x)\frac{1}{\sigma(x)} f\left(\frac{y - w(x)a^T}{\sigma(x)}\right),$$

[1] $$p(z;\kappa_{[1]}) = g(x) f\left(y - w(x)a^T\right).$$

Corresponding distributions are denoted by P_κ and $P_{\kappa_{[1]}}$.

Calculating efficient scores is related to differentiation of the model densities and projecting resulting derivatives onto some tangent spaces. Typical approach is to consider square root of densities and treat them as elements of appropriate L_2 space. To write an increment of $p^{1/2}(z;\cdot)$ a common practice is to introduce some paths throughout which one approaches the model density. The next step then is to modify classical notions of differentiability such as e.g. that of Fréchet or Hadamard to provide some pathwise variants of them. This works, but introduces many complications. Probabilistic and analytical arguments are mixed up. Many authors consider different variants of differentiability and introduce new notions which often are not very consistent each to other. This causes that many existing results are in fact not easily accessible.

Our impact to the story is an observation that it is much more convenient to separate analytical and probabilistic arguments whenever possible. Namely, instead to solve problems in probability space with distribution indexed by few parameters [infinite dimesional also] we propose to solve related problems defined in one standard measurable space with the Lebesgue measure. This considerably simplifies the setting. To be more specific we shall present below the idea in application to the model [0].

Instead to introduce paths consider more general setting by introducing more general map than $\kappa \to p^{1/2}(z;\kappa)$ and differentiate this general map in a standard way. More precisely, $p^{1/2}(z;\kappa)$, seen as a function of κ, is a map from $\Omega \to \mathcal{H}$, where $\Omega = \mathcal{A} \times \mathcal{B} \times \mathcal{C} \times \mathcal{S}$ while $\mathcal{A} = R^{k+q}$, $\mathcal{B} = L_2(I, \lambda)$, $\mathcal{C} = L_2(R, \lambda)$, $\mathcal{S} = L_\infty(I, \lambda)$, while $\mathcal{H} = L_2(I \times R, \lambda \times \lambda)$. The specific structure of $p^{1/2}(z;\kappa)$, given above, motivates introducing an abstract map

$$\Psi : \Omega \to \mathcal{H}, \ \ \Psi(\omega) = \Delta_{wa^T}\Upsilon_s(bc),$$

where Υ_s is some scale operator while Δ_{aw^T} is some shift operator. These operators are precisely described and investigated in Section A. Roughly speaking this section generalizes widely used Hájek and Šidák (1967), pp. 210-214, results on classical location/scale family. Having defined Ψ we apply classical notion of Hadamard differentiability in Banach spaces. The derivative $\dot{\Psi}_{(a,b,c,s)}$ of Ψ is obtained in Section B. Moreover, Theorem B.7 relates the standard in statistical literature derivative of $p^{1/2}(z;\kappa)$ with $\dot{\Psi}_\kappa$. The range of $\dot{\Psi}_{(a,b,c,s)}$ for (a,b,c,s) coming from $< M[0]1 >$ is investigated in Section B[0].2. Further, Remark B.10 indicates that, having fixed this, we can again forget about probability and derive necessary projections in the restricted [by probabilistic arguments] subspaces of the introduced abstract spaces [cf. Remark B.10 and Sections C[0].1.1 - C[0].1.3].

Obviously, similar approach can be applied to model [1]. Related results are presented in section labeled by B[1] and C[1]. A reason to including this simpler case as well is our application of these results to construct a data driven smooth test we considered in Inglot and Ledwina (2003).

Closing, we would like to emphasize that primary goal of this paper is to simplify existing approach to understand carefully necessary assumptions and derivation of existing results. Basic ingredients [i.e. the residuals r^0 and $r^0_{[1]}$] of efficient scores we derived, can be found in the literature [cf. Schick (1997)]. Anyway, no derivations were provided.

A. SOME AUXILIARY PROPERTIES OF SHIFT AND SCALE OPERATORS

We prove here some auxiliary results which will be used in Section B. Set $\mathcal{H} = L_2(I \times R, \lambda \times \lambda)$ and $\mathcal{C} = L_2(R, \lambda)$, where λ is the Lebesgue measure on the corresponding space, and $I = [0,1]$. Denote by $||\cdot||_{\mathcal{H}}$ the norm in $\mathcal{H}$ and by $||\cdot||_{\mathcal{C}}$ the norm in $\mathcal{C}$. Throughout the paper all integrals are taken with respect to the Lebesgue measure if not otherwise indicated.

Theorem A.1. (Rudin 1974, Theorem 9.5) For $c \in \mathcal{C}$ and $t \in R$ put $\Delta_t^* c(y) = c(y-t)$, $y \in R$. Then for each $c \in \mathcal{C}$, $\Delta_\bullet^*$ is a uniformly continuous transformation of R into $\mathcal{C}$. In particular, for each $c \in \mathcal{C}$, $||\Delta_t^* c - c||_{\mathcal{C}} \to 0$ as $t \to 0$.

Corollary A.2. Let φ be an arbitrary, fixed measurable function on $[0,1]$. Define $\Delta_\varphi : \mathcal{H} \to \mathcal{H}$ by

$$\Delta_\varphi h(x,y) = h(x, y - \varphi(x)),\ x \in [0,1],\ y \in R,\ h \in \mathcal{H}. \tag{A.1}$$

Then

(i) Δ_φ is an isometry on $\mathcal{H}$,

(ii) for each $h \in \mathcal{H}$ it holds

$$\lim_{t\to 0} ||\Delta_{t\varphi} h - h||_{\mathcal{H}} = 0. \tag{A.2}$$

If, additionally, $\{\varphi_t,\ t \in R\}$ is a family of measurable functions on $[0,1]$ satisfying $\lim_{t\to 0} t\varphi_t(x) = 0$ for almost all x, then

(iii) for each $h \in \mathcal{H}$ it holds

$$\lim_{t\to 0} ||\Delta_{t\varphi_t} h - h||_{\mathcal{H}} = 0. \tag{A.3}$$

Proof of Corollary A.2. The property (i) is a consequence of the following equalities:

$$||\Delta_\varphi h||^2_{\mathcal{H}} = \int_0^1 \left[\int_R h^2(x, y - \varphi(x))dy\right] dx = \int_0^1 \left[\int_R h^2(x, y)dy\right] dx = ||h||^2_{\mathcal{H}}.$$

To prove (ii) observe that for almost all $x \in I$, $h(x, \cdot) \in \mathcal{C}$ and by Theorem A.1

$$\lim_{t\to 0} ||\Delta^*_{t\varphi(x)} h(x, \cdot) - h(x, \cdot)||_{\mathcal{C}} = 0 \tag{A.4}$$

for almost all x. On the other hand, as $\Delta^*_{t\varphi(x)}$ is an isometry on $\mathcal{C}$ for every $t \in R$ and $x \in I$, we have

$$||\Delta^*_{t\varphi(x)} h(x, \cdot) - h(x, \cdot)||^2_{\mathcal{C}} \le 4||h(x, \cdot)||^2_{\mathcal{C}}. \tag{A.5}$$

Hence, by (A.4), (A.5) and Lebesgue Dominated Convergence Theorem we get

$$||\Delta_{t\varphi} h - h||^2_{\mathcal{H}} = \int_0^1 \left[||\Delta^*_{t\varphi(x)} h(x, \cdot) - h(x, \cdot)||^2_{\mathcal{C}}\right] dx \to 0$$

as $t \to 0$ which gives (A.2).

The proof of (A.3) is similar. □

Before stating the next result introduce the following useful notation: $\frac{\partial}{\partial y} h = \dot{h}_y$.

Corollary A.3. Let φ be arbitrary, fixed measurable function on $[0, 1]$ and $h \in \mathcal{H}$ such that $\dot{h}_y$ exists for every $y \in R$ and almost all x and $\dot{h}_y \in \mathcal{H}$. Then

$$||\Delta_\varphi h - h + \varphi \dot{h}_y||^2_{\mathcal{H}} \le \int_0^1 ||\varphi[\Delta_{t\varphi} \dot{h}_y - \dot{h}_y]||^2_{\mathcal{H}} dt, \tag{A.6}$$

where both sides of (A.6) or only the right hand side are allowed to be equal $+\infty$.

Proof of Corollary A.3. Actually, the assumptions on h imply that for almost all $x \in I$ h, is absolutely continuous with respect to $y \in R$ (see Theorem 8.21 in Rudin 1974). Hence for almost all $x \in I$ and every $y \in R$

$$\Delta_\varphi h(x, y) - h(x, y) + \varphi(x)\dot{h}_y(x, y) = \int_y^{y-\varphi(x)} \dot{h}_y(x, \tau)d\tau + \varphi(x)\dot{h}_y(x, y) =$$

$$= \int_0^1 (-\varphi(x))\dot{h}_y(x, y - t\varphi(x))dt - (-\varphi(x))\dot{h}_y(x, y) =$$

$$= \int_0^1 (-\varphi(x)) \left[\dot{h}_y(x, y - t\varphi(x)) - \dot{h}_y(x, y)\right] dt.$$

Putting on the norm of $\mathcal{H}$ we get by Fubini Theorem

$$||\Delta_\varphi h - h + \varphi \dot{h}_y||^2_{\mathcal{H}} = \int_0^1 \int_R \left(\int_0^1 (-\varphi(x)) \left[\dot{h}_y(x, y - t\varphi(x)) - \dot{h}_y(x, y)\right] dt\right)^2 dxdy \le$$

$$\int_0^1 \int_R \int_0^1 \varphi^2(x) \left[\dot{h}_y(x, y - t\varphi(x)) - \dot{h}_y(x, y)\right]^2 dxdydt = \int_0^1 ||\varphi \left[\Delta_{t\varphi}\dot{h}_y - \dot{h}_y\right] ||^2_{\mathcal{H}} dt$$

which proves (A.6). □

Remark A.4. If $\varphi \in L_\infty(I)$ then by (i) of Corollary A.2 we have for every $t \in I$

$$||\varphi\left[\Delta_{t\varphi}\dot{h}_y - \dot{h}_y\right]||_{\mathcal{H}}^2 \leq 4||\phi||_\infty^2||\dot{h}_y||_{\mathcal{H}}^2$$

and both sides of (A.6) are finite.

Remark A.5. In the proof of Corollary A.3 we have exploited only that h is absolutely continuous with respect to y for almost all $x \in I$. So, in Corollary A.3 it is enough to assume e.g. that for almost all x the derivative $\dot{h}_y$ exists except a finite set of points in which there exist right hand side and left hand side derivatives and $\dot{h}_y \in \mathcal{H}$.

In the above statements we have established some properties of the shift operator Δ. Now, we shall formulate and prove the analogous results for scale operators Υ and Υ^*.

Theorem A.6. For $c \in \mathcal{C}$ and $t \in R$ put $\Upsilon_t^* c(y) = c(yt),\ y \in R$. For arbitrary $\tau > 0$ set $T = [\tau, \infty)$. Then for each $c \in \mathcal{C}$ and $\tau > 0$ $\Upsilon_\bullet^*$ is uniformly continuous transformation of T into $\mathcal{C}$. In particular, for each $c \in \mathcal{C}$ $||\Upsilon_t^* c - c||_{\mathcal{C}} \to 0$ as $t \to 1$.

The proof of Theorem A.6 is analogous to that of Theorem A.1 so we omit it.

Corollary A.7. Let φ be an arbitrary fixed measurable function on $[0,1]$. Define $\Upsilon_\varphi : \mathcal{H} \to \mathcal{H}$ by

$$\Upsilon_\varphi h(x,y) = \left[h(x, ye^{\varphi(x)})\right] e^{\varphi(x)/2}, \quad x \in I,\ y \in R,\ h \in \mathcal{H}. \tag{A.7}$$

Then
(i) Υ_φ is an isometry on $\mathcal{H}$,
(ii) for each $h \in \mathcal{H}$ it holds

$$\lim_{t\to 0} ||\Upsilon_{t\varphi} h - h||_{\mathcal{H}} = 0. \tag{A.8}$$

If, additionally, $\{\varphi_t,\ t \in R\}$ is a family of measurable functions on $[0,1]$ satisfying $\lim_{t\to 0} t\varphi_t(x) = 0$ for almost all x, then
(iii) for each $h \in \mathcal{H}$ it holds

$$\lim_{t\to 0} ||\Upsilon_{t\varphi_t} h - h||_{\mathcal{H}} = 0. \tag{A.9}$$

Proof of Corollary A.7. The property (i) can be proved similarly as previously by a suitable change of variables in the integral. To prove (ii) observe that by Theorem A.6 for almost all $x \in I$

$$||\Upsilon^*_{\exp\{t\varphi(x)\}} h(x,\cdot)e^{t\varphi(x)/2} - h(x,\cdot)||_{\mathcal{C}} \leq \tag{A.10}$$

$$\leq \left[||\Upsilon^*_{\exp\{t\varphi(x)\}} h(x,\cdot) - h(x,\cdot)||_{\mathcal{C}}\right] e^{t\varphi(x)/2} + ||h(x,\cdot)||_{\mathcal{C}} \left|e^{t\varphi(x)/2} - 1\right| \to 0 \text{ as } t \to 0.$$

On the other hand, for every $t \in R,\ x \in I$, we have

$$||\Upsilon^*_{\exp\{t\varphi(x)\}} h(x,\cdot)e^{t\varphi(x)/2} - h(x,\cdot)||_{\mathcal{C}}^2 \leq \tag{A.11}$$

$$\leq 2\int_R h^2(x, ye^{t\varphi(x)})e^{t\varphi(x)}dy + 2||h(x,\cdot)||_{\mathcal{C}}^2 = 4||h(x,\cdot)||_{\mathcal{C}}^2.$$

Hence, by (A.10), (A.11) and Lebesgue Dominated Convergence Theorem we get

$$||\Upsilon_{t\varphi}h - h||^2_{\mathcal{H}} = \int_0^1 \left[||\Upsilon^*_{\exp\{t\varphi(x)\}}h(x,\cdot)e^{t\varphi(x)/2} - h(x,\cdot)||^2_{\mathcal{C}}\right] dx \to 0$$

as $t \to 0$ which proves (A.8).

The proof of (A.9) is similar. □

Before formulating the next result introduce the following notation:

$$(A.12) \qquad i(y) = y \ \text{ and } \ h^{(1)}(x,y) = i(y)\dot{h}_y(x,y), \ x \in I, \ y \in R.$$

In short, $h^{(1)} = i\dot{h}_y$.

Corollary A.8. Let φ be an arbitrary, fixed measurable function on $[o,1]$ and $h \in \mathcal{H}$ such that $\frac{\partial}{\partial y}h = \dot{h}_y$ exists for every $y \in R$ and almost all x and $h^{(1)} = i\dot{h}_y \in \mathcal{H}$. Then

$$(A.13) \qquad ||\Upsilon_\varphi h - h - \varphi h^{(1)} - \frac{1}{2}\varphi h||^2_{\mathcal{H}} \leq$$

$$\leq 2\int_0^1 ||\varphi\left[\Upsilon_{t\varphi}h^{(1)} - h^{(1)}\right]||^2_{\mathcal{H}}dt + \frac{1}{2}\int_0^1 ||\varphi\left[\Upsilon_{t\varphi}h - h\right]||^2_{\mathcal{H}}dt,$$

where both sides of (A.13) or only the right hand side may be equal to $+\infty$.

Proof of Corollary A.8. For fixed x and y define $\delta(\tau) = h(x, ye^\tau)e^{\tau/2}$. By the assumption on h it follows that $\delta(\tau)$ is differentiable at each $\tau \in R$ and

$$\delta'(\tau) = ye^{3\tau/2}\dot{h}_y(x, ye^\tau) + \frac{1}{2}e^{\tau/2}h(x, ye^\tau).$$

Moreover, $\delta'(\tau)$ is integrable on each finite interval. Consequently, $\delta(\tau)$ is absolutely continuous and for almost all x and all y it holds

$$\Upsilon_\varphi h(x,y) - h(x,y) = h(x, ye^{\varphi(x)})e^{\varphi(x)/2} - h(x,y) = \delta(\varphi(x)) - \delta(0) =$$

$$\int_0^{\varphi(x)} \delta'(\tau)d\tau = \int_0^{\varphi(x)} ye^{3\tau/2}\dot{h}_y(x, ye^\tau)d\tau + \frac{1}{2}\int_0^{\varphi(x)} e^{\tau/2}h(x, ye^\tau)d\tau.$$

Inserting $\tau = t\varphi(x)$ we obtain

$$\Upsilon_\varphi h(x,y) - h(x,y) = \int_0^1 \varphi(x)ye^{3t\varphi(x)/2}\dot{h}_y(x, ye^{t\varphi(x)})dt + \frac{1}{2}\int_0^1 \varphi(x)e^{t\varphi(x)/2}h(x, ye^{t\varphi(x)})dt =$$

$$= \varphi(x)\int_0^1 \Upsilon_{t\varphi}h^{(1)}(x,y)dt + \frac{1}{2}\varphi(x)\int_0^1 \Upsilon_{t\varphi}h(x,y)dt$$

for almost all x and all y. Putting on the norm of $\mathcal{H}$ we get similarly as in Corollary A.3

$$||\Upsilon_\varphi h - h - \varphi h^{(1)} - \frac{1}{2}\varphi h||^2_{\mathcal{H}} =$$

$$\int_0^1\int_R \left(\varphi(x)\int_0^1 \left[\Upsilon_{t\varphi}h^{(1)}(x,y) - h^{(1)}(x,y)\right]dt + \frac{1}{2}\varphi(x)\int_0^1 \left[\Upsilon_{t\varphi}h(x,y) - h(x,y)\right]dt\right)^2 dxdy \leq$$

$$\int_0^1\int_R\int_0^1 2\varphi^2(x)\left[\Upsilon_{t\varphi}h^{(1)}(x,y) - h^{(1)}(x,y)\right]^2 dxdydt +$$

$$+\int_0^1\int_R\int_0^1 \frac{1}{2}\varphi^2(x)\left[\Upsilon_{t\varphi}h(x,y)-h(x,y)\right]^2 dxdydt =$$

$$2\int_0^1 ||\varphi\left[\Upsilon_{t\varphi}h^{(1)}-h^{(1)}\right]||^2_{\mathcal{H}}dt+\frac{1}{2}\int_0^1 ||\varphi\left[\Upsilon_{t\varphi}h-h\right]||^2_{\mathcal{H}}dt$$

which proves (A.13). □

Remark A.9 If $\varphi \in L_\infty(I)$ then by (i) of Corollary A.7 we have for $t \in I$

$$||\varphi\left[\Upsilon_{t\varphi}h^{(1)}-h^{(1)}\right]||^2_{\mathcal{H}} \le 4||\varphi||^2_\infty ||h^{(1)}||^2_{\mathcal{H}}$$

and

$$||\varphi\left[\Upsilon_{t\varphi}h-h\right]||^2_{\mathcal{H}} \le 4||\varphi||^2_\infty ||h||^2_{\mathcal{H}}$$

and consequently both sides of (A.13) are finite.

B. TOWARDS DIFFERENTIABILITY OF THE MODEL DENSITY

B[0]. Heteroscedastic regression

Define the following spaces $\mathcal{A}=R^{k+q}$, $\mathcal{B}=L_2(I,\lambda)$, $\mathcal{C}=L_2(R,\lambda)$, $\mathcal{H}=L_2(I\times R,\lambda\times\lambda)$, $\mathcal{S}=L_\infty(I,\lambda)$ and set $\Omega=\mathcal{A}\times\mathcal{B}\times\mathcal{C}\times\mathcal{S}$. Related norms in respective spaces shall be denoted by $||\cdot||_{\mathcal{A}}$, $||\cdot||_{\mathcal{B}}$, etc. Moreover, for a_1 and a_2 from $\mathcal{A}$, $a_1\cdot a_2$ stands for their scalar product. In some places, to improve readability, we shall use previous notation $a_1\cdot a_2=a_1a_2^T$.

Square root of the model density $p(z;\kappa)$, $z=(x,y)\in I\times R$, seen as a function of κ, can be considered as a map from Ω to $\mathcal{H}$. Indeed, putting $\tilde a = a$, $\tilde b=\sqrt{g}$, $\tilde c=\sqrt{f}$ and $\tilde s = s = -\log\sigma$ we get

$$p^{1/2}((x,y);\kappa)=\Delta_{w(x)\cdot\tilde a}\Upsilon_{\tilde s(x)}\left(\tilde b(x)\tilde c(y)\right).$$

This observation naturally leads to introducing the map

$$\Psi:\Omega\to\mathcal{H},\quad \Psi(\omega)=\Delta_{w\cdot a}\Upsilon_s(bc),$$

with $\omega=(a,b,c,s)$ and $w=(w_1,\ldots,w_{k+q})$, where $w_i\in\mathcal{S}$, $i=1,\ldots,k+q$, and deciding on its differentiability. In our application Hadamard differentiability is useful and sufficiently strong notion.

B[0].1. Hadamard differentiability of Ψ

To fix some notations and terminology we recall now an adjusted definition of Hadamard differentiability.

Let $(\mathcal{D}_1,||\cdot||_{\mathcal{D}_1})$ and $(\mathcal{D}_2,||\cdot||_{\mathcal{D}_2})$ be two Banach spaces and let $T:\mathcal{D}_1\to\mathcal{D}_2$ be a map.

Definition B.1. Let $d\in\mathcal{D}_1$. We say that T is Hadamard differentiable at d if there exists a continuous linear operator $\dot T_d:\mathcal{D}_1\to\mathcal{D}_2$ such that for all $t_n\in R$, $t_n\to 0$, all $d_0\in\mathcal{D}_1$ and all $\{d_n\}\subset\mathcal{D}_1$ with $||d_n-d_0||_{\mathcal{D}_1}\to 0$ it holds

$$\left|\left|\frac{T(d+t_nd_n)-T(d)-t_n\dot T_d(d_n)}{t_n}\right|\right|_{\mathcal{D}_2}\to 0. \qquad (B.1)$$

Remark B.2. Obviously, equivalently one can write in (B.1) $\dot{T}_d(d_0)$ instead of $\dot{T}_d(d_n)$. One can also say that $d_0 \in \mathcal{D}_1$ defines a direction at which we approach d. It is worth noticing that considering square roots of the probability densities $p(z;\kappa)$ and the model $Y = a[w(X)]^T + \sigma(X)\epsilon$ there are some natural restrictions on the directions we may consider [cf. (B.9), (B.10), below]. Anyway, in this section we consider simply an abstract setting and no restrictions on the set of directions shall be imposed.

When checking (B.1) it is convenient to introduce

$$\mathcal{R}_n(T) = \frac{T(d + t_n d_n) - T(d) - t_n \dot{T}_d(d_n)}{t_n}.$$

Define now some auxiliary mappings.

$$\Psi^{(1)} : \Omega \to \mathcal{A} \times \mathcal{H} \times \mathcal{S}, \quad \Psi^{(1)}(a,b,c,s) = (a, bc, s), \tag{B.2}$$

$$\Psi^{(2)} : \mathcal{A} \times \mathcal{H} \times \mathcal{S} \to \mathcal{A} \times \mathcal{H}, \quad \Psi^{(2)}(a,h,s) = (a, \Upsilon_s h), \tag{B.3}$$

and

$$\Psi^{(3)} : \mathcal{A} \times \mathcal{H} \to \mathcal{H}, \quad \Psi^{(3)}(a,h) = \Delta_{w \cdot a} h. \tag{B.4}$$

With these notations for $\Psi(a,b,c,s) = \Delta_{w \cdot a} \Upsilon_s(bc)$ it holds that

$$\Psi = \Psi^{(3)} \circ \Psi^{(2)} \circ \Psi^{(1)},$$

where $\circ$ denotes the superposition.

To prove Hadamard differentiability of Ψ we shall exploit the well known fact that superposition of Hadamard differentiable mappings is Hadamard differentiable and the chain rule holds i.e.

$$\dot{\Psi}_{(a,b,c,s)} = \dot{\Psi}^{(3)}_{\Psi^{(2)} \circ \Psi^{(1)}(a,b,c,s)} \circ \dot{\Psi}^{(2)}_{\Psi^{(1)}(a,b,c,s)} \circ \dot{\Psi}^{(1)}_{(a,b,c,s)}$$

[cf. e.g. Bickel et al. (1993), Proposition 1, p. 455]. Therefore in successive propositions we prove Hadamard differentiability of $\Psi^{(j)}$, $j = 1,2,3$.

Proposition B.3. Let (a,b,c,s) and (a_0,b_0,c_0,s_0) be two arbitrary points from Ω. Then the map $\Psi^{(1)}$ is Hadamard differentiable at each (a,b,c,s) and

$$\dot{\Psi}^{(1)}_{(a,b,c,s)}(a_0,b_0,c_0,s_0) = (a_0, bc_0 + b_0 c, s_0). \tag{B.5}$$

Proof. Let $(a_n,b_n,c_n,s_n) \to (a_0,b_0,c_0,s_0)$ and $t_n \to 0$. Then

$$\mathcal{R}_n(\Psi^{(1)}) = \frac{1}{t_n}\left(0, t_n^2 b_n c_n, 0\right).$$

Hence it follows $\mathcal{R}_n(\Psi^{(1)}) \to 0$ in $\mathcal{A} \times \mathcal{H} \times \mathcal{S}$. □

Before formulating the next result recall that in (A.12) we have defined $h^{(1)} = i\dot{h}_y$.

Proposition B.4. Let (a_0, h_0, s_0) be an arbitrary point from $\mathcal{A} \times \mathcal{H} \times \mathcal{S}$. Then the map $\Psi^{(2)}$ is Hadamard differentiable at each (a, h, s) such that $\dot{h}_y(x, y)$ exists for every y and almost all x and $h^{(1)} \in \mathcal{H}$. Moreover,

$$\dot{\Psi}^{(2)}_{(a,h,s)}(a_0, h_0, s_0) = \left(a_0, s_0 \Upsilon_s(h^{(1)} + \frac{1}{2}h) + \Upsilon_s h_0\right). \tag{B.6}$$

Proof. Let $(a_n, h_n, s_n) \to (a_0, h_0, s_0)$ and $t_n \to 0$. Then

$$\mathcal{R}_n(\Psi^{(2)}) = \frac{1}{t_n}\left(0, \Psi^{(2)}(h + t_n h_n, s + t_n s_n) - \Psi^{(2)}(h, s) - t_n \dot{\Psi}^{(2)}_{(h,s)}(h_n, s_n)\right) =$$

$$= (0, \mathcal{R}_{1n}(\Psi^{(2)})) + (0, \mathcal{R}_{2n}(\Psi^{(2)})),$$

where

$$\mathcal{R}_{1n}(\Psi^{(2)}) = \frac{1}{t_n}\left[\Upsilon_s\left(\Upsilon_{t_n s_n} h - h - t_n s_n h^{(1)} - \frac{1}{2} t_n s_n h\right)\right],$$

$$\mathcal{R}_{2n}(\Psi^{(2)}) = \frac{1}{t_n}\left[\Upsilon_s\{\Upsilon_{t_n s_n}(t_n h_n) - t_n h_n\}\right].$$

By Corollary A.7 (i), Υ_s is an isometry. Therefore, by (A.13) applied to $\varphi = t_n s_n$ we get

$$||\mathcal{R}_{1n}(\Psi^{(2)})||^2_{\mathcal{H}} = \frac{1}{t_n^2}||\Upsilon_{t_n s_n} h - h - t_n s_n h^{(1)} - \frac{1}{2} t_n s_n h||^2_{\mathcal{H}} \le$$

$$2\int_0^1 ||s_n(\Upsilon_{\tau t_n s_n} h^{(1)} - h^{(1)})||^2_{\mathcal{H}} d\tau + \frac{1}{2}\int_0^1 ||s_n(\Upsilon_{\tau t_n s_n} h - h)||^2_{\mathcal{H}} d\tau.$$

By (A.9), Remark A.9 and Lebesgue Dominated Convergence Theorem we obtain that $\mathcal{R}_{1n}(\Psi^{(2)})$ tends to 0 in $\mathcal{H}$.

Similarly we have

$$||\mathcal{R}_{2n}(\Psi^{(2)})||_{\mathcal{H}} = ||\Upsilon_s(\Upsilon_{t_n s_n} h_n - h_n)||_{\mathcal{H}} = ||\Upsilon_{t_n s_n} h_n - h_n||_{\mathcal{H}} \le ||\Upsilon_{t_n s_n} h_0 - h_0||_{\mathcal{H}} + 2||h_n - h_0||_{\mathcal{H}}.$$

Again by (A.9) and the assumption $||h_n - h_0||_{\mathcal{H}} \to 0$ it follows that $\mathcal{R}_{2n}(\Psi^{(2)}) \to 0$ in $\mathcal{H}$. In view of the form of $\mathcal{R}_n(\Psi^{(2)})$ the proof is concluded. □

Proposition B.5. The map $\Psi^{(3)}$ is Hadamard differentiable at each point (a, h) such that $\dot{h}_y(x, y)$ exists for every y and almost all x and $\dot{h}_y \in \mathcal{H}$. Moreover, for each $(a_0, h_0) \in \mathcal{A} \times \mathcal{H}$

$$\dot{\Psi}^{(3)}_{(a,h)}(a_0, h_0) = -\Delta_{w\cdot a}\dot{h}_y[w \cdot a_0] + \Delta_{w\cdot a} h_0. \tag{B.7}$$

Proof. Let $(a_n, h_n) \to (a_0, h_0)$ and $t_n \to 0$. Then

$$\mathcal{R}_n(\Psi^{(3)}) = \frac{1}{t_n}\left[\Psi^{(3)}(a + t_n a_n, h + t_n h_n) - \Psi^{(3)}(a, h) - t_n \dot{\Psi}^{(3)}_{(a,h)}(a_n, h_n)\right] =$$

$$= \mathcal{R}_{1n}(\Psi^{(3)}) + \mathcal{R}_{2n}(\Psi^{(3)}),$$

where

$$\mathcal{R}_{1n}(\Psi^{(3)}) = \frac{1}{t_n}\left[\Delta_{w\cdot a}(\Delta_{t_n w\cdot a_n} h - h + t_n \dot{h}_y[w \cdot a_n])\right],$$

$$\mathcal{R}_{2n}(\Psi^{(3)}) = \frac{1}{t_n}\left[\Delta_{w\cdot a}\{\Delta_{t_n w\cdot a_n}(t_n h_n) - t_n h_n\}\right].$$

Since $\Delta_{w\cdot a}$ is an isometry we get by (A.6) applied to $\varphi = t_n w \cdot a_n$ the following

$$||\mathcal{R}_{1n}(\Psi^{(3)})||^2_{\mathcal{H}} = \frac{1}{t_n^2}\int_0^1 \left|\left|(t_n w\cdot a_n)\left[\Delta_{\tau t_n w\cdot a_n}\dot{h}_y - \dot{h}_y\right]\right|\right|^2_{\mathcal{H}} d\tau.$$

Using now boundedness of the coordinate functions of w, Remark A.4 and (A.3) we infer again by Lebesgue Dominated Convergence Theorem that $||\mathcal{R}_{1n}(\Psi^{(3)})||_{\mathcal{H}}$ tends to 0.

To prove that $||\mathcal{R}_{2n}(\Psi^{(3)})||_{\mathcal{H}} \to 0$ we exploit again the isometry property of $\Delta_{w\cdot a}$ and $\Delta_{t_n w\cdot a_n}$ and the triangle inequality thus obtaining

$$||\mathcal{R}_{2n}(\Psi^{(3)})||_{\mathcal{H}} = ||\Delta_{t_n w\cdot a_n}h_n - h_n||_{\mathcal{H}} \leq ||\Delta_{t_n w\cdot a_n}h_0 - h_0||_{\mathcal{H}} + 2||h_n - h_0||_{\mathcal{H}}.$$

As $||h_n - h_0||_{\mathcal{H}} \to 0$, (A.3) implies $||\mathcal{R}_{2n}(\Psi^{(3)})||_{\mathcal{H}} \to 0$. □

Theorem B.6. The map $\Psi = \Psi^{(3)} \circ \Psi^{(2)} \circ \Psi^{(1)} : \Omega \to \mathcal{H}$ is Hadamard differentiable at each point $\omega = (a, b, c, s)$ such that c is differentiable for every $y \in R$, $\int_R [c'(y)]^2 dy < \infty$ and $\int_R [yc'(y)]^2 dy < \infty$. Moreover, for any $(a_0, b_0, c_0, s_0) \in \Omega$ it holds

$$(B.8) \qquad \dot{\Psi}^{(3)}_{(a,b,c,s)}(a_0, b_0, c_0, s_0) =$$
$$= \Delta_{w\cdot a}\Upsilon_s\left(-e^s bc'[w\cdot a_0] + [ic' + \frac{1}{2}c]bs_0 + bc_0 + b_0 c\right).$$

Proof. By Propositions B.3, B.4 and B.5 and the chain rule of Hadamard derivative we obtain

$$\dot{\Psi}_{(a,b,c,s)}(a_0, b_0, c_0, s_0) = \dot{\Psi}^{(3)}_{\Psi^{(2)}(a,bc,s)} \circ \dot{\Psi}^{(2)}_{(a,bc,s)}(a_0, bc_0 + b_0 c, s_0) =$$
$$= \dot{\Psi}^{(3)}_{(a,\Upsilon_s(bc))}\left(a_0, s_0\Upsilon_s([ic' + \frac{1}{2}c]b) + \Upsilon_s(bc_0 + b_0 c)\right) =$$
$$-\left[\Delta_{w\cdot a}\left(\frac{\partial}{\partial y}\Upsilon_s(bc)\right)[w\cdot a_0] + \Delta_{w\cdot a}\left(\Upsilon_s([ic' + \frac{1}{2}c]bs_0) + \Upsilon_s(bc_0 + b_0 c)\right)\right].$$

Since $\frac{\partial}{\partial y}\Upsilon_s(bc) = [\Upsilon_s(bc')]e^s$ the proof is concluded. □

B[0].2. Hadamard differentiability of $p^{1/2}(z;\kappa)$

Recall that our model is as follows

$$Y = a[w(X)]^T + \sigma(X)\epsilon,$$

where $a = (\theta, \beta)$, ϵ has a density f while X has a density g. Basic model assumptions $< M[0]1 >$ have been stated in Introduction.

Now we shall discuss how we can approach such model in the space of densities and how it reflects the related directions in $\mathcal{B}$ and $\mathcal{C}$.

Obviously f and g always satisfy $\int_R f d\lambda = \int_I g d\lambda = 1$. So, if one wants to approach $p^{1/2}(z;\kappa)$ throughout some "paths" within the space of densities, natural restriction is as follows. Take $b = \sqrt{g}$ and disturb b by $b_n \in \mathcal{B}$, $b_n \to b_0$. Let $\{t_n\}$ be a real sequence tending to 0. Then we require: for large n, $[b + t_n b_n]^2$ is a probability density [with respect to λ in our setting]. Equivalently, $\int_I [b+t_nb_n]^2 d\lambda = 1$ for large n. This implies that b_0 has to satisfy $\int_I b_0 b d\lambda = 0$ or alternatively $\int_I b_0\sqrt{g}d\lambda = 0$. Therefore, for a given fixed $b \in \mathcal{B}$ we introduce $\mathcal{B}_0 \subset \mathcal{B}$

$$(B.9) \qquad \mathcal{B}_0 = \left\{ b_0 \in \mathcal{B} : \int_I b_0 b d\lambda = 0 \right\}.$$

Analogously, setting $c = \sqrt{f}$ and asking $[c+t_nc_n]^2$, $c_n \to c_0$, $t_n \to 0$ to be a probability density obeying model assumptions $< M[0]1 >$ we get the following set $\mathcal{C}_0$ of directions from which we can approach the given fixed element $c \in \mathcal{C}$.

$$(B.10) \qquad \mathcal{C}_0 = \left\{ c_0 \in \mathcal{C} : \int_R c_0 c d\lambda = \int_R i c_0 c d\lambda = \int_R |i| c_0 c d\lambda = 0 \right\},$$

where i is the identity function introduced in (A.12). Since no further restrictions on a_0 and s_0 are necessary, we finally define $\Omega_0 = \mathcal{A} \times \mathcal{B}_0 \times \mathcal{C}_0 \times \mathcal{S}$.

Complete now the model assumptions $< M[0]1 >$ imposed in Introduction by the following ones

$< \mathrm{\mathbf{M[0]2}} >$ $\qquad f'(y)$ exists for all $y \in R$, $J = \int_R \frac{[f']^2}{f} d\lambda < \infty$, $\int_R \frac{[if']^2}{f} d\lambda < \infty$.

Take now f and g satisfying $< M[0]1 >$ and $< M[0]2 >$ and

$$\omega = \kappa = (a, \sqrt{g}, \sqrt{f}, s).$$

Moreover, consider a sequence $\{\omega_n\} \subset \Omega$, $\omega_n \to \omega_0 \in \Omega_0$ and $t_n \to 0$. The assumption $< M[0]2 >$ ensures that Theorem B.6 is applicable for $\omega = \kappa$. Therefore we can write

$$(B.11) \qquad \frac{1}{t_n}||p^{1/2}(\bullet;\kappa+t_n\omega_n) - p^{1/2}(\bullet;\kappa) - \frac{1}{2}t_n\left[\frac{\dot{\Psi}_\kappa(\omega_n)}{\frac{1}{2}p^{1/2}(\bullet;\kappa)}\right] p^{1/2}(\bullet;\kappa)||_{\mathcal{H}} \to 0.$$

(B.11) shows that $\dot{\Psi}_\kappa(\bullet)/[\frac{1}{2}p^{1/2}(\bullet;\kappa)]$ is the standard form of Hadamard derivative of $p^{1/2}(\bullet;\kappa)$, according to the convention accepted in most of statistical literature [cf. van der Vaart (1991), e.g.]. Let us denote this derivative by $\dot{p}_\kappa$. So, we have

$$(B.12) \qquad \dot{p}_\kappa(\bullet) = \frac{\dot{\Psi}_\kappa(\bullet)}{\frac{1}{2}p^{1/2}(\bullet;\kappa)}.$$

Following van der Vaart (1991) we like to emphasize also that (B.11) "is related to but weaker than the assumption of Hellinger differentiability in Begun, Hall, Huang and Wellner (1983)... The condition in Begun, Hall, Huang and Wellner is Fréchet differentiability."

Finally observe that (B.8) yields for any $\omega_0 \in \Omega_0$

$$(B.13) \qquad \dot{p}_\kappa(\omega_0) = \Delta_{w\cdot a}\Upsilon_s\left(-[w\cdot a_0]\left[\frac{f'}{f}\right]e^s + \left[1+i\frac{f'}{f}\right]s_0 + \frac{b_0}{\sqrt{g}} + \frac{c_0}{\sqrt{f}}\right).$$

We summarize the above as follows.

Theorem B.7. Under $< M[0]1 >$ and $< M[0]2 >$ $p^{1/2}(\bullet;\kappa)$ is Hadamard differentiable with the derivative $\dot{p}_\kappa$ given by (B.13).

Remark B.8. The linear operator $\dot{p}_\kappa(\bullet)$ is defined by the vector

$$(B.14) \qquad \Delta_{w\cdot a}\Upsilon_s\left(-u\left[\frac{f'}{f}\right]e^s, -v\left[\frac{f'}{f}\right]e^s, \left[1+i\frac{f'}{f}\right], \frac{1}{\sqrt{f}}, \frac{1}{\sqrt{g}}\right).$$

We shall call (B.14) the score vector. Observe that the form of (B.14) is not effected by the restrictions on directions Ω_0 from which we can approach the model density.

Remark B.9. Note that $\dot{p}_\kappa(\omega_0) \in \mathcal{T}$, where $\mathcal{T}$ is the tangent space defined on p. 376 of Schick (1997). The fact that the tangent space is spanned by the restricted set of directions Ω_0 plays essential role when calculating projections of some components of (B.14) onto the subspace spanned by the remaining components of (B.14). This point shall be treated in Section C[0].1.

We close this section by the following remark concerning projections.

Remark B.10. The argument relating $\dot{\Psi}_\kappa$ to $\dot{p}_\kappa$ shows also that to calculate projections of components of (B.14) onto some sets in $L_2(I \times R, \mu_\kappa)$,
$d\mu_\kappa/d\lambda \times \lambda = p(\bullet;\kappa)$, it is enough to calculate projections of $\dot{\Psi}_\kappa$ onto related subspaces in $\mathcal{H} = L_2(I \times R, \lambda \times \lambda)$ and, at a final stage, to divide the resulting expressions by $\frac{1}{2}p^{1/2}(\bullet;\kappa)$. We shall proceed in this way in Section C.

B[1]. Homoscedastic regression

We shall exploit notations, ideas and results of Section B[0] whenever possible.

B[1].1. Hadamard differentiability of Φ

Set

$$(B.15) \qquad \Omega_{[1]} = \mathcal{A} \times \mathcal{B} \times \mathcal{C}, \quad \Phi : \Omega_{[1]} \to \mathcal{H}, \quad \Phi(\omega) = \Delta_{w\cdot a}(bc).$$

Additionally set

$$\Phi^{(1)} : \Omega_{[1]} \to \mathcal{A} \times \mathcal{H}, \quad \Phi^{(1)}(a,b,c) = (a, bc),$$
$$\Phi^{(3)} : \mathcal{A} \times \mathcal{H} \to \mathcal{H}, \quad \Phi^{(3)}(a,h) = \Delta_{a\cdot w}h.$$

Hence $\Phi = \Phi^{(3)} \circ \Phi^{(1)}$ and arguing similarly as in the case of (B.5) and (B.7) we get

Theorem B.11. The map $\Phi = \Phi^{(3)} \circ \Phi^{(1)} : \Omega_{[1]} \to \mathcal{H}$ is Hadamard differentiable at each point $\omega_{[1]} = (a,b,c)$ such that $c(y)$ is differentiable for every $y \in R$ and $\int_R [c']^2 d\lambda < \infty$. Moreover, for any $(a_0, b_0, c_0) \in \Omega_{[1]}$ it holds that

$$(B.16) \qquad \dot{\Phi}_{(a,b,c)}(a_0,b_0,c_0) = \dot{\Phi}^{(3)}_{\Phi^{(1)}(a,b,c)} \circ \dot{\Phi}^{(1)}_{(a,b,c)}(a_0,b_0,c_0) =$$
$$\Delta_{w\cdot a}\left(-bc'[w \cdot a_0] + b_0c + bc_0\right).$$

B[1].2. Hadamard differentiability of $p^{1/2}(z;\kappa_{[1]})$

Now the model is: $Y = a[w(X)]^T + \epsilon$, $a = (\theta, \beta)$, and the basic assumptions are listed in Introduction under the label $< M[1]1 >$. Arguing as in Section B[0].2 we get the following sets of directions from which we can approach $b = \sqrt{g}$ and $c = \sqrt{f}$

$$\mathcal{B}_0 = \{b_0 \in \mathcal{B} : \int_I b_0 b \, d\lambda = 0\},$$

$$\mathcal{C}_{0[1]} = \{c_0 \in \mathcal{C} : \int_R c_0 c \, d\lambda = \int_R i c_0 c \, d\lambda = 0\}.$$

Take now $\Omega_{0[1]} = \mathcal{A} \times \mathcal{B}_0 \times \mathcal{C}_{0[1]}$ and $\omega_{0[1]} \in \Omega_{0[1]}$. Additionally to $< M[1]1 >$ assume

< M[1]2 > $\qquad f'(y)$ exists for all $y \in R$ and $J = \int_R \frac{[f']^2}{f} d\lambda < \infty.$

Under such assumptions Theorem B.11 holds for $\omega_{[1]} = (\theta, \beta, \sqrt{g}, \sqrt{f}) = \kappa_{[1]}$. Hence

$$\dot{p}_{\kappa_{[1]}}(\omega_{0[1]}) = \Delta_{w \cdot a}\left(-[w \cdot a_0]\frac{f'}{f} + \frac{b_0}{\sqrt{g}} + \frac{c_0}{\sqrt{f}}\right). \tag{B.17}$$

The above yields

Theorem B.12. Under $< M[1]1 >$ and $< M[1]2 >$ $p^{1/2}(\cdot; \kappa_{[1]})$ is Hellinger differentiable with the derivative given in (B.17). Therefore the related score vector has the form

$$\Delta_{w \cdot a}\left(-u\left[\frac{f'}{f}\right], -v\left[\frac{f'}{f}\right], \frac{1}{\sqrt{f}}, \frac{1}{\sqrt{g}}\right). \tag{B.18}$$

C[0]. General heteroscedastic model

Take now $a = (\theta, \beta)$, $\kappa = (\theta, \eta)$, $\eta = (\beta, \sqrt{g}, \sqrt{f}, s)$. Recall that the score vector (B.14) has the form

$$(C.1) \qquad \Delta_{w \cdot a}\Upsilon_s\left(-u\left[\frac{f'}{f}\right]e^s, -v\left[\frac{f'}{f}\right]e^s, [1+i\frac{f'}{f}], \frac{1}{\sqrt{f}}, \frac{1}{\sqrt{g}}\right)$$

where $i(y) = y$, $y \in R$. Moreover, observe that under the null hypothesis $\theta = 0$ we have $a = (0, \beta)$ in (C.1).

As in a parametric case, under $a = (0, \beta)$, we are searching for the projection of components of $\Delta_{w \cdot a}\Upsilon_s(-u[\frac{f'}{f}]e^s)$ onto the space spanned by the remaining components of (C.1). To find them we shall apply Remark B.10 and find first related projections in $\mathcal{H}$. Obviously, the considerations shall be restricted to elements of $\mathcal{H}$ satisfying model assumptions, only. So, first of all we are reformulating $< M[0]1 >$ and $< M[0]2 >$ in terms of functions b, c, s, where $b \in \mathcal{B} = L_2(I, \lambda)$, $c \in \mathcal{C} = L_2(R, \lambda)$, $s \in \mathcal{S} = L_\infty(I, \lambda)$. For completeness recall also that $\mathcal{A} = R^{k+q}$ while $\mathcal{H} = L_2(R \times I, \lambda \times \lambda)$.

Throughout the Section C[0] we assume that given $b \in \mathcal{B}$ and $c \in \mathcal{C}$ satisfy

< **A1** > $\qquad b > 0 \ \lambda\text{ - a.e.}, \ \int_I b^2 d\lambda = 1.$

< **A2** > $\qquad c > 0 \ \lambda\text{ - a.e.}, \ \int_R c^2 d\lambda = 1.$

< **A3** > $\qquad \int_R ic^2 d\lambda = 0, \ \int_R |i| c^2 d\lambda = 1, \int_R i^2 c^2 d\lambda = \tau, \ \tau \in (0, \infty).$

< **A4** > $\qquad c'(y)$ exists for all $y \in R$, $c' \in \mathcal{C}$ and $c_* \in \mathcal{C}$, where

$$(C.2) \qquad c_* = ic' + \frac{1}{2}c.$$

Moreover, given functions $u_1(x), ..., u_k(x), v_1(x), ..., v_q(x)$ and the function $\sigma(x)$, $x \in I$, satisfy

< **A5** > $\qquad u_1, ..., u_k, v_1, ..., v_q$ are linearly independent and belong to $\mathcal{S}$,

$$s = -\log \sigma \in \mathcal{S}.$$

Following Section B[0].2, Remark B.9, we shall restrict in C[0] our attention to the directions from $\Omega_0 = \mathcal{A} \times \mathcal{B}_0 \times \mathcal{C}_0 \times \mathcal{S}$, only, where

$$\mathcal{B}_0 = \{b_0 \in \mathcal{B} : \int_I b_0 b d\lambda = 0\},$$

$$\mathcal{C}_0 = \{c_0 \in \mathcal{C} : \int_R c_0 c d\lambda = \int_R ic_0 c d\lambda = \int_R |i| c_0 c d\lambda = 0\}.$$

The assumption < $A4$ > allows to apply Theorem B.6 which yields [cf. (B.8) and (C.2)]

$$\dot{\Psi}_{(a,b,c,s)}(a_0, b_0, c_0, s_0) = \Delta_{w\cdot a}\Upsilon_s(-e^s bc'[w \cdot a_0] + bc_* s_0 + bc_0 + b_0 c).$$

The structure of $\dot{\Psi}$ and our task suggest to introduce the following subspaces of $\mathcal{H}$

$$\mathcal{J} = \{b_0 c + bc_0 + bc_* s_0 : b_0 \in \mathcal{B}_0, c_0 \in \mathcal{C}_0, s_0 \in \mathcal{S}\}$$

$$\mathcal{F}_1 = \{e^s bc' u \cdot \theta_0 : \theta_0 \in R^k\} = \{bc'\frac{u}{\sigma} \cdot \theta_0 : \theta_0 \in R^k\},$$

$$\mathcal{F}_2 = \{e^s bc' v \cdot \beta_0 : \beta_0 \in R^q\} = \{bc'\frac{v}{\sigma} \cdot \beta_0 : \beta_0 \in R^q\}.$$

In Section C[0]1 we are searching for orthogonal projections of components of $-\Delta_{v\cdot\beta}\Upsilon_s(bc'\frac{u}{\sigma})$ onto the space $\Delta_{v\cdot\beta}\Upsilon_s(\mathcal{J}+\mathcal{F}_2)$. Since $-\Delta_{v\cdot\beta}\Upsilon_s$ is the isometry, it is therefore enough to find

$$\Pi(bc'\frac{u}{\sigma}|\mathcal{J} + \mathcal{F}_2), \tag{C.3}$$

where $\Pi(\cdot \mid \cdot)$ denotes the projection of components of a given vector onto an indicated space. The projection (C.3) shall be found in several steps. The basic problem of finding $\Pi(bc'\frac{u}{\sigma}|\mathcal{J})$ is solved in Section C[0].1.2.

To simplify further reading we first introduce some auxiliary notations and collect some simple useful formulae.

C[0].1. Projections of chosen components of $\dot{\Psi}_{(a,b,c,s)}$

C[0].1.1. Notations and auxiliary facts

For short we shall abbreviate $\int_R c(y)dy$, $\int_I \int_R h(x,y)dxdy$, etc. or $\int_R cd\lambda$, $\int_I \int_R hd\lambda$, etc. to $\int_R c$, $\int\int_{IR} h$ etc. For $h \in \mathcal{H}$ we set

$$\bar{h} = \int\int_{IR} hbc. \tag{C.4}$$

Define the measure μ on I by

$$\frac{d\mu}{d\lambda} = b^2, \tag{C.5}$$

and set

$$m = E_\mu \frac{w(X)}{\sigma(X)} = E_\mu \left(\frac{u(X)}{\sigma(X)}, \frac{v(X)}{\sigma(X)}\right) = (m_1, m_2). \tag{C.6}$$

By < $A2$ > it follows that c and related c_* [cf. (C.2)] are linearly independent. However, this assumption can be weakened. For a discussion see Remark C.12. Hence the matrix

$$\mathbf{C} = \int_R (c',\ c_*)^T(c',\ c_*) = \begin{pmatrix} \mathbf{c}_{11} & \mathbf{c}_{12} \\ \mathbf{c}_{21} & \mathbf{c}_{22} \end{pmatrix} \tag{C.7}$$

is positive definite. Additionally set

$$\rho = \frac{\mathbf{c}_{12}}{\mathbf{c}_{22}}. \tag{C.8}$$

Next define

$$\varsigma = \int_R i|i|c^2 \tag{C.9}$$

and

$$\mathbf{T} = \begin{pmatrix} \tau & \varsigma \\ \varsigma & \tau - 1 \end{pmatrix} \tag{C.10}$$

The assumptions < $A1$ > - < $A4$ > and an elementary integration yield the relations

$$\int_R c'c = \int_R c_* c = \int_R i c_* c = \int_R i^2 c' c = 0, \quad \int_R |i| c_* c = \int_R i c' c = -\frac{1}{2},$$

$$\int_R i|i|c'c = -1, \ \int_R ic'c_* = \mathbf{c_{22}}, \ \int_R i(c')^2 = \mathbf{c_{12}}, \tag{C.11}$$

$$\int_R (|i|-1)c'c = -\frac{1}{2}\phi, \ \phi = \int_R [\text{sign}(y)]c^2(y)dy.$$

As follows from the begining of Section C[0] special role is played by the vector

$$r = (r_1, r_2) = (-bc'\frac{u}{\sigma}, -bc'\frac{v}{\sigma}) \in \mathcal{H}^{k+q}. \tag{C.12}$$

Set

$$r^* = (r_1^*, r_2^*) = \Pi(r|\mathcal{J}) = (\Pi(r_1|\mathcal{J}), \Pi(r_2|\mathcal{J}))$$

and define

$$r^0 = r - r^* = (r_1^0, r_2^0). \tag{C.13}$$

We close this section by stating two simple propositions which shall be exploited in the sequel.

Proposition C.1. Let $\mathcal{Y}$ be a Hilbert space and $\mathcal{D}$ its subspace [not necessary closed]. Let $y, f_1, ..., f_j \in \mathcal{Y}$ and set $\mathcal{F} = \text{lin}\{f_1, ... f_j\}$, $\mathcal{Y}_0 = \mathcal{D} + \mathcal{F}$. Suppose there exist projections $d = \Pi(y|\mathcal{D})$ and $\Pi(f_i|\mathcal{D})$, $i = 1, ..., j$. Define $f_i^0 = f_i - \Pi(f_i|\mathcal{D})$ and $\mathcal{F}^0 = \text{lin}\{f_1^0, ..., f_j^0\}$. Then, $\Pi(y|\mathcal{Y}_0)$ exists and $\Pi(y|\mathcal{Y}_0) = \Pi(y|\mathcal{D}) + \Pi(y - d|\mathcal{F}^0)$.

Proof of this lemma is similar to the proof of Theorem 5, p. 444, of Bickel et al. (1993) [BKRW in what follows] and therefore omitted.

Proposition C.2. Let $\Phi_1 = (f_1, ... f_j)$ and $\Phi_2 = (f_{j+1}, ... f_l)$ be two vectors with components from a Hilbert space $\mathcal{Y}$. Suppose the matrix of scalar products

$$\mathbf{F} = ((\Phi_1, \Phi_2), (\Phi_1, \Phi_2))_{\mathcal{Y}} = \begin{pmatrix} \mathbf{F_{11}} & \mathbf{F_{12}} \\ \mathbf{F_{21}} & \mathbf{F_{22}} \end{pmatrix}$$

is positive definite. Set $\mathcal{F}_2 = \text{lin}\{f_{j+1}, ..., f_l\}$. Then $\Pi(\Phi_1|\mathcal{F}_2) = \Phi_2 \mathbf{F_{22}}^{-1} \mathbf{F_{12}}$.

C[0].1.2. Derivation of $\Pi(r|\mathcal{J})$

The space $\mathcal{J}$ is not a closed subspace of $\mathcal{H}$. Therefore extend $\mathcal{S}$ to

$$\mathcal{S}_0 = L_2(I, \mu),$$

with μ defined by (C.4), and introduce the space

$$\mathcal{J}_0 = \{b_0 c + bc_0 + bc_* s_0 : b_0 \in \mathcal{B}_0, c_0 \in \mathcal{C}_0, s_0 \in \mathcal{S}_0\}$$

and the following subspaces of $\mathcal{J}_0$

$$\mathcal{H}_1 = \{b_0 c : b_0 \in \mathcal{B}_0\},$$

$$\mathcal{H}_2 = \{bc_0 : c_0 \in \mathcal{C}_0\},$$

$$\mathcal{H}_3 = \{bc_* s_0 : s_0 \in \mathcal{S}_0\}.$$

Obviously we have $\mathcal{J}_0 = \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{H}_3$. In succeeding steps we shall investigate properties of the introduced spaces and find $\Pi(\cdot\,|\mathcal{H}_1)$, $\Pi(\cdot\,|\mathcal{H}_3)$, $\Pi(\cdot\,|\mathcal{H}_1 + \mathcal{H}_3)$, $\Pi(\cdot\,|\mathcal{H}_2)$. Finally $\Pi(\cdot\,|\mathcal{J}_0)$ shall be derived as a combination of $\Pi(\cdot\,|\mathcal{H}_1 + \mathcal{H}_3)$ and $\Pi(\cdot\,|\mathcal{H}_2)$. An observation that for r given in (C.12) it holds $\Pi(r\,|\mathcal{J}_0) \in \mathcal{J}$ [see (C.24), below] obviously implies $r^* = \Pi(r\,|\mathcal{J})$ exists [cf. (C.13)] and $r^* = \Pi(r\,|\mathcal{J}_0)$.

Lemma C.3. Subspaces $\mathcal{H}_i$, $i = 1, 2, 3$ are closed in $\mathcal{H}$. Moreover, $\mathcal{H}_1 \perp \mathcal{H}_3$, $\mathcal{H}_1 \perp \mathcal{H}_2$ and $\mathcal{H}_2 \cap \mathcal{H}_3 = \{0\}$.

Proof. Since $b \in \mathcal{B}$ and c satifies $< A4 >$, it follows that $\mathcal{B}_0$ and $\mathcal{C}_0$ are closed in $\mathcal{B}$ and $\mathcal{C}$, respectively. This and completeness of $\mathcal{S}_0$ imply that the spaces $\mathcal{H}_i$, $i = 1, 2, 3$ are closed in $\mathcal{H}$.

By (C.11),

$$\int\int_{IR} (b_0 c)(bc_* s_0) = \left[\int_I b_0 b s_0\right]\left[\int_R c_* c\right] = 0.$$

So, $\mathcal{H}_1 \perp \mathcal{H}_3$. Moreover, by the definition of $\mathcal{B}_0$ and $\mathcal{C}_0$

$$\int\int_{IR} (b_0 c)(bc_0) = \left[\int_I b_0 b\right]\left[\int_R c_0 c\right] = 0.$$

This proves $\mathcal{H}_1 \perp \mathcal{H}_2$.

Suppose now $h \in \mathcal{H}_2 \cap \mathcal{H}_3$. Hence $h = bc_* s_0 = bc_0$ $\lambda \times \lambda$-a.e. for some $s_0 \in \mathcal{S}_0$ and $c_0 \in \mathcal{C}_0$. So, on the set $\{b > 0\}$ we have $c_* s_0 = c_0$ $\lambda \times \lambda$-a.e. This, however, implies that s_0 is constant on $\{b > 0\}$ and $c_0 = (\text{const})c_*$. By (C.11), $\int_R |i| c_* c = -\frac{1}{2}$ which means that $c_* \notin \mathcal{C}_0$ and the only allowable s_0 and c_0 are such that $c_0 = 0$, $bs_0 = 0$. Hence, $\mathcal{H}_2 \cap \mathcal{H}_3 = \{0\}$. □

Lemma C.4. For any $h \in \mathcal{H}$ we have

$$\Pi(h|\mathcal{H}_1) = \left[\int_R hc - \overline{h}b\right] c, \tag{C.14}$$

where $\overline{h}$ is defined in (C.3),

$$\Pi(h|\mathcal{H}_2) = \left[\int_I hb - \overline{h}c - \left(\overline{ih},\ \overline{(|i|-1)h}\right) \mathrm{T}^{-1}(i,\ |i|-1)^T\right] b, \tag{C.15}$$

$$(C.16) \qquad \Pi(h|\mathcal{H}_3) = \frac{1}{\mathbf{c}_{22}} \left[\int_R hc_* \right] c_*,$$

where $\mathbf{c}_{22}$ is given in (C.7),

$$(C.17) \qquad \Pi(h|\mathcal{H}_1 \oplus \mathcal{H}_3) = \Pi(h|\mathcal{H}_1) + \Pi(h|\mathcal{H}_3).$$

Proof. Observe that $\mathcal{B} = \mathcal{B}_0 \oplus \text{lin}\{b\}$. Let $B : \mathcal{B} \to \mathcal{H}$ be a linear operator given by $B\tilde{b} = \tilde{b}c$. Then the range of B has the form $\mathcal{H}_B = \mathcal{H}_1 \oplus \text{lin}\{bc\}$. Therefore $\Pi(h|\mathcal{H}_B)$ can be easily derived by use Theorem 2, p. 428 of BKRW and Proposition 3, p. 427 of BKRW. The relations (C.15) and (C.16) follow in a similar way. Note only that proving (C.15) we deal with $\text{lin}\{bc,\ ibc,\ |i|bc\}$ instead of $\text{lin}\{bc\}$ in the case of (C.14). □

Corollary C.5. The subspace $\mathcal{J}_0 = \mathcal{H}_2 + (\mathcal{H}_1 \oplus \mathcal{H}_3)$ is closed in $\mathcal{H}$.

Proof. By Proposition 2, point B, p. 441 of BKRW, it is enough to check that $\Pi(\mathcal{H}_2\ |\mathcal{H}_1 \oplus \mathcal{H}_3)$ is finite dimensional. Indeed, for any $bc_0 \in \mathcal{H}_2$, in view of (C.14), (C.16) and (C.17), we get

$$\Pi(bc_0|\mathcal{H}_1 \oplus \mathcal{H}_3) = -(\overline{bc_0})bc + \frac{1}{\mathbf{c}_{22}} \left[\int_R c_* c_0 \right] bc_* \in \text{lin}\{bc,\ bc_*\}.$$

The proof is complete. □

Theorem C.6. The vector of projections of components of r onto $\mathcal{J}_0$ has the form

$$(C.18) \qquad \Pi(r|\mathcal{J}_0) = -[\rho bc_*]\frac{w}{\sigma} - [(c' - \rho c_*)b + \frac{1}{2\tau} ibc]m,$$

where ρ, m and τ are given by (C.8), (C.5) and $< A3 >$, respectively.

Proof. To simplify the notation we shall consider projections of $r \cdot a_0$, where a_0 is any vector in R^{k+q} and, in particular, any element of the standard basis in R^{k+q}.

By Proposition 1, p. 439 of BKRW, $\Pi(r \cdot a_0|\mathcal{J}_0) = h_2^* + h_4^*$, provided that the functions $h_2^* \in \mathcal{H}_2$ and $h_4^* \in \mathcal{H}_4 = \mathcal{H}_1 \oplus \mathcal{H}_3$ satisfy two following equations

$$(C.19) \qquad h_2^* = \Pi(r \cdot a_0 - h_4^*|\mathcal{H}_2)$$

and

$$(C.20) \qquad h_4^* = \Pi(r \cdot a_0 - h_2^*|\mathcal{H}_4).$$

Write h_2^* and h_4^* in the form $h_2^* = bc_0^*,\ \ h_4^* = b_0^*c + bc_*s_0^*$ for some $c_0^* \in \mathcal{C}_0,\ b_0^* \in \mathcal{B}_0$ and $s_0^* \in \mathcal{S}_0$. Since, by Lemma C.3, $\mathcal{H}_4$ is closed, then using (C.11), (C.16) and the definition of $\mathcal{C}_0$, (C.20) can be written in the equivalent form as

$$(C.21) \qquad h_4^* = b_0^*\dot{c} + bc_*s_0^* = -[\rho bc_*]\frac{w \cdot a_0}{\sigma} - \frac{\chi}{\mathbf{c}_{22}} bc_*,$$

where $\chi = \int_R ic' c_0^*$. Since $\mathcal{H}_4 = \mathcal{H}_1 \oplus \mathcal{H}_3$ the representation of h_4^* is unique and therefore we infer that

$$(C.22) \qquad b_0^* = 0,\ s_0^* = -\rho\frac{w \cdot a_0}{\sigma} - \frac{\chi}{\mathbf{c}_{22}}.$$

Inserting (C.21) and (C.22) into (C.19) and applying (C.15) we obtain the following equivalent form of (C.19)

$$bc_0^* = b\left\{-[m\cdot a_0]\left[c' - \rho c_* + \frac{1}{2}(1,\ \phi-\rho)\mathbf{T}^{-1}(i,\ \ |i|-1)^T c\right] + \right.$$

$$\left. + \frac{\chi}{\mathbf{c_{22}}}[c_* + \frac{1}{2}(0\ 1)\mathbf{T}^{-1}(i,\ |i|-1)^T c]\right\}. \tag{C.23}$$

Dividing both sides of (C.23) by b [cf. $< A1 >$], multiplying by ic' and integrating over R, by (C.10) and (C.11) we get the following relation

$$\frac{\chi}{\mathbf{c_{22}}} = [m\cdot a_0][\phi - \rho - \frac{\varsigma}{\tau}].$$

This formula, the expressions for b_0^*, s_0^* and c_0^* and simple algebra yield the required form of $h_2^* + h_4^*$. □

Corollary C.7. Observe that, by the above,

$$\Pi(r\cdot a_0|\mathcal{J}_0) = \left\{-[\rho b c_*]\frac{w}{\sigma} - [(c'-\rho c_*)b + \frac{1}{2\tau}ibc]m\right\}\cdot a_0 = b_0 c + bc_0 + bc_* s_0$$

with

$$b_0 = 0,\quad s_0 = -\left[\rho\frac{w}{\sigma} - (\phi - \rho - \frac{\varsigma}{\tau})m\right]\cdot a_0,$$

$$c_0 = c' - \rho c_* + \frac{1}{2\tau}ic - \left(\phi - \rho - \frac{\varsigma}{\tau}\right)c_*.$$

By $< A5 >$ it implies $s_0 \in \mathcal{S}$. Moreover, by (C.11) we infer that $c_0 \in \mathcal{C}_0$. This proves that

$$\Pi(r|\mathcal{J}_0) = r^* = \Pi(r|\mathcal{J}). \tag{C.24}$$

By Theorem C.6 and Corollary C.7 it follows

$$r^0 = r - r^* = \left(\frac{w}{\sigma} - m\right)[\rho c_* - c']b + m\left[\frac{1}{2\tau}ibc\right].$$

Remark C.8. Note that for the choice $b = \sqrt{g},\ \ c = \sqrt{f}$, the expression $\frac{2r^0}{\sqrt{fg}}$ coincides, for the model under consideration, with the vector map L_b introduced on p. 375 of Schick (1997) [cf. also p. 382, ibidem]. Note also that $\frac{\Delta_{w\cdot a}\Upsilon_s 2r^0}{\Delta_{w\cdot a}\Upsilon_s\sqrt{fg}}$ is the efficient score in the terminology of our paper.

C[0].1.3. Projections of components of r_1 onto $\mathcal{J} + \mathcal{F}_2$

Recall that for $r = (r_1, r_2) = -(bc'\frac{u}{\sigma}, bc'\frac{v}{\sigma})$ we set $r^* = \Pi(r|\mathcal{J})$ and $r^0 = r - r^* = (r_1^0, r_2^0)$ [cf. (C.13)]. Define $\mathcal{F}_2^0$ to be the space spanned by the components of r_2^0. By Propositions C.1 and C.2 we infer

$$(C.25) \qquad \Pi(r_1|\mathcal{J} + \mathcal{F}_2) = \Pi(r_1|\mathcal{J}) + \Pi(r_1^0|\mathcal{F}_2^0), \quad \Pi(r_1^0|\mathcal{F}_2^0) = r_2^0\mathbf{V}^{-1}\mathbf{M},$$

where the matrices $\mathbf{M}$ and $\mathbf{V}$ are blocks of

$$(C.26) \qquad \mathbf{W} = \int_{IR} (r_1^0,\ r_2^0)^T (r_1^0,\ r_2^0) = \begin{pmatrix} \mathbf{U} & \mathbf{M}^T \\ \mathbf{M} & \mathbf{V} \end{pmatrix}.$$

Theorem C.6 implies the following result:

Proposition C.9. The matrix $\mathbf{W}$ defined in (C.26) has the form

$$\mathbf{W} = \frac{\det \mathbf{C}}{\mathbf{c_{22}}} E_\mu \left(\frac{w}{\sigma} - m\right)^T \left(\frac{w}{\sigma} - m\right) + \frac{1}{4\tau} m^T m.$$

$< A1 >$, $< A5 >$ and (C.5) ensure that $\mathbf{W}$ is positive definite and $\mathbf{V}^{-1}$ exists. Therefore (C.25) follows by Proposition C.2.

Set now

$$(C.27) \qquad \tilde{w} = (\tilde{u}, \tilde{v}) = \frac{w}{\sigma} - m = \left(\frac{u}{\sigma} - m_1, \frac{v}{\sigma} - m_2\right).$$

Corollary C.10. By (C.18), (C.24), (C.25) and (C.26) we get

$$\ell_1^0 = r_1 - \Pi(r_1|\mathcal{J} + \mathcal{F}_2) = r_1^0 - r_2^0\mathbf{V}^{-1}\mathbf{M} = r^0(\mathbf{I} - \mathbf{V}^{-1}\mathbf{M})^T =$$

$$= b[\rho c_* - c'](\tilde{u} - \tilde{v}\mathbf{V}^{-1}\mathbf{M}) + \frac{i}{2\tau} bc(m_1 - m_2\mathbf{V}^{-1}\mathbf{M}),$$

where $\mathbf{I}$ is $k \times k$ identity matrix.

C[0].2. Efficient score vector for testing $\theta = 0$ in $P_{\theta,\eta}$

Following comments given at the begining of Section C[0], take $b = \sqrt{g}$, $c = \sqrt{f}$ and r^0 defined with the use of them. Recall also that $\kappa = (\theta, \eta)$, $\eta = (\beta, \sqrt{g}, \sqrt{f}, s)$ and set $\kappa_0 = (0, \eta)$.

Recall that hitherto needed model assumptions $< M[0]1 >$ and $< M[0]2 >$ were stated in Introduction and Section B[0].2, respectively. Now, we impose the following one [cf. $< A1 >$ and $< A2 >$].

$< \mathrm{M}[0]3 >$ $\qquad g > 0\ \ \lambda - a.e., \quad f > 0\ \ \lambda - a.e.$

Theorem C.11. Suppose the assumptions $< M[0]1 >$, $< M[0]2 >$ and $< M[0]3 >$ hold. Then the efficient score vector for testing $\kappa = \kappa_0$ in P_κ is given by

$$\ell^* = \frac{\Delta_{v\cdot\beta}\Upsilon_s 2\ell_1^0}{\Delta_{v\cdot\beta}\Upsilon_s\sqrt{fg}}. \tag{C.28}$$

Consequently,

$$\ell^*(z;\kappa_0) = [\Psi^*(x,y)]\left(\tilde{u}(x) - \tilde{v}(x)\mathrm{V}^{-1}\mathrm{M}\right) + \frac{y - v(x)\beta^T}{\tau\sigma(x)}(m_1 - m_2\mathrm{V}^{-1}\mathrm{M}),$$

where

$$\Psi^*(x,y) = -\frac{f'}{f}\left(\frac{y - v(x)\beta^T}{\sigma(x)}\right) + \rho\left[1 + \frac{y - v(x)\beta^T}{\sigma(x)} \times \frac{f'}{f}\left(\frac{y - v(x)\beta^T}{\sigma(x)}\right)\right], \tag{C.29}$$

$\tilde{u}$ and $\tilde{v}$ are given in (C.27) while $m = (m_1, m_2)$ by (C.6).

Remark C.12. The assumption $f > 0$ λ-a.e. [cf. $< A2 >$ and $< M[0]3 >$] guarantees that with $c = \sqrt{f}$, c' and the related c_* [see (C.2)] are linearly independent and the matrix C given in (C.7) is positive definite. However, this assumption can be considerably relaxed. Indeed, suppose there exist α_1, $\alpha_2 \in R$ such that $\alpha_1 c' + \alpha_2 c_* = 0$. Consider the set $C^+ = \{y \in R : c(y) > 0\}$. On C^+ an equivalent form of the above equation is

$$\alpha_1 cc' + \alpha_2 cc_* = \frac{1}{2}\left[(\alpha_1 + \alpha_2 i)c^2\right]' = 0.$$

This yields $(\alpha_1 + \alpha_2 i)c^2$ to be constant λ-a.e. on each interval contained in C^+. Therefore, as long as $f = c^2 \neq \alpha_0/(\alpha_1 + \alpha_2 i)$ on some intervalcontained in C^+, we are getting a contradiction. The above yields e.g. that Theorem C.11 holds for all error distributions with compact support on R and densities different from a homographic function $\alpha_0/(\alpha_1 + \alpha_2 y)$ on some interval for any constant α_0.

The assumption $g > 0$ λ -a.e. implies $< A1 >$ which was needed for the existence of V^{-1} [cf. Proposition C.9].

In what follows E_κ stands for the expectation calculated under the distribution P_κ. Straightforward calculation yield

$$E_{\kappa_0}\ell^*(Z;\kappa_0) = 0,$$

$$E_{\kappa_0}[\ell^*(Z;\kappa_0)]^T\ell^*(Z;\kappa_0) = 4\int_{IR}[\ell_1^0]^T[\ell_1^0] = 4(\mathrm{U} - \mathrm{M}^T\mathrm{V}^{-1}\mathrm{M}). \tag{C.30}$$

Remark C.13. If $< M[0]3 >$ holds then by Proposition C.9 the matrix $\mathbf{W}$ defined in (C.26) is positive definite. Hence $\mathbf{W}^{-1}$ is positive definite as well and it holds

$$\mathbf{W}^{-1} = \begin{pmatrix} \mathbf{W}^{11} & \mathbf{W}^{12} \\ \mathbf{W}^{21} & \mathbf{W}^{22} \end{pmatrix} \text{ with } \mathbf{W}^{11} = \left(\mathrm{U} - \mathrm{M}^T\mathrm{V}^{-1}\mathrm{M}\right)^{-1}.$$

[cf. Rao 1973, Exercise 1, Ch. 1]. This shows that the covariance matrix of $\ell^*(Z;\kappa_0)$ defined in (C.30) is positive definite.

Set

$$\mathbf{L} = \left\{E_{\kappa_0}[\ell^*(Z;\kappa_0)]^T \ell^*(Z;\kappa_0)\right\}^{-1} = \frac{1}{4}\mathbf{W}^{11}.$$

Corollary C.14. Suppose the assumptions of Theorem C.11 hold. Let $Z_1, ..., Z_n$ be i.i.d. with distribution P_{κ_0}. Set

$$W_k(\kappa_0) = \left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \ell^*(Z_j;\kappa_0)\right\} \mathbf{L} \left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \ell^*(Z_j;\kappa_0)\right\}^T.$$

By (C.30), Remark C.13 and multivariate central limit theorem it holds under $H_0 : \kappa = \kappa_0 = (0, \eta)$ that

$$(C.31) \qquad W_k(\kappa_0) \xrightarrow{D} \chi_k^2,$$

where χ_k^2 denotes a random variable with central chi-square distribution with k degrees of freedom.

C[1]. Homoscedastic model

Below we are keeping the notations and the scheme of the Section C[0] whenever possible. These notations which have to be changed have added the label [1].

C[1].1. Projection of chosen components of $\dot{\Phi}_{(a,b,c)}$

We keep the assumptions $< A1 > - < A2 >$ of Section C[0] while $< A3 >$ - $< A5 >$ are restricted to

$< \mathbf{A3[1]} >$ $$\int_R ic^2 d\lambda = 0, \; \int_R i^2 c^2 d\lambda = \tau, \; \tau \in (0, \infty).$$

$< \mathbf{A4[1]} >$ $c'(y)$ exists for all $y \in R$ and $c' \in \mathcal{C}$.

$< \mathbf{A5[1]} >$ the functions $u_1, ...u_k, v_1, ..., v_q$ are linearly independent and belong to $\mathcal{S}$.

As before we consider

$$\mathcal{B}_0 = \{b_0 \in \mathcal{B} : \int_I b_0 b = 0\} \text{ and } \mathcal{H}_1 = \{b_0 c : b_0 \in \mathcal{B}_0\}.$$

We set

$$\mathcal{C}_{0[1]} = \{c_0 \in \mathcal{C} : \int_R c_0 c = \int_R i c_0 c = 0\}, \; \mathcal{H}_{2[1]} = \{bc_0 : c_0 \in \mathcal{C}_{0[1]}\}.$$

By the definition of $\mathcal{B}_0$ we have $\mathcal{H}_2 \perp \mathcal{H}_{2[1]}$. We put

$$\mathcal{J}_{[1]} = \mathcal{H}_1 \oplus \mathcal{H}_{2[1]} = \{bc_0 + b_0 c : b_0 \in \mathcal{B}_0, \; c_0 \in \mathcal{C}_{0[1]}\}$$

$$\mathcal{F}_{1[1]} = \{bc'[u \cdot \theta_0] : \theta_0 \in R^k\},$$

$$\mathcal{F}_{2[1]} = \{bc'[v \cdot \beta_0] : \beta_0 \in R^q\}.$$

By Lemma C.4, for any $h \in \mathcal{H} = L_2(I \times R, \lambda \times \lambda)$ we have

$$\Pi(h|\mathcal{H}_1) = \left[\int_R hc - \overline{h}b\right] c, \tag{C.32}$$

where $\overline{h}$ is defined in (C.4). Analogusly as (C.15) we get

$$\Pi(h|\mathcal{H}_{2[1]}) = \left[\int_R hb - \overline{h}c - \frac{1}{\tau}(\overline{ih})ic\right] b. \tag{C.33}$$

C[1].1.1. Projections of components of $r_{1[1]}$

As before, for convenience, we shall find projections of $-bc'[w \cdot a_0]$, where a_0 is an arbitrary element of $\mathcal{A} = R^{k+q}$.

By (C.32), the property $\int_R c'c = 0$ [cf. (C.11)] and the definition of $\overline{h}$ we infer that

$$\Pi(-bc'[w \cdot a_0]|\mathcal{H}_1) = 0.$$

Hence

$$\Pi(-bc'[w \cdot a_0]|\mathcal{J}_{[1]}) = \Pi(-bc'[w \cdot a_0]|\mathcal{H}_{2[1]}). \tag{C.34}$$

With μ defined in (C.5) set now

$$m_{[1]} = (m_{1[1]}, m_{2[1]}) = E_\mu w(X).$$

Taking into account the relations $\int_R c'c = 0$ and $\int_R ic'c = -\frac{1}{2}$, [cf. (C.11)], the formulae (C.33) and (C.34) yield

$$\Pi(-bc'[w \cdot a_0]|\mathcal{J}_{[1]}) = -[m_{[1]} \cdot a_0]bc' - \frac{1}{2\tau}[m_{[1]} \cdot a_0]ibc.$$

Set now $r_{[1]} = (r_{1[1]}, r_{2[1]}) = -(ubc', vbc')$. With these notations

$$r^*_{[1]} = \Pi(r_{[1]}|\mathcal{J}_{[1]}) = -[bc' + \frac{1}{2\tau}ibc]m_{[1]}.$$

Additionally define

$$r^0_{[1]} = r_{[1]} - r^*_{[1]}.$$

Remark C.15. Note that taking $b = \sqrt{g}$, $c = \sqrt{f}$ we get

$$\frac{\Delta_{w \cdot a} 2r^0_{[1]}}{\Delta_{w \cdot a} bc}(\bullet) = -(w - m_{[1]}) \left[\frac{f'}{f}\right] (\bullet - w \cdot a) + \frac{i(\bullet - w \cdot a)}{\tau} m_{[1]},$$

where $i(y) = y$. This is efficient score in the terminology of the present paper. Our derivation shows that in the formula $L(x, \theta, \gamma)$ on p. 93 of Schick (1987) the expression $\overline{w}r(x, \theta)$ should be replaced by $\overline{w}r(x, \theta)/\sigma_\gamma$ in the notations of that paper. Observe also that, in contrast to Schick (1997), the standardized function $L(x, \theta, \gamma)$ is named in Schick (1987) efficient score function.

Introduce now

$$\mathbf{W}_{[1]} = \int_{IR} [r^0_{[1]}]^T [r^0_{[1]}] = \begin{pmatrix} \mathbf{U}_{[1]} & \mathbf{M}^T_{[1]} \\ \mathbf{M}_{[1]} & \mathbf{V}_{[1]} \end{pmatrix} =$$

$$\mathbf{c}_{11} E_\mu (w - m_{[1]})^T (w - m_{[1]}) + \frac{1}{4\tau} [m_{[1]}]^T [m_{[1]}].$$

Recall that hitherto imposed model assumptions $< M[1]1 >$ and $< M[1]2 >$ can be found in Introduction and Section B[1].2, respectively. Now, we introduce next model assumption $< M[1]3 >$.

$< \mathrm{M[1]3} >$ $$g > 0 \quad \lambda - a.e.$$

This assumption allows to apply Remark C.13.

Remark C.16. Since $< A5[1] >$ and $< M[1]3 >$ hold $\mathbf{W}_{[1]}$ is positive definite and conclussions of Remark C.13 apply to $\mathbf{W}_{[1]}$ as well.

With the above notations, as in Section C[0].1.3, we define $\mathcal{F}^0_{2[1]}$ and calculate

$$\ell^0_{1[1]} = r_{1[1]} - \Pi\left(r_{1[1]} | \mathcal{J}_{[1]} + \mathcal{F}_{2[1]}\right) = r^0_{1[1]} - r^0_{2[1]} \mathbf{V}^{-1}_{[1]} \mathbf{M}_{[1]} =$$

$$-bc' \left[\tilde{u}_{[1]} - \tilde{v}_{[1]} \mathbf{V}^{-1}_{[1]} \mathbf{M}_{[1]}\right] + \frac{1}{\tau} bc \left[m_{1[1]} - m_{2[1]} \mathbf{V}^{-1}_{[1]} \mathbf{M}_{[1]}\right],$$

where $\tilde{w}_{[1]} = (\tilde{u}_{[1]}, \tilde{v}_{[1]}) = (u - m_{1[1]}, v - m_{2[1]})$.

C[1].2. Efficient score vector for testing $\theta = 0$ in $P_{\kappa_{[1]}}$

Take $b = \sqrt{g}$, $c = \sqrt{f}$, $\kappa_{0[1]} = (0, \eta_{[1]})$, $\eta_{[1]} = (\beta, \sqrt{g}, \sqrt{f})$.

Theorem C.17. Assume $< M[1]1 >$, $< M[1]2 >$ and $< M[1]3 >$ hold. Then the efficient score vector for testing $\kappa_{[1]} = \kappa_{0[1]}$ in $P_{\kappa_{[1]}}$ is given by

$$\ell^*_{[1]} = \frac{\Delta_{v \cdot \beta} 2\ell^0_{1[1]}}{\Delta_{v \cdot \beta} \sqrt{fg}}.$$

Consequently,

$$\ell^*_{[1]}(z; \kappa_{0[1]}) = \left[-\frac{f'}{f}(y - v(x)\beta^T)\right] \left[\tilde{u}_{[1]} - \tilde{v}_{[1]} \mathbf{V}^{-1}_{[1]} \mathbf{M}_{[1]}\right] + \frac{1}{\tau} \left[y - v(x)\beta^T\right] \left[m_{1[1]} - m_{2[1]} \mathbf{V}^{-1}_{[1]} \mathbf{M}_{[1]}\right].$$

Corollary C.18. Suppose the assumptions of Theorem C.17 hold. Let $Z_1, ..., Z_n$ be i.i.d. random variables with distribution $P_{\kappa_{0[1]}}$. Set

$$\mathbf{L}_{[1]} = \frac{1}{4} (\mathbf{U}_{[1]} - \mathbf{M}^T_{[1]} \mathbf{V}^{-1}_{[1]} \mathbf{M}_{[1]})^{-1}$$

and

$$W_k(\kappa_{0[1]}) = \left\{\frac{1}{\sqrt{n}} \sum_{j=1}^n \ell^*_{[1]}(Z_j; \kappa_{0[1]})\right\} \mathbf{L}_{[1]} \left\{\frac{1}{\sqrt{n}} \sum_{j=1}^n \ell^*_{[1]}(Z_j; \kappa_{0[1]})\right\}^T.$$

Then, under $H_{0[1]} : \kappa = \kappa_{0[1]} = (0, \eta_{[1]})$ it holds that $E_{\kappa_{0[1]}} \ell^*_{[1]} = 0$, $\{E_{\kappa_{0[1]}}[\ell^*_{[1]}]^T [\ell^*_{[1]}]\}^{-1} = \mathbf{L}_{[1]}$ and

$$W_k(\kappa_{0[1]}) \xrightarrow{D} \chi^2_k.$$

REFERENCES

Begun, J. M., Hall, W. J., Huang, W.-M. and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric - nonparametric models. *Ann. Statist.* **11** 432-452.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* John Hopkins Univ. Press, Baltimore.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests.* Academia, Prague.

Inglot, T. and Ledwina, T. (2003). Data driven score test of fit for semiparametric homoscedastic linear regression model. *Manuscript.*

Rudin, W. (1974). *Real and Complex Analysis.* McGraw-Hill, New York.

Schick, A. (1987). A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inference* **16** 89-105. Correction (1989) **22** 269-270.

Schick, A. (1994). Efficient estimation in regression models with unknown scale functions. *Math. Meth. Statist.* **3** 171-212.

Schick, A. (1997). Efficient estimates in linear and nonlinear regression with heteroscedastic error. *J. Statist. Plann. Inference* **58** 371-387.

van der Vaart, A. W. (1991). On differentiable functionals. *Ann. Statist.* **19** 178-204.

TADEUSZ INGLOT, INSTITUTE OF MATHEMATICS, WROCŁAW UNIVERSITY OF TECHNOLOGY, WYB. WYSPIAŃSKIEGO 27, 50-370 WROCŁAW, POLAND
e-mail address: inglot@neyman.im.pwr.wroc.pl

TERESA LEDWINA, INSTITUTE OF MATHEMATICS, POLISH ACADEMY OF SCIENCES, UL. KOPERNIKA 18, 51-617 WROCŁAW, POLAND
e-mail address: ledwina@impan.pan.wroc.pl